UNIVERZITET U NOVOM SADU
FAKULTET TEHNIČKIH NAUKA

Lidija Krstanović

MERA SLIČNOSTI IZMEĐU MODELA GAUSOVIH SMEŠA ZASNOVANA NA TRANSFORMACIJI PROSTORA PARAMETARA

DOKTORSKA DISERTACIJA

Novi Sad, 2017.

УНИВЕРЗИТЕТ У НОВОМ САДУ ● **ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА**
21000 НОВИ САД, Трг Доситеја Обрадовића 6

КЉУЧНА ДОКУМЕНТАЦИЈСКА ИНФОРМАЦИЈА

Редни број, **РБР**:		
Идентификациони број, **ИБР**:		
Тип документације, **ТД**:		Монографска документација
Тип записа, **ТЗ**:		Текстуални штампани материјал
Врста рада, **ВР**:		Докторска дисертација
Аутор, **АУ**:		Лидија Крстановић
Ментор, **МН**:		Проф. др Небојша Ралевић
Наслов рада, **НР**:		Мера сличности између модела Гаусових смеша заснована на трансформацији простора параметара
Језик публикације, **ЈП**:		Српски
Језик извода, **ЈИ**:		Српски, Енглески
Земља публиковања, **ЗП**:		Република Србија
Уже географско подручје, **УГП**:		Војводина
Година, **ГО**:		2017
Издавач, **ИЗ**:		Ауторски репринт
Место и адреса, **МА**:		Нови Сад, Факултет техничких наука, Трг Доситеја Обрадовића 6
Физички опис рада, **ФО**: (поглавља/страна/цитата/табела/слика/графика/прилога)		7/145/44/16/98/4/2
Научна област, **НО**:		Примењена математика
Научна дисциплина, **НД**:		Препознавање облика
Предметна одредница/Кључне речи, **ПО**:		Гаусове смеше, Мере сличности, Редукција димензионалности, КЛ-дивергенца
УДК		
Чува се, **ЧУ**:		Библиотека Факултета техничких наука, Трг Доситеја Обрадовића 6, Нови Сад
Важна напомена, **ВН**:		
Извод, **ИЗ**:		Предмет истраживања овог рада је истраживање и експлоатација могућности да параметри Гаусових компоненти коришћених Gaussian mixture модела (ГММ) апроксимативно леже на ниже димензионалној површи уметнутој у конусу позитивно дефинитних матрица. У ту сврху уводимо нову, много ефикаснију меру сличности између ГММ-ова пројектовањем ЛПП-типа параметара компоненти из више димензионалног параметарског оригинално конфигурацијског простора у простор значајно ниже димензионалности. Према томе, налажење дистанце између два ГММ-а из оригиналног простора се редукује на налажење дистанце између два скупа ниже димензионалних еуклидских вектора, пондерисаних одговарајућим тежинама. Предложена мера је погодна за примене које захтевају високо димензионални простор обележја и/или велики укупан број Гаусових компоненти. Разрађена методологија је примењена како на синтетичким тако и на реалним експерименталним подацима.
Датум прихватања теме, **ДП**:		26.01.2017.
Датум одбране, **ДО**:		
Чланови комисије, **КО**:	Председник:	др Ратко Обрадовић,редовни професор
	Члан:	др Лидија Чомић, доцент
	Члан:	др Братислав Иричанин, доцент
	Члан:	др Срђан Попов, ванредни професор
	Члан:	др Владимир Злоколица, доцент
	Члан, ментор:	др Небојша Ралевић, редовни професор

Потпис ментора

УНИВЕРЗИТЕТ У НОВОМ САДУ ● **ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА**
21000 НОВИ САД, Трг Доситеја Обрадовића 6

КЉУЧНА ДОКУМЕНТАЦИЈСКА ИНФОРМАЦИЈА

Accession number, **ANO**:		
Identification number, **INO**:		
Document type, **DT**:		Monographic publication
Type of record, **TR**:		Textual printed material
Contents code, **CC**:		PhD thesis
Author, **AU**:		Lidija Krstanović
Mentor, **MN**:		Professor Nebojša Ralević, PhD
Title, **TI**:		GMMs similarity measure based on transformation of the parameter space
Language of text, **LT**:		Serbian
Language of abstract, **LA**:		Serbian, English
Country of publication, **CP**:		Republic of Serbia
Locality of publication, **LP**:		Province of Vojvodina
Publication year, **PY**:		2017
Publisher, **PB**:		Author's reprint
Publication place, **PP**:		Novi Sad, Faculty of Technical Sciences, Trg Dositeja Obradovića 6
Physical description, **PD**: (chapters/pages/ref./tables/pictures/graphs/appendixes)		7/145/44/16/98/4/2
Scientific field, **SF**:		Applied Mathematics
Scientific discipline, **SD**:		Pattern recognition
Subject/Key words, **S**/**KW**:		Gaussian mixture model, Similarity measures, Dimensionality reduction, KL-divergence
UC		
Holding data, **HD**:		Library of the Faculty of Technical Sciences, Trg Dositeja Obradovića 6, Novi Sad
Note, **N**:		
Abstract, **AB**:		This thesis studies the possibility that the parameters of Gaussian components of a particular Gaussian Mixture Model (GMM) lie approximately on a lower-dimensional surface embedded in the cone of positive definite matrices. For that case, we deliver novel, more efficient similarity measure between GMMs, by LPP-like projecting the components of a particular GMM, from the high dimensional original parameter space, to a much lower dimensional space. Thus, finding the distance between two GMMs in the original space is reduced to finding the distance between sets of lower dimensional euclidian vectors, pondered by corresponding weights. The proposed measure is suitable for applications that utilize high dimensional feature spaces and/or large overall number of Gaussian components. We confirm our results on artificial, as well as real experimental data.
Accepted by the Scientific Board on, **ASB**:		26.01.2017.
Defended on, **DE**:		
Defended Board, **DB**:	President:	Ratko Obradović, PhD, full professor
	Member:	Lidija Čomić, PhD, assistant professor
	Member:	Bratislav Iričanin, assistant professor
	Member:	Srđan Popov, associate professor
	Member:	Vladimir Zlokolica, PhD, assistant professor
	Member, Mentor:	Nebojša Ralević, PhD, full professor

Menthor's sign

Zahvalnica

Pre svega, želim da se zahvalim svom mentoru prof. dr Nebojši Raleviću što me je uveo u za mene nepoznate i neistražene oblasti matematike. Veliko hvala i na tome što mi je dao slobodu u istraživačkom radu.

Veliku zahvalnost dugujem dr Marku Janevu na nesebično prenetom naučno-istraživačkom iskustvu i bezgraničnoj podršci koju mi je pružio u toku izrade rada.

Želim posebno da se zahvalim prof. dr Ratku Obradoviću na interesovanju za moj rad, razumevanju i podršci koju mi pruža.

Hvala dr Vladimiru Zlokolici na diskusijama, podršci i prijateljskoj pomoći.

Jedno veliko hvala i prof. dr Srđanu Popovu koji me je uveo u svet programiranja, i koji mi pruža podršku i daje savete kada mi je to najpotrebnije.

Dr Lidiji Čomić zahvaljujem na savesnom čitanju disertacije, korisnim savetima i sugestijama.

Zahvaljujem se dr Bratislavu Iričaninu na uloženom trudu, stručnim savetima i pomoći koju mi je pružio tokom pisanja ove disertacije.

Hvala svim kolegama sa fakulteta i katedre na ugodnoj radnoj atmosferi, uzajamnoj podršci i pomoći u nastavi i naučno-istraživačkom radu.

Ipak najveću zahvalnost dugujem svojim roditeljima koji su uvek verovali u mene i pružali mi bezuslovnu ljubav i podršku u svemu što radim.

Lidija Krstanović

Predgovor

Potreba za upoređivanjem dve Gausove smeše (eng. *Gaussian Mixture Models*)(GMM) igra bitnu ulogu u različitim zadacima prepoznavanja oblika i predstavlja glavnu komponentu u mnogim ekspertskim sistema i sistemima veštačke inteligencije (eng. *artificial intelligence*)(AI) koji rešavaju probleme iz svakodnevnog života. Kako ovi sistemi često rade sa velikim skupom podataka i koriste vektore osobina koji su velike dimenzionalnosti, od velikog je značaja za njihovu komponentu prepoznavanja da budu računski efikasni u odnosu na dobru tačnost prepoznavanja.

U ovoj tezi je data nova mera sličnosti između GMM-ova, pomoću projektovanja komponenti pojedinačnog Gausijana projektovanjem LPP tipa iz visoko dimenzionalnog originalnog parametarskog prostora u niže dimenzionalni prostor. Dakle, distanca između dva GMM-a u originalnom prostoru je redukovana na nalaženje distance između skupova niže dimenzionalnih Euklidskih vektora, koji su ponderisani odgovarajućim težinama. Na ovaj način je dobijen bolji odnos između tačnosti prepoznavanja i računske složenosti, u odnosu na mere između GMM-ova koje koriste distance između Gausovih komponenti iz originalnog parametarskog prostora.

GMM mera koja je data u ovoj tezi je pogodna za primene u AI sistemima koji koriste GMM-ove i njihove zadatke prepoznavanja, kao i rad sa velikim skupovima podataka i nezaobilaznim velikim brojem svih Gausovih komponenti koje su uključene u rad tih sistema. Data GMM mera je primenjena kako na veštačkim tako i na realnim podacima i daje odličan odnos između tačnosti prepoznavanja i računske složenosti, u poređenju sa ostalim GMM merama sličnosti koje su testirane.

Abstract

The need for a comparison between two Gaussian Mixture Models (GMMs) plays a crucial role in various pattern recognition tasks and is involved as a key components in many expert and artificial intelligence (AI) systems dealing with real-life problems. As those system often operate on large data-sets and use high dimensional features, it is crucial for their recognition component to be computationally efficient in addition to its good recognition accuracy.

In this thesis we deliver the novel similarity measure between GMMs, by LPP-like projecting the components of a particular GMM, from the high dimensional original parameter space, to a much lower dimensional space. Thus, finding the distance between two GMMs in the original space is reduced to finding the distance between sets of lower dimensional Euclidian vectors, pondered by corresponding weights. By doing so, we manage to obtain much better trade-off between the recognition accuracy and the computational complexity, in comparison to the measures between GMMs utilizing distances between Gaussian components evaluated in the original parameter space.

Thus, the GMM measure that we propose is suitable for applications in AI systems that use GMMs in their recognition tasks and operate on large data sets, as the required number of overall Gaussian components involved in such systems is always large. We evaluate the proposed GMM measure on artificial, as well as real-world experimental data obtaining a much better trade-off between recognition accuracy and the computational complexity, in comparison to all baseline GMM similarity measures tested.

Sadržaj

Glava 1

Uvod

1.1 Pregled disertacije

Ova teza je organizovana u sedam glava.

U prvoj glavi je dat pregled disertacije, predmet i cilj istraživanja teze. Ukratko je data motivacija za nastajanje teze na zadatu temu i u kratkim crtama je napravljen osvrt na prethodna naučna istraživanja.

U drugoj glavi su date postojeće metode koje su potrebne za razumevanje ostatka teze. Ovo poglavlje je podeljeno na tri sekcije. U prvoj sekciji su iznete teorijske osnove o Gausovim smešama i EM(eng. *Expectation-Maximization*) algoritmu. U okviru ove sekcije je napravljen osvrt na primenu EM algoritma i Jensenove nejednakosti na Gausove smeše. U drugoj sekciji su obrađene metode redukcije dimenzionalnosti. U ovoj sekciji je posebna pažnja posvećena dobro poznatoj LPP redukciji dimenzionalnosti, jer u ovoj tezi konstruišemo projektovanje LPP tipa za transformaciju originalnog prostora parametara. Treća sekcija je posvećena merama sličnosti između Gausovih smeša. Posebna pažnja je posvećena EMD (eng. *Earth Mover's distance*) i KL-divergenci (eng. *Kullback-Leiber divergence*).

Nova mera sličnosti između modela Gausovih smeša koja je razvijena u okviru ove teze je tema treće glave. U prvoj sekciji je dat ponovni osvrt na neke metode iz grupe naučnih radova u kojima je razmatran problem razvoja efikasne mere sličnosti između Gausovih smeša. U drugoj sekciji je prikazana dobro poznata LPP tehnika učenja površi,

koja će biti modifikovana u ovoj tezi i primenjena na parametarski slučaj Gausovih smeša. U ovoj sekciji je uvedena i nova mera sličnosti između Gausovih smeša, bazirana na transformaciji LPP tipa iz više-dimenzionalnog originalnog prostora parametara u niže-dimenzionalni transformisani prostor. Takođe je prikazana i računska složenost u fazi prepoznavanja za uvedenu meru sličnosti kao i za već poznate mere sličnosti između Gausovih smeša s' kojim sa poredi.

U četvrtoj glavi su prikazani eksperimantalni rezultati kako na sintetičkim, tako i na realnim podacima. Eksperimenti na realnim podacima su izvedeni na tri baze teksture sa zadatkom prepoznavanja, u kojem su korišćene i nekoliko izabranih mera sličnosti zasnovanih na KL-divergenci i EMD. Takođe je izvršeno i poređenje metode prepoznavanja zasnovane na uvedenoj meri sličnosti između Gausovih smeša, s' nekim *state-of-the-art* metodama prepoznavanja teksture. U svim slučajevima uvedena metoda daje mnogo bolji odnos (eng. *trade-off*) između tačnosti prepoznavanja i računske složenosti, u poređenju sa ostalim metodama sa kojima se poredi.

U petoj glavi prikazane su baze tekstura koje su korišćene za izvođenje ekseprimenata na realnim podacima. Pored samih slika tekstura dat je i kratak opis svake baze u smislu ukupnog broja slika, broja klasa kao i specifičnosti svake baze.

Predlozi i ideje za buduća istraživanja, sa kratkim teorijskim uvodom su dati u šestoj glavi.

Zaključak disertacije dat je u sedmoj glavi.

Posle zaključka se nalaze dva priloga sa osnovnim pojmovima iz linearne algebre i verovatnoće koji su nepohodni za razumevanje ove teze. Na kraju disertacije dat je i pregled literature.

1.2 Predmet istraživanja

Predmet istraživanja ove teze je eksploatacija mogućnosti da parametri Gausovih komponenti korišćenih *Gaussian mixture* modela (GMM) aproksimativno leže na niže-dimenzionalnoj površi umetnutoj u konusu pozitivno definitnih matrica. U tu svrhu uvodimo novu, mnogo efikasniju meru sličnosti između GMM-ova projektovanjem LPP-tipa parametara komponenti iz više-dimenzionalnog parametarskog konfiguracijskog prostora u prostor značajno niže dimenzionalnosti.

Dakle, nalaženje distance između dva GMM-a iz originalnog prostora se redukuje na nalaženje distance između dva skupa niže dimenzionalnih euklidskih vektora, ponderisanih odgovarajućim težinama. Na taj način želimo da postignemo dva cilja:

- bolju diskriminativnost između različitih klasa gde pripadajući GMM-ovi zadovoljavaju gore navedeni uslov;
- mera sličnosti između GMM-ova koju predlažemo ima znatno manju računsku složenost u poređenju sa merama sličnosti između GMM-ova, koje koriste distancu između komponenti računatu u originalnom konfiguracijskom domenu.

Prema tome, uvedena mera je pogodna za primene koje zahtevaju visoko dimanzionalni prostor obeležja i/ili veliki ukupan broj Gausovih komponenti. Razrađena metodologija je primenjena kako na sintetičkim, tako i na realnim eksperimentalnim podacima.

Potreba za upoređivanjem dva GMM-a igra bitnu ulogu u različitim zadacima prepoznavanja oblika kao što su verifikacija govornika, upoređivanje slika na osnovu sadržaja (*content based image matching*), prepoznavanje teksture, itd.

Mnogi autori su razmatrali problem konstruisanja efikasne mere sličnosti između GMM-ova kao i odabira različitih deskriptora radi primene u sličnim problemima [9],[11],[12],[27],[1]. Kao najprirodnija informaciona distanca između raspodela p, q, a samim tim i najprikladnija mera sličnosti između GMM-ova, prihvaćena je *Kullback-Leibler* divergenca (KL) [24]. Problem koji nastaje kod KL divergence je što ne

postoji izraz u zatvorenoj formi za KL divergencu između proizvoljnih GMM-ova. Direktan, ali računarski neprihvatljivo zahtevan način računanja KL diveregenci između GMM-ova je primena Monte-Carlo metode [9]. Većina mera koje su prijavljene od strane eksperata zasnivaju se na aproksimaciji pomenute KL divergence. U [11], predložena je jedna aproksimacija KL diveregence između dva GMM-a i primenjena kao efikasna mera sličnosti između slika u problemu *image retrieval*. Isti autori su u [9] predložili gornju i donju granicu za aproksimaciju KL-divergence između GMM-ova, i eksperimentalno potvrdili njihov rezultat na sintetičkim podacima kao i na zadatku verifikacije govornika. Takođe, u [12] isti autori su predložili preciznu i u isto vreme računski efikasnu aproksimaciju KL-divergence baziranu na *Unscented Transform*, s' primenama na zadatak prepoznavanja govornika. U [31] razmatran je prostor multivarijabilnih Gausijana kao Rimanova mnogostrukost i predložena je procedura za uleganje iste u Liovu grupu simetričnih pozitivno definitnih matrica (SPD). Nedavno su autori u [27], motivisani efikasnom primenom *Earth Movers Distance* (EMD) metodologije u raznim zadacima prepoznavanja, prilagodili EMD za GMM-ove u primeni u *Image Matching* zadataku. Oni su prvi predložili metodu SR-EMD retkih reprezentacija (eng. *sparse representations*) zasnovanih na EMD, koristeći osobinu retkosti parametara Gausijana u konfiguracionom prostoru i samim tim dobili efikasniji i robusniji algoritam. Takođe su uveli novu metriku između komponenti Gausijana zasnovanih na informacionoj geometriji.

S' druge strane, jedan od centralnih problema u prepoznavanju oblika, kao i mašinskom učenju u celini, jeste razviti odgovarajući reprezentaciju složenih podataka. Naime, u mnogim primenama mašinskog učenja, kao npr. prepoznavanje objekata, *image retrieval*, prepoznavanje teksture, kategorizacije teksta, *information retrieval*, itd., podaci su predstavljeni u visoko-dimenzionalnom prostoru obeležja. Tehnike redukcije dimenzionalnosti prostora obeležja, kao što su Linerana Diskriminantna Analiza (LDA) [2], kriterijum maksimalne margine (MMC) [27], [29], koje su sa supervizorom, ili npr. Analiza Glavnih Komponenti (*Principal Component Analysis*) (PCA), koje su bez supervizora, bave se tom problematikom pokušavajući da izbegnu probleme kao što su "prokletstvo dimenzionalnosti", računska složenost, kao i diskriminativnost transformisanih obeležja. Redukcija dimenzionalnosti je posebno prisutna u problemima reprezentacije

podataka, koji leže na površi manje dimenzionalnosti, ulegnute u više-dimenzionalni euklidski prostor, tzv. *manifold learning* (ML). Neke od najčešćih ML tehnika su Isomap [37] i *Laplacian Eigenmaps* (LE) [3], *Local Linear Embedding* (LLE) [34]. Metode koje su bazirane na LE koriste Spektralnu teoriju grafova, odnosno vezu između Laplas-Beltrami operatora i Laplasijan grafa, da bi se konstruisala reprezentacija koja očuvava lokalne osobine. S obzirom da je metoda definisana samo na podacima za obuku, ona je dizajnirana samo za primene u raznim zadacima za spektralno klasterovanje. Ipak, metoda bazirana na LE, pod nazivom *Locality Preserving Projections* (LPP) (videti [16]) uspeva da nauči matricu projekcije na način da najbolje "naleže" na pomenutu niže-dimenzionalnu površ, i samim tim na najbolji način očuvava osobinu lokalnosti, tj. informacije o lokalnim susedima su očuvane u transformacionom prostoru. Na taj način dobijamo da ne samo podaci za obuku već i neopservirani podaci mogu biti transformisani u nisko-dimenzionalni prostor čineći tako da je metoda primenljiva u raznim zadacima prepoznavanja oblika i mašinskog učenja. U [10], autori su uopštili neke od pomenutih metoda, kao što su LE i LLE na probleme klasterovanja na proizvoljnim Rimanovim površima i dali primer LE na SPD Rimanovoj površi za primene u zadatku segmentacije slika.

1.3 Cilj istraživanja

Inspirisani sa obe grupe istraživanja, koja su izložena u prethodnoj sekciji, razvili smo novu meru sličnosti između GMM-ova.

Ona koristi tehniku baziranu na LPP da bi naučila projektivnu matricu W koja projektuje parametre GMM-ova (konfiguracioni prostor) na niže-dimenzionalni prostor parametra (transformisani prostor).

U tu svrhu posmatramo vektorizovane parametre (μ_i, Σ_i) koji odgovaraju Gausovim komponentama $\mathcal{N}(\mu_i, \Sigma_i)$ $i = 1, \ldots, M$, gde je M ukupan broj Gausovih komponenti. Oni pripadaju visoko dimenzionalnom euklidskom konfiguracionom prostoru parametara.

Dodeljujemo svaki pomenuti vektor čvoru neorijentisanog ponderisanog grafa. Biramo težine p_{ij} grafa da očuvaju lokalnost parametara koji leže u konusu pozitivno definitnih matrica.

Umesto euklidske distance koristimo uvedenu meru sličnosti između Gausovih komponenti koje odgovaraju čvorovima i i j. Koristimo rastojanja između Gausovih komponenti $N(\mu_i, \Sigma_i)$ i $N(\mu_j, \Sigma_j)$ zasnovana na KL-divergenci ili informacionoj geometriji predloženoj u [27] i [31].

Želimo da postignemo dva cilja:

- Prvo, pretpostavimo da parametri Gausijana koji pripadaju GMM-ovima korišćenim u nekom određenom problemu, leže na nekoj niže-dimenzionalnoj površi umetnutoj u više-dimenzionalnom konfiguracijskom prostoru parametara. Želimo da metod očuva lokalnost indukovanu strukturom površi ulegnutoj u originalni prostor parametara. Tako je moguće postići veću diskriminativnost u problemima prepoznavanja gde parametri GMM-ova zadovoljavaju pretpostavljeni uslov.

- Drugo, mera sličnosti između GMM-ova treba da ima mnogo manju računsku složenost od mera sličnosti između GMM-ova koje koriste distancu između Gausijana računatu u originalnom prostoru parametara. Ovo smo pokazali procenom računske složenosti u etapi prepoznavanja predloženog algoritma kao i bazičnih algoritama, tj. algoritama s' kojima se poredimo. Iz tog razloga, mera sličnosti koju smo uveli je efikasna u problemima koji rade sa visoko dimenzionalnim prostorom obeležja nad kojima se obučavaju GMM-ovi i/ili slučaja kada postoji veliki broj Gausovih komponenti.

Glava 2

Teorijske osnove

2.1 Mašinsko učenje

Mašinsko učenje spada u istraživanja koja se bave formalnim proučavanjem sistema učenja. S obzirom da je ovo veoma interdisciplinarna oblast koriste se znanja iz statistike, računarstva, inženjerstva, kognitivnih nauka, teorije optimizacije i mnogih drugih naučnih disciplina.

Samo mašinsko učenje se može definisati na više različitih načina. Može se definisati kao disciplina koja se bavi pravljenjem prilagodljivih računarskih sistema koji mogu da poboljšaju svoje preformanse učeći iz iskustva. Ali, može se definisati i kao disciplina koja proučava generalizaciju, konstrukciju i analizu algoritama za generalizaciju. Mašinsko učenje, u stvari, teži da napravi teorijski model za ljudsko učenje, pri čemu se trudi da taj model bude što efikasniji i da ga što bolje objasni.

Iako postoje razne vrste procesa učenja i primena mašinskog učenja, ipak postoje zajedničke karakteristike procesa i zadataka učenja. Osnovna klasifikacija problema učenja je na nadgeldano učenje (eng. *supervised learning*) i nenadgledano učenje (eng. *unsupervised learning*).

Kada se algoritmu zajedno sa podacima iz kojih se uči daju i željeni izlazi, tj. ciljne promenljive, govorimo o nadgledanom učenju. Algoritam treba da nauči da za date podatke daje odgovarajuće

izlaze, kao i da za podatke nad kojima nije vršeno učenje izlazi budu takođe dobri. Kada se algoritmu koji uči daju samo podaci bez izlaza govorimo o nenadgledanom učenju. U ovom slučaju algoritam treba da sam uoči neke zakonitosti u podacima koji su mu dati.

Učenje uvek kreće od podataka koji su mu dati. Podatke na osnovu kojih se vrši generalizacija zovemo podacima za trening. Odgovarajući skup onda zovemo trening skup. Pre upotrebe naučenog znanja neophodno je proceniti kvalitet tog znanja. Kvalitet naučnog znanja se testira u odnosu na neke podatke koji su unapred zadati. Te podatke zovemo podaci za testiranje. Podaci za testiranje čine test skup. Test skup i trening skup treba da budu disjunktni.

Nadgledano učenje

Za početak ustanovimo notaciju koju ćemo koristiti. Sa $x^{(i)}$ označavamo "ulazne" promenljive, koje se često zovu i ulazni vektori osobina. Sa $y^{(i)}$ označavamo "izlazne" promenljive koje želimo da predvidimo. Par $(x^{(i)}, y^{(i)})$ zovemo trening primer, a skup m takvih trening primera zaovemo trening skup. Sa Y ćemo označiti prostor svih ulaznih vrednosti, a sa Y svih izlaznih vrednosti (videti [32]).

Cilj je da se nauči funkcija $h : X \to Y$ tako da $h(x)$ dobro predviđa odgovarajuću vrednost y. Funkcija h se zove hipoteza.

Jedan primer nadgledanog učenja, koji ćemo sada ukratko opisati, je linearna regresija (videti [32]). Da bismo sproveli nadgledano učenje moramo da se opredelimo za reprezentaciju funkcije/hipoteze h . Inicijalno izberimo da to bude linearna funkcija po h

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2.$$

Ovde su θ-e parametri (težine) koji pramatrizuju prostor linarnom funkcijom koja preslikava X u Y. Radi lakše notacije, ako ne dolazi do zabune, pišemo $h(x)$ umesto $h_\theta(x)$. Takođe, uzimamo da je $x_0 = 1$. Znači,

$$h(x) = \sum_{i=0}^{n} \theta_i x_i = \theta^T x,$$

gde su s' desne strane jednakosti θ i x vektori, a n je broj ulaznih promenljivih ne računajući x_0.

Neka je dat trening skup. Kako ćemo izabrati, odnosno naučiti, parametar "θ"? Najrazumljiviji metod je da $h(x)$ bude što bliže y, bar za primere trening skupa. Da bismo ovo formalizovali, definišimo funkciju koja će "meriti" to rastojanje za svako θ. U tu svrhu definišemo funkciju troškova

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2.$$

Nenadgledano učenje

Izraz nenadgledano učenje ili učenje bez supervizora se generalno odnosi na to da se koriste opservacije $X_1, \ldots, X_n$ koje su opservirane kao uzorci raspodele $p(X)$ sa ciljem da se opiše ta raspodela. Ova definicija je jako uopštena (videti [32]).

Primeri nenadgledanog učenja su:

- klasterovanje: algoritam kojim se identifikuju "grupe" podataka,
- učenje smeša(eng. *mixture larning*): bitna oblast statistike koja se bavi identifikovanjem parametarskih gustina pojedinačnih populacija,
- PCA (eng. *principal component analysis*): videti 2.3.1,
- *association rule discovery*: *data mining* tehnika koja nalazi kolekcije osobina koje se pojavljuju zajedno u opservacijama,
- *multidimensional scaling*: algoritam koji identifikuje Euklidski prostor malih dimenzija, i moguće nelinearno preslikavanje koje preslikava originalni prostor u novi prostor, u smislu da su rastojanja između parova trening tačaka u originalnom prostoru skoro jednaka rastojanjima između njihovih projekcija.

2.2 Gausove smeše

Konačan model smeše je raspodela oblika

$$p(x) = \sum_{i=1}^{g} \pi_j p(x; \theta_j), \tag{2.1}$$

gde je g broj komponenti smeše, $\pi_j \geq 0$ su "težine" komponeti ($\sum_{j=1}^{g} \pi_j = 1$) i $p(x; \theta_j)$, $j = 1, \ldots, g$, su funkcije gustine komponenti smeše koje zavise od parametra θ. Dakle, imamo tri skupa parametara za estimaciju: π_j, θ_j i g. Gustine komponenti mogu da budu različitog parametarskog oblika. U modelu Gausovih smeša koristi se multivarijabilna normalna raspodela (videti Prilog A).

2.2.1 Jensenova nejednakost

Neka je f funkcija čiji je domen skup realnih brojeva. Podsetimo se teoreme koja u slučaju dvaput diferencijabilne funkcije f daje dovoljan uslov konveksnosti, tj. f je *konveksna funkcija* ako je $f''(x) \geq 0$ (za svako $x \in \mathbb{R}$). Ako f uzima vektor vrednosti kao ulaze ($f : \mathbb{R}^m \to \mathbb{R}^n$), onda je f konveksna ako je njena Hessian matrica H pozitivno semi-definitna ($H \geq 0$). Ako je $f''(x) > 0$ za svako $x \in \mathbb{R}$, tada kažemo da je f *strogo konveksna* funkcija (u slučaju da f uzima vektor vrednosti, odgovarajući uslov je da H mora biti pozitivno definitna, tj. $H > 0$). *Jensenova nejednakost* glasi kao što sledi u teoremi (videti [32]).

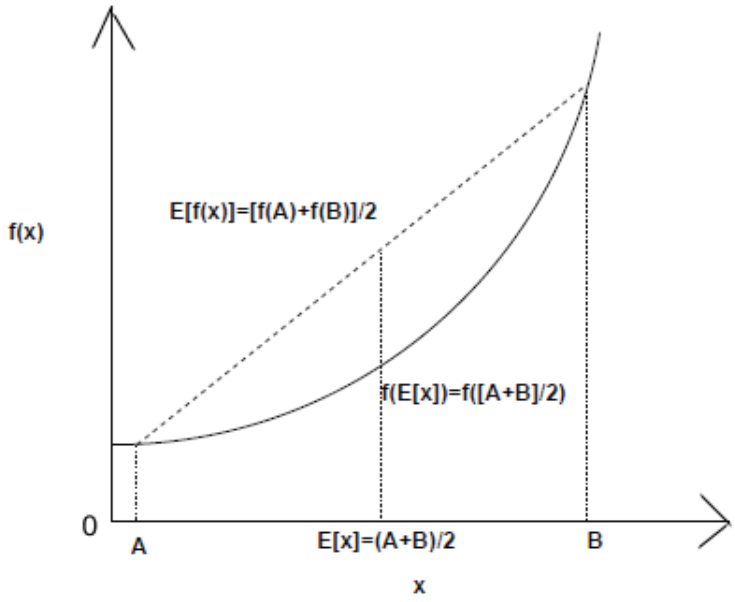
2.2.1 Teorema *Neka je f konveksna funkcija, i neka je X slučajna promenljiva. Tada:*

$$E[f(X)] \geq f(EX).$$

Štaviše, ako je f strogo konveksna, onda $E[f(X)] = f(EX)$ važi ako i samo ako je $X = E[X]$ sa verovatnoćom 1 (tj. ako je X konstanta).

U prethodnoj teoremi smo koristili oznaku $f(EX) = f(E[X])$. Za našu interpretaciju teoreme razmotrimo sliku (2.1).

Ovde je konveksna funkcija f prikazana kao puna linija. Takođe, X je slučajna promenljiva koja ima verovatnoću 0.5 da uzme vrednost A, i 0.5 da uzme vrednost B (prikazane na x-osi). Prema tome, očekivana vrednost X je na sredini između A i B.



Slika 2.1: Jensenova nejednakost

Vrednosti $f(A)$, $f(B)$ i $f(E[X])$ se nalaze na y-osi (na slici označeno sa $f(x)$). Kako je $E[X]$ na sredini između A i B, važi da je $E[f(X)]$ na sredini između $f(A)$ i $f(B)$. Kako je f konveksna funkcija, mora važiti da je $E[f(X)] \geq f(EX)$.

Napomenimo da je f *konkavna* ako i samo ako je $-f$ konveksna (tj. $f'' \leq 0$ ili $H \leq 0$). Odnosno, f *strogo konkavna* ako i samo ako je $-f$ strogo konveksna (tj. $f'' < 0$ ili $H < 0$). Jensenova nejednakost takođe važi za konkavne funkcije, ali je smisao nejednakosti promenjena

$$E[f(X)] \leq f(EX).$$

2.2.2 EM algoritam

Pretpostavimo da imamo problem ocene parametara s' trening skupom $\{x^{(1)}, \ldots, x^{(m)}\}$ koji se sastoji od m nezavisnih opservacija. Mi želimo da napravimo pogodan (eng. *fit*) parametarski model $p(x, z)$ podataka, gde je verodostojnost (eng. *likelihood*) data sa (videti [32])

$$l(\theta) = \sum_x \log p(x; \theta) = \sum_x \log \sum_z p(x, z; \theta).$$

Međutim, teško je eksplicitno dobiti ocenu maksimuma verodostojnosti (eng. *maximum likelihood*) parametra θ. Sa $z^{(i)}$-ovima su označene latentne slučajne promenljive. Često je slučaj da je lako naći ocenu maksimuma verodostojnosti ako su već uočeni $z^{(i)}$-ovi.

EM algoritam predstavlja efikasan metod za estimaciju maksimuma verodostojnosti (videti [4]). Eksplicitno nalaženje maksimuma $l(\theta)$ je teško, tako da se umesto toga u više iteracija traži donja granica l ($E-$korak), a zatim vrši optimizacija te donje granice ($M-$korak).

Neka je Q_i za svako i neka raspodela po z-ovima ($\sum_z Q_i(z) = 1$, $Q_i(z) \geq 0$). Razmotrimo sledeće:

$$\begin{aligned} \textstyle\sum_i \log p(x^{(i)}; \theta) &= \textstyle\sum_i \log \sum_{z^{(i)}} p(x^{(i)}, z^{(i)}; \theta) \\ &= \textstyle\sum_i \log \sum_{z^{(i)}} Q_i(z^{(i)}) \frac{p(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})} \\ &\geq \textstyle\sum_i \sum_{z^{(i)}} Q_i(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})}. \end{aligned}$$

U poslednjem koraku 2.2.2 ovog izvođenju korišćena je Jensenova nejednakost, kao i činjenica da je funkcija $f(x) = \log x$ konkavna (jer je $f''(x) = -\frac{1}{x^2} < 0$ na njenom domenu $x \in \mathbb{R}^+$). Izraz

$$\sum_{z^{(i)}} Q_i(z^{(i)}) \left[\frac{p(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})} \right]$$

unutar sume je samo očekivanje izraza $\left[\frac{p(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})} \right]$ u odnosu na $z^{(i)}$ čiji je slučajan odabir vršen po raspodeli $Q_i(z^{(i)})$. Koristeći Jensenovu nejednakost, dobijamo

$$f\left(E_{z^{(i)} \sim Q_i} \left[\frac{p(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})} \right] \right) \geq E_{z^{(i)} \sim Q_i} \left[f\left(\frac{p(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})} \right) \right],$$

gde indeks "$z^{(i)} \sim Q_i$" označava da se radi o očekivanju u odnosu na $z^{(i)}$ koje odgovara raspodeli $Q_i(z^{(i)})$. Sve ovo nam dozvoljava izvodjenje iz koraka 2.2.2 u korak 2.2.2.

Sada, za bilo koji skup raspodela Q_i, formula 2.2.2 daje donju granicu $l(\theta)$. Postoji mnogo mogućih izbora za Q_i-ove; pitanje je koje ćemo od njih izabrati. Ako imamo neki trenutni pretpostavljeni parametar θ čini se prirodnim da pokušamo da dobijemo donju granicu za fiksnu vrednost θ, tj. "pravićemo" nejednakost koristeći jednakosti koje važe za našu određenu vrednost parametra θ.[1]

Da bismo dobili granicu za fiksiranu određenu vrednost θ, potreban nam je korak u našem izvođenju u kom ćemo primeniti Jensenovu nejednakost. Da bi ovo mogli da izvedemo, znamo da je dovoljno da očekivanje bude po "konstantnoj" slučajnoj promenljivoj, tj. zahtevamo da je

$$\frac{p(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})} = c$$

za neku konstantu c koja ne zavisi od $z^{(i)}$. Ovo se jednostavno postiže izborom da je

$$Q_i(z^{(i)}) \propto p(x^{(i)}, z^{(i)}; \theta).$$

Zapravo, kako znamo da je $\sum_z Q_i(z^{(i)}) = 1$ (jer je raspodela), ovo nam dalje daje da važi

$$\begin{aligned} Q_i(z^{(i)}) &= \frac{p(x^{(i)}, z^{(i)}; \theta)}{\sum_z p(x^{(i)}, z; \theta)} \\ &= \frac{p(x^{(i)}, z^{(i)}; \theta)}{p(x^{(i)}; \theta)} \\ &= p(z^{(i)} | x^{(i)}; \theta). \end{aligned}$$

Prema tome, lako smo dobili da su Q_i-ovi *a posterior* rapodele po $z^{(i)}-$ovima za dato $x^{(i)}$ i za postavljen parametar θ.

Za ovaj izbor Q_i-ova 2.2.2 daje donju granicu verodostojnosti l koju pokušavamo da maksimizujemo. Ovo je $E-$ korak. U $M-$koraku algoritma maksimizujemo 2.2.2 u odnosu na parametre da bismo dobili

[1] Videćemo kasnije kako nam ovo omogućuje da dokažemo da $l(\theta)$ monotono raste pri uzastopnim iteracijama EM algoritma.

novo θ. Ponavljamo u više navrata ova dva koraka EM algoritma, kao što sledi:

Ponavljaj do konvergencije {
(E-korak) Za svako i, neka je

$$Q_i(z^{(i)}) := p(z^{(i)}|x^{(i)};\theta)$$

(M-korak) Neka je

$$\theta := \arg\max_{\theta} \sum_i \sum_{z^{(i)}} Q_i(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)};\theta)}{Q_i(z^{(i)})}.$$

}

Šta nam obezbeđuje konvergenciju ovog algoritma (videti [32])? Pretpostavimo da su $\theta^{(t)}$ i $\theta^{(t+1)}$ dva parametra za dve uzastopne iteracije EM algoritma. Dokažimo da je $l(\theta^{(t)}) \leq l(\theta^{(t+1)})$, jer ćemo na taj način pokazati da EM algoritam uvek monotono poboljšava log-verodostojnost. Ključ pokazivanja ovog rezultata leži u našem izboru Q_i-ova. Posebno, u iteraciji EM algoritma u kojoj se naiđe na parametar $\theta^{(t)}$ mi ćemo izabrati da je $Q_i(z^{(i)}) := p(z^{(i)}|x^{(i)};\theta)$. Ranije smo pokazali da ovakav naš izbor osigurava jednakost u Jensenovoj nejednakosti, koja će biti primenjena u 2.2.2. Prema tome, važi

$$l(\theta^{(t)}) = \sum_i \sum_{z^{(i)}} Q_i^{(t)}(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)};\theta^{(t)})}{Q_i^{(t)}(z^{(i)})}.$$

Parametar $\theta^{(t+1)}$ se onda dobija maksimizovanjem desne strane gornje jednakosti. Dakle,

$$\begin{aligned} l(\theta^{(t+1)}) &\geq \sum_i \sum_{z^{(i)}} Q_i^{(t)}(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)};\theta^{(t+1)})}{Q_i^{(t)}(z^{(i)})} && (2.2)\\ &\geq \sum_i \sum_{z^{(i)}} Q_i^{(t)}(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)};\theta^{(t)})}{Q_i^{(t)}(z^{(i)})} && (2.3)\\ &= l(\theta^{(t)}). && (2.4) \end{aligned}$$

Prva nejednakost proizilazi iz činjenice da

$$l(\theta) \geq \sum_i \sum_{z^{(i)}} Q_i(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)};\theta)}{Q_i(z^{(i)})}$$

važi za svaku vrednost Q_i i θ, i posebno važi da je $Q_i = Q_i^{(t)}$, $\theta = \theta^{(t+1)}$. Da bismo dobili jednakost 2.3, koristimo činjenicu da je $\theta^{(t+1)}$ ekspilicitno izabrano da bude

$$\arg\max_{\theta} \sum_i \sum_{z^{(i)}} Q_i(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})}$$

i prema tome ovako dobijena formula za $\theta^{(t+1)}$ mora da bude veća ili jednaka od iste formule dobijene za $\theta^{(t)}$. Konačno, korak koji se koristi da bi se dobilo 2.4 je pokazan ranije, i proizilazi iz toga što smo izabrali $Q_i = Q_i^{(t)}$, $\theta = \theta^{(t+1)}$.

Stoga, EM algoritam dovodi do toga da verodostojnost konvergira monotono. Do konvergencije dolazi ako za neko $\varepsilon > 0$ važi $|l(\theta^{t+1}) - l(\theta^t)| < \varepsilon$.

Napomena: Ako definišemo

$$J(Q, \theta) = \sum_i \sum_{z^{(i)}} Q_i(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})},$$

tada iz našeg prethodnog izvođenja znamo da je $l(\theta) \geq J(Q, \theta)$. EM algoritam se može takođe videti kao koordiniranje uspona od J, u kom je $E-$korak maksimizovanje u odnosu na Q, a $M-$korak maksimizovanje u odnosu na θ.

2.2.3 Gausove smeše i EM algoritam

U ovom poglavlju razmatramo EM (eng. *Expectation-Maximization*) za ocenu parametara gustine raspodele Gausovih smeša (videti [4], [32], Prilog A).

Pretpostavimo da je dat trening skup $\{x^{(1)}, \ldots, x^{(m)}\}$. S obzirom da je u pitanju podučavanje bez supervizora, ove opservacije nisu labelirane.

Želimo model podataka koji je određen raspodelom $p(x^{(i)}, z^{(i)}) = p(x^{(i)}|z^{(i)})p(z^{(i)})$, pri čemu je: $z^{(i)} \sim$ *Multinominalna raspodela* (ϕ) (videti A.12, Prilog A) (gde je $\phi_j \geq 0$, $\sum_{j=1}^{k} \phi_j = 1$, i za parametar ϕ_j je dat sa $p(z^{(i)} = j)$), i $x^{(i)}|z^{(i)} = j \sim N(\mu_j, \Sigma_j)$ (videti A.11, Prilog A). Označimo sa k broj vrednosti koje $z^(i)-$ovi mogu da uzmu. Prema tome, za naš model postavljamo da je svako $x^{(i)}$ generisano slučajnim odabirom $z^{(i)}$ iz $\{1, \ldots, k\}$. Tada $x^{(i)}$ odgovara nekom od k Gausijana koji zavisi od $z^{(i)}$. Ovo odgovara **modelu Gausove smeše** koji je uveden ranije formulom 2.1, pri čemu važi $\pi_j \equiv \phi_j$. $z^{(i)}-$ovi su **latentne** slučajne promenljive, što znači da su one skrivene/neposmatrane.

Parametri našeg modela su ϕ,μ i Σ. Da bismo procenili vrednosti ovih parametara zapišimo verodostojnost (eng. *likelihood*):

$$\begin{aligned} l(\phi, \mu, \Sigma) &= \sum_{1}^{m} \log p(x^{(i)}; \phi, \mu, \Sigma) \\ &= \sum_{1}^{m} \log \sum_{z^{(i)}=1}^{k} p(x^{(i)}|z^{(i)}; \mu, \Sigma)p(z^{(i)}; \phi). \end{aligned}$$

U poslednjem koraku gornjeg izraza smo koristili osobinu A.9 iz Priloga A.

Međutim, ako stavimo da su izvodi ove formule po parametrima i pokušamo to da rešimo, uočićemo da nije moguće naći maksimalnu verodostojnost (eng. *maximum likelihood*) parametara u zatvorenoj formi.

Slučajne promenljive $z^{(i)}$ ukazuju na to kom od k Gausijana odgovara $x(i)$. Napomenimo da ako znamo $z^{(i)}$ problem procene maksimuma verodostojnosti postaje jednostavan. Tada možemo verodostojnost zapisati kao:

$$l(\phi, \mu, \Sigma) = \sum_{1}^{m} \log p(x^{(i)}|z^{(i)}, \mu, \Sigma) + \log p(z^{(i)}; \phi).$$

Maksimizovanjem ovog u donosu na ϕ, μ i Σ daje parametre:

$$\begin{aligned}
\phi_j &= \frac{1}{m}\sum_{i=1}^{m} 1\{z^{(i)} = j\}, \\
\mu_j &= \frac{\sum_{i=1}^{m} 1\{z^{(i)} = j\}x(i)}{\sum_{i=1}^{m} 1\{z^{(i)} = j\}}, \\
\Sigma_j &= \frac{\sum_{i=1}^{m} 1\{z^{(i)} = j\}(x^{(i)} - \mu_j)(x^{(i)} - \mu_j)^T}{\sum_{i=1}^{m} 1\{z^{(i)} = j\}},
\end{aligned}$$

gde je

$$1\{z^{(i)} = j\} = \begin{cases} 1, & z^{(i)} = j, \\ 0, & z^{(i)} \neq j. \end{cases}$$

Uvodimo EM algoritam. To je iterativni algoritam koji se sastoji iz dva koraka. $E-$korak pokušava da "pogodi" vrednosti $z^{(i)}-$ova. $M-$korak vrši ponovnu ocenu modela baziranu na pretpostavci iz $E-$koraka. Kako $M-$korak pretpostavlja da je pogađanje iz $E-$koraka korektno problem maksimizacije lak. Algoritam je dat sa:

Ponavljaj do konvergencije {
(E-korak) Za svako i, j*, neka je*

$$w_j^{(i)} := p(z^{(i)}|x^{(i)}; \theta, \mu, \Sigma)$$

(M-korak) Ponovna ocena parametara

$$\phi_j := \frac{1}{m}\sum_{i=1}^{m} w_j^{(i)},$$

$$\mu_j := \frac{\sum_{i=1}^{m} w_j^{(i)} x^{(i)}}{\sum_{i=1}^{m} w_j^{(i)}}$$

$$\Sigma_j := \frac{\sum_{i=1}^{m} w_j^{(i)}(x^{(i)} - \mu_j)(x^{(i)} - \mu_j)^T}{\sum_{i=1}^{m} w_j^{(i)}}$$

}

U $E-$koraku računamo *posterior* verovatnoće parametara $z^{(i)}$, za zadato $x^{(i)}$, koristeći trenutne vrednosti parametara. Koristeći Bajesovo pravilo, imamo:

$$p(z^{(i)} = j|x^{(i)}; \phi, \mu, \Sigma) = \frac{p(x^{(i)}|z^{(i)} = j; \mu, \Sigma)p(z^{(i)} = j; \phi)}{\sum_{l=1}^{k} p(x^{(i)}|z^{(i)} = l; \mu, \Sigma)p(z^{(i)} = l; \phi)},$$

gde je $p(x^{(i)}|z^{(i)} = j; \mu, \Sigma)p(z^{(i)} = j; \phi)$ dobijeno evaluacijom Gausijana sa centroidom μ_j i kovarijansom Σ_j u tački $x^{(i)}$; $p(z^{(i)} = j; \phi)$ je dato sa ϕ_j. Vrednosti $w_j^{(i)}$ su određene u $E-$koraku.

Označimo

$$w_j^{(i)} = Q_i(z^{(i)} = j) = P(z^{(i)} = j|x^{(i)}; \phi, \mu, \Sigma).$$

U $M-$koraku vršimo maksimizaciju po parametrima ϕ, μ, Σ

$$\begin{aligned}
&\sum_{i=1}^{m} \sum_{z^{(i)}} Q_i(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; \phi, \mu, \Sigma)}{Q_i(z^{(i)})} \\
&= \sum_{i=1}^{m} \sum_{z^{(i)}} Q_i(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; \mu, \Sigma)p(z^{(i)} = j; \phi)}{Q_i(z^{(i)})} \\
&= \sum_{i=1}^{m} \sum_{z^{(i)}} w_j^{(i)} \log \frac{\frac{1}{(2\pi)^{n/2}|\Sigma_j|^{1/2}} \exp(-\frac{1}{2}(x^{(i)} - \mu_j)^T \Sigma_j^{-1}(x^{(i)} - \mu_j))\phi_j}{w_j^{(i)}}.
\end{aligned}$$

Kada nađemo prvi izvod prethodnog izraza po μ_l dobijamo

$$\begin{aligned}
&\nabla_{\mu_l} \sum_{i=1}^{m} \sum_{z^{(i)}} w_j^{(i)} \log \frac{\frac{1}{(2\pi)^{n/2}|\Sigma_j|^{1/2}} \exp(-\frac{1}{2}(x^{(i)} - \mu_j)^T \Sigma_j^{-1}(x^{(i)} - \mu_j))\phi_j}{w_j^{(i)}} \\
&= -\nabla_{\mu_l} \sum_{i=1}^{m} \sum_{z^{(i)}} w_j^{(i)} \frac{1}{2}(x^{(i)} - \mu_j)^T \Sigma_j^{-1}(x^{(i)} - \mu_j) \\
&= \frac{1}{2} \sum_{i=1}^{m} w_l^{(i)} \nabla_{\mu_l}(2\mu_l^T \Sigma_l^{-1} x^{(i)} - \mu_l^T \Sigma_l^{-1} \mu_l) \\
&= \sum_{i=1}^{m} w_l^{(i)}(\Sigma_l^{-1} x^{(i)} - \Sigma_l^{-1} \mu_l).
\end{aligned}$$

Kada izjednačimo ovo sa nulom i rešimo po μ_l, dobijamo:

$$\mu_l := \frac{\sum_{i=1}^{m} w_l^{(i)} x^{(i)}}{\sum_{i=1}^{m} w_l^{(i)}}.$$

Uradimo sad $M-$korak za parametar ϕ_j. Grupisanjem izraza koji zavise samo od ϕ_j, dobijamo da treba da maksimizujemo izraz

$$\sum_{i=1}^{m} \sum_{j=1}^{k} w_j^{(i)} \log \phi_j.$$

Međutim, kako važi da je suma po ϕ_j-ovima jednaka jedan, sledi da je $\phi_j = p(z^{(i)} = j; \phi)$. Koristeći uslov da je $\sum_{j=1}^{k} \phi_j = 1$ dobijamo Lagranžijan

$$L(\phi) = \sum_{i=1}^{m} \sum_{j=1}^{k} w_j^{(i)} \log \phi_j + \beta(\sum_{j=1}^{k} \phi_j - 1),$$

gde je β Lagranžov množitelj. Kad prethodni izraz diferenciramo dobijamo

$$\frac{\partial}{\partial \phi_j} L(\phi) = \sum_{i=1}^{m} \frac{w_j^{(i)}}{\phi_j} + 1.$$

Izjednačavanjem sa nulom, dobijamo

$$\phi_j = \frac{\sum_{i=1}^{m} w_j^{(i)}}{-\beta},$$

tj. $\phi_j \propto \sum_{i=1}^{m} w_j^{(i)}$. Kad iskoristimo uslov $\sum_{j=1}^{k} \phi_j = 1$ dobijamo da je $-\beta = \sum_{i=1}^{m} \sum_{j=1}^{k} w_j^{(i)} = \sum_{i=1}^{m} 1 = m$.[2] Dakle, naš $M-$ korak za parametar ϕ_j glasi

$$\phi_j := \frac{1}{m} \sum_{i=1}^{m} w_j^{(i)}.$$

Postupak je analogan za Σ_j.

[2]Ovde koristmo činjenicu da je $w_j^{(i)} = Q_i(z^{(i)} = j)$ i $\sum_{j=1}^{k} \phi_j = 1$.

2.3 Metode redukcije dimenzionalnosti

U mnogim oblastima veštačke inteligencije, pronalaženju informacija i podataka, često se dešava da podaci suštinski leže u niže-dimenzionalnom prostoru više-dimenzionalnog prostora.

Posmatrajmo npr. crno-belu sliku objekta koja je slikana pri fiksnim uslovima svetla sa kamerom koja se pomera. Svaka slika je obično reprezentovana vrednostima luminanse za svaki piksel. Ako slika ima n^2 piksela (slika je dimenzije $n \times n$), tada svaka slika daje tačke podataka u $\mathbb{R}^{n^2}$. Unutrašnja dimenzionalnost prostora svih slika istog objekta je broj stepeni slobode kamere. U ovom slučaju, prostor koji posmatramo ima prirodnu strukturu niže-dimenzionalne površi koja je ulegnuta u $\mathbb{R}^{n^2}$.

Problem redukcije dimenzionalnosti ima dugu istoriju. Uobičajen pristup je analiza glavnih komponenti (PCA, eng. *Principal component analysis*) i multidimenzionalno skaliranje (eng. *Multidimensional Scaling*)(videti [32]). Razmatrane su i različite metode nelinearnog mapiranja. Najviše njih se svodi na nelinearan problem čije je rešenje dobijeno opadanjem gradijenta, što garantuje dobijanje lokalnog minimuma. Nalaženje globalnog minimuma na efikasan način je teško. Napomenimo da skoriji pristupi generalizacije PCA koje se baziraju na kernelu nemaju ovu manu. Najviše ovih metoda eksplicitno ne razmatraju strukturu površi u kojoj su verovatno smešteni podaci.

2.3.1 Analiza glavnih komponenti-PCA

Ukratko uvodimo PCA algoritam (videti [4],[32]). Pretpostavimo da je dat trening skup $\{x^{(1)}, \ldots, x^{(m)}\}$, $x^{(i)} \in \mathbb{R}^n$. Pre primene PCA treba uraditi predhodnu obradu podataka, tj. normalizaciju srednje vrednosti i varijanse kao što sledi:

- Neka je $\mu = \frac{1}{m} \sum_{i=1}^{m} x^{(i)}$;
- Zamenimo svako $x^{(i)}$ sa $x^{(i)} - \mu$;
- Neka je $\sigma_j^2 = \frac{1}{m} \sigma_i (x_j^{(i)})^2$;
- Zamenimo svako $x_j^{(i)}$ sa $x_j^{(i)}/\sigma_j$.

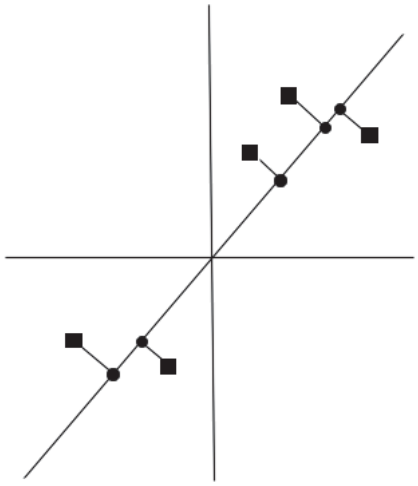
Prva dva koraka obezbeđuju da centorida bude u nuli i ova dva koraka se mogu izbaciti ukoliko se dogodi da je centorida podataka već nula. Druga dva koraka preračunavaju svaku koordinatu tako da ima jediničnu varijansu, što omogućava da se različite osobine tretiraju na isti način, tj. na istoj "skali". Druga dva koraka mogu se izostaviti ako znamo da se različite osobine mere na istoj skali.

Kada smo završili sa normalizacijom, treba da odredimo "glavnu osu", odnosno njen pravac, na kojoj podaci aproksimativno leže. Jedan način nalaženja pravca "glavne ose" je da se nađe jedinični vektor u tako da je varijansa maksimalna kada se podaci projektuju na pravu koja odgovara ovom vektoru.

Razmotrimo skup podataka kao na slici 2.2. Kvadratići predstavljaju originalne podatke, a kružići predtavljaju projekcije tih podataka na izabrani pravac.

Neka je već urađena normalizacija. Pretpostavimo da smo proizvoljno izabrali pravac u kao na slici 2.2. Vidimo da je varijansa zaista velika, ali su kružići daleko od nule (tj. koordinatnog početka). Ako ipak izaberemo pravac kao što je prikazan na slici 2.3, dobijamo da su podaci bliže koordinatnom početku, a da je varijansa mnogo manja.

Želimo da automatski izaberemo pravac u kao što je to zadato na slici 2.2. Da bismo ovo formalizovali, napomenimo da za dati vektor u i dužina projekcije na u je data sa $x^T u$. Ako je $x^{(i)}$ tačka našeg skupa podataka (koja je na prethodnima slikama prikazana kao kvadratić), onda njena projekcija na u (koja je prikazana kružićem) jeste rastojanje $x^T u$ od koordinatnog početka. Prema tome, da bismo



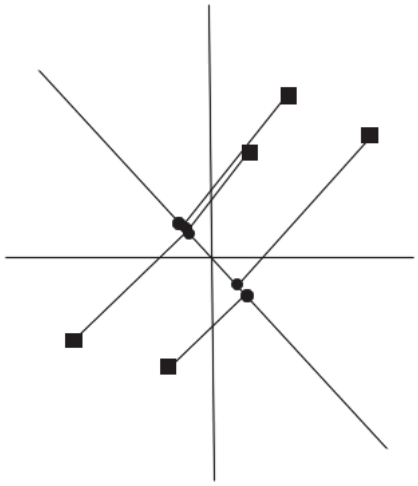
Slika 2.2: PCA primer

izvršili maksimizovanje varijanse projektovanih podataka, biramo jedinični vektor u tako da maksimizujemo sledeće (videti [4], [32]):

$$\begin{aligned}\frac{1}{m}\sum_{i=1}^{m}\left(x^{(i)^T}u\right)^2 &= \frac{1}{m}\sum_{i=1}^{m}u^T x^{(i)}x^{(i)^T}u \\ &= u^T\left(\frac{1}{m}\sum_{i=1}^{m}x^{(i)}x^{(i)^T}\right)u.\end{aligned}$$

Lako prepoznajemo da maksimiziranjem gornjeg izraza pri uslovu $\|u\|_2 = 1$ daje glavni karakteristični vektor $\Sigma = \frac{1}{m}\sum_{i=1}^{m}x^{(i)}x^{(i)^T}$, što je kovarijansna matrica podataka (ako pretpostavimo da je centroida nula, videti Prilog B).

Dakle, ako želimo da nađemo 1-dimenzionalani podprostor na kom aproksimativno leže podaci, biramo u da bude glavni karakteristični vektor Σ. Ako želimo da projektujemo naše podatke na k-dimenzionalni podprostor ($k < n$) treba da izaberemo $u_1, \ldots, u_k$



Slika 2.3: PCA primer

koji odgovaraju k najvećim karakterističnim vektorima. Sada u_i-ovi formiraju novu, ortonormiranu bazu podataka.

Da bismo reprezentovali $x^{(i)}$ u ovoj bazi, potrebno je da samo izračunamo odgovarajući vektor (videti [32])

$$y^{(i)} = \begin{bmatrix} u_1^T x^{(i)} \\ u_2^T x^{(i)} \\ \cdots \\ u_k^T x^{(i)} \end{bmatrix}$$

koje je iz $\mathbb{R}^k$. Kada je $x^{(i)} \in \mathbb{R}^n$, vektor $y^{(i)}$ daje nižu k-dimenzionalnu reprezentaciju/aproksimaciju za $x^{(i)}$. PCA se, dakle, često koristi kao algoritam za redukciju dimenzionalnosti. Vektori $u_1, \ldots, u_k$ se nazivaju prvim k glavnim komponentama podataka.

2.3.2 LPP

Dok PCA teži da očuva globalnu strukturu podataka, *Locality preserving projections* (LPP) teži da očuva lokalnost (tj. susedstvo) strukture podataka (videti [16]). Intuitivno, LPP zadržava veću diskriminativnost informacija nego PCA, pretpostavljajući da uzorci iz iste klase su približno blizu jedan drugom kao što je to u početnom prostoru. Ako koristimo istu notaciju kao za PCA, ciljna funkcija koju koristi LPP je definisana kao što sledi

$$\min_{\mathbf{w}} \sum_{i,j} (y_i - y_j)^2 p_{i,j}, \tag{2.5}$$

gde je $y_i = \mathbf{w^T x_i}$, $i = 1, 2, \ldots, n$ i $\mathbf{P} = (p_{i,j})_{n \times n}$ je matrica sličnosti definisana na sledeći način

$$p_{i,j} = \begin{cases} e^(- \|\mathbf{x}_i - \mathbf{x}_j\|^2/t, & \text{ako je } x_i \text{ u } kNN \text{ od } x_j \text{ ili je } x_j \text{ u } kNN \text{ od } x_i \\ 0, & \text{inače.} \end{cases}$$

Minimizovanjem 2.5 želi se postići da ako su dve tačke $\mathbf{x}_i$ i $\mathbf{x}_j$ blizu jedna drugoj u početnom prostoru, onda će biti blizu jedna drugoj u korespodentnom transformisanom prostoru. Kada izvršimo jednostavno preformulisanje ciljne funkcije dobijamo da treba minimizovati sledeće

$$\begin{aligned} \frac{1}{2} \sum_{i,j} (y_i - y_j)^2 p_{i,j} &= \frac{1}{2} \sum_{i,j} (\mathbf{w}^T \mathbf{x}_i - \mathbf{w}^T \mathbf{x}_j)^2 p_{i,j} \\ &= \mathbf{w}^T \mathbf{X}(\mathbf{D} - \mathbf{P})\mathbf{X}^T \mathbf{w} = \mathbf{w}^T \mathbf{X}\mathbf{L}\mathbf{X}^T \mathbf{w}, \end{aligned} \tag{2.6}$$

gde je $\mathbf{D}$ dijagonalna matrica čiji su ulazi u suštini sume vrsta (ili kolona ako je $\mathbf{P}$ simetrična) $\mathbf{P}$, t.j. $d_{i,i} = \sum_j p_{i,j}$ i $\mathbf{L} = \mathbf{D} - \mathbf{P}$ je Laplasijan matrica. Kada iskoristimo uslov $\mathbf{w}^T \mathbf{X}\mathbf{D}\mathbf{X}^T \mathbf{w} = 1$, dobijamo

$$\min_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{X}\mathbf{L}\mathbf{X}^T \mathbf{w}}{\mathbf{w}^T \mathbf{X}\mathbf{D}\mathbf{X}^T \mathbf{w}} \tag{2.7}$$

Optimalno $\mathbf{w}$ se dobija kao karakteristični vektor koji odgovara minimalnoj karakterisitčnoj vrednosti sledećeg generalizovanog karakterističnog problema

$$\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{w} = \lambda\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{w}. \tag{2.8}$$

U opštem slučaju kada $\mathbb{R}^n$ projektujemo na $\mathbb{R}^k$, $k \ll n$, uzimamo k najmanjih karakterističnih vrednosti. Ovde je opisan slučaj kada je $k = 1$.

2.4 Mere sličnosti

2.4.1 Mere sličnosti zasnovane na aproksimaciji KL divergence

Gausove smeše (eng. *Gaussian Mixture Models*, GMM) se naširoko primenjuju kod modela sa nepoznatom funkcijom gustine (eng. *probability density functions*, PDFs). *Kullback-Leibler* (KL) divergenca izmedju dva PDFs f i g, $D_{KL}(f||g)$ se može koristiti da se uporede raspodele. Zato je česta njena primena u mnogim oblastima, kao što su klasifikacija govornika, procena parametara itd.

Neka su f i g dva PDFs, definisana na $\mathbb{R}^d$, gde je d dimenzija posmatranog vektora x. KL divergenca između f i g je definisana na sledeći način:

$$D_{KL}(f||g) = \int_{\mathbb{R}^d} f(x) \log \frac{f(x)}{g(x)} dx. \tag{2.9}$$

Kada su f i g PDFs za normalne slučajne promenljive (tj. kada su normalne raspodele), tada je

$$\log f(x) = -\frac{1}{2} \log \left((2\pi)^d |\Sigma^f| \right) - \frac{1}{2} (x - \mu^f)^T (\Sigma^f)^{-1} (x - \mu^f),$$

$$f(x) \doteq N(x; \mu^f, \Sigma^f),$$
$$g(x) \doteq N(x; \mu^g, \Sigma^g),$$

gde su μ^f i Σ^g (μ^g i Σ^g, redom) srednja vrednost i kovarijansna matrica od f (repektivno g), T je operator transponovanja i $|\Sigma^f|$ je deteriminanta od Σ^f. KL divergenca za f i g u ovom slučaju ima zatvorenu formu (videti [9]):

$$D_{KL} = \frac{1}{2} \log \frac{|\Sigma^g|}{|\Sigma^f|} + \frac{1}{2} Tr((\Sigma^g)^{-1} \Sigma^f) + \frac{1}{2} (\mu^f - \mu^g)^T (\Sigma^g)^{-1} (\mu^f - \mu^g) - \frac{d}{2}. \tag{2.10}$$

Za GMM-ove, KL-diveregenca nema zatvorenu formu. Stavimo da su f i g PDFs za dva GMM-a. Tada je izraz za f dat sa (analogno se dobija i izraz za g):

$$f(x) = \sum_{a=1}^{A} w_a^f f_a(x) = \sum_{a=1}^{A} w_a^f N(x; \mu_a^f, \Sigma_a^f). \tag{2.11}$$

Sa A i B označavamo broj komponenti GMM za f i g redom, a sa f_a i g_b, $\forall a, b$ pojedinačne normalne PDFs. Moguće je dobiti veoma tačnu aproksimaciju KL divergence između f i g pomoću Monte Karlo metode, ali samo uz veliku računsku složenost. Računski dosta manje složena i dovoljno dobra aproksimacija KL divergence je data u [18]. Ovde ćemo dati gornju i donju granicu KL divergence između dva GMM-a (videti [9]).

Monte Karlo procena (MC). Kao što smo rekli Monte Karlo procenom možemo dobiti veoma tačnu aproksimaciju KL-divergence. Ona se može izraziti kao očekivanje logaritma količnika između f i g. Neka je X (multivarijabilna) slučajna promenljiva, sa PDF f. Tada je:

$$D_{KL}(f||g) = E_X[\log(f(X)/g(X))]. \tag{2.12}$$

MC se može primeniti na procenu ovog očekivanja na sledeći način:

- Izaberi n nezavisnih uzoraka x_i iz PDF f.
- Izračunaj $D_{MC,n}(f||g) = \frac{1}{n}\sum_i \log(f(x_i)/g(x_i))$.

Po zakonu velikih brojeva, $D_{MC,n}(f||g)$ konvergira ka $D_{KL}(f||g)$ kad n teži beskonačnosti.

Aproksimacija proizvoda Gausijana. U [18] je predložena dekompozicija KL-divergence koja daje nekoliko aproksimacija, uključujući i ovu koju ćemo ovde izložiti. Neka je $L_f(g) = E_x[\log g(X)]$, gde je $X \sim f$. KL-divergenca se može dekomponovati na sledeći način:

$$D_{KL}(f||g) = L_f(f) - L_f(g). \tag{2.13}$$

Koristeći Jensenovu nejednakost dobijamo gornju granicu za $L_f(f)$:

$$\begin{aligned} L_f(g) &= \sum_a w_a^f \int_x f_a(x) \log(\sum_b w_b^g g_b(x)) dx \\ &\leq \sum_a w_a^f \log \left(\sum_b w_b^g \int_x f_a(x) g_b(x) dx \right), \\ L_f(g) &\leq \sum_a w_a^f \log \left(\sum_b w_b^g t_{ab} \right), \end{aligned} \tag{2.14}$$

gde je $t_{ab} \doteq \int_x f_a(x) g_b(x) dx$ normalizovana konstanta proizvoda Gausijana. Slično dobijamo

$$\begin{aligned} L_f(f) &\leq \sum_a w_a^f \log \left(\sum_\alpha w_\alpha^f z_{a\alpha} \right), \\ z_{a\alpha} &\doteq \int_x f_a(x) f_\alpha(x) dx. \end{aligned} \tag{2.15}$$

Pretpostavimo da su gornje granice iz 2.14 i 2.15 dovoljno blizu $L_f(g)$ i $L_f(f)$, redom. Koristeći ovu pretpostavku i 2.13, dobijamo aproksimaciju "proizvoda Gausijana", D_{prod} (videti [18]):

$$D_{prod}(f||g) \doteq \sum_a w_a^f \log \frac{\sum_\alpha w_\alpha^f z_{a\alpha}}{\sum_b w_b^g t_{ab}}. \tag{2.16}$$

Varijaciona aproksimacija. Na sličan način se može dobiti i donja granica za $L_f(g)$ i $L_f(f)$ (videti [18]):

$$\begin{aligned} L_f(g) &= E_X[\log(\sum_b w_b^g g_b(x))] \\ &= \sum_a w_a^f \int_x f_a(x) \log \left(\sum_b w_b^g \phi_{ba} \frac{g_b(x)}{\phi_{ba}} \right) \\ &\geq \sum_{ab} w_a^f \phi_{ba} \int_x f_a(x) \log \frac{w_b^g g_b(x)}{\phi_{ba}} dx, \end{aligned}$$

gde je $\phi_{ba} \geq 0$ i $\sum_b \phi_{ba} = 1$, za sve a, b. Maksimizovanjem desne strane gornje jednakosti, u donosu na ϕ_{ba}, dobijamo donju granicu za $L_f(g)$:

$$L_f(g) \geq \sum_a w_a^f \log \sum_b w_b^g e^{-D_{KL}(f_a||g_b)} - \sum_a w_a^f H(f_a) \tag{2.17}$$

gde je $H(f_a)$ entropija f_a data sa $H(f) = \int_{\mathbb{R}} f \ln f dx$, a $D_{KL}(f_a||g_b)$ je dato u zatvorenoj formi u 2.10. Slično, za $L_f(f)$ dobijamo

$$L_f(f) \geq \sum_a w_a^f \log \sum_\alpha w_\alpha^f e^{-D_{KL}(f_a||f_\alpha)} - \sum_a w_a^f H(f_a). \tag{2.18}$$

Prethodne dve granice se kao i ranije mogu iskoristiti da se dobije "varijaciona aproksimacija":

$$D_{var}(f||g) = \sum_a w_a^f \log \frac{\sum_\alpha w_\alpha^f e^{-D_{KL}(f_a||f_\alpha)}}{\sum_b w_b^g e^{-D_{KL}(f_a||g_b)}}. \tag{2.19}$$

Gornja i donja granica za KL divergencu. Koristeći prethodno navedene granice koje su predložene u [18] autori u [9] predlažu malo drugačije granice. Kombinovanjem 2.14 i 2.18 predlažu sledeću donju granicu:

$$\begin{aligned} D_{lower}(f||g) &= \sum_a w_a^f \log \frac{\sum_\alpha w_\alpha^f e^{-D_{KL}(f_a||f_\alpha)}}{\sum_b w_b^g t_{ab}} - \sum_a w_a^f H(f_a) \\ &\leq D_{KL}(f||g). \end{aligned} \tag{2.20}$$

Slično, koriteći 2.15 i 2.17, dobija se izraz za gornju granicu [9]:

$$\begin{aligned} D_{KL}(f||g) &\leq \sum_a w_a^f \log \frac{\sum_\alpha w_\alpha^f z_{a\alpha}}{\sum_b w_b^g e^{-D_{KL}(f_a||g_b)}} + \sum_a w_a^f H(f_a) \\ &\leq D_{KL}(f||g) = D_{upper}(f||g). \end{aligned} \tag{2.21}$$

Srednja vrednost od ove gornje i donje granice je ustvari jednaka srednjoj vrednosti od D_{prod} i D_{var}:

$$\begin{aligned} D_{mean}(f||g) &\doteq [D_{upper}(f||g) + D_{lower}(f||g)]/2 \\ &= [D_{prod}(f||g) + D_{var}(f||g)]/2. \end{aligned} \tag{2.22}$$

2.4.2 EMD rastojanje

Više-dimenzionalne raspodele se često koriste u kompjuterskoj viziji da bi se opisala slika i njene različite osobine; npr. jednodimenzionalna raspodela luminanse slike opisuje oštrinu crno-bele slike, dok tro-dimenzionalna raspodela igra sličnu ulogu kod slike u boji.

Često se vrši "kompresija" ili aproksimacija originalne raspodele sa nekom drugom raspodelom. Više-dimenzionalna raspodela se obično kompresuje particijom potprostora u fiksan broj binova. Tako se dobija histogram. Obično mali broj binova sadrži značajne informacije. Npr. ako imamo sliku nekog pejzaža u kom preovlađuju plavi pikseli za nebo i zeleno-žuti za ostatak, finalni histogram je u tom slučaju neefikasan. S' druge strane, ako imamo sliku karnevala u Riu koja ima obilje boja, odgovarajući histogram će biti takođe neadekvatan. Dakle, histogrami su fiksne strukture i ne mogu da naprave balans između efikasnosti i ekspresivnosti.

Autori u [35] predlažu *deskriptore promenljive dimenzije* koje zovu obeležja (eng. *signatures*). Predlažu da se dominanti klasteri ekstrahuju iz originalne raspodele i iskoriste za kompresovanu reprezentaciju. Opis raspodele je dat preko skupa glavnih klastera ili modova raspodele. Svaki klaster je reprezentovan sa jednom tačkom (centrom klastera) u odgovarajućem prostoru, zajedno sa težinom koja predstavlja veličinu klastera.

Neka su date dve raspodele preko histograma ili obeležja. Korisno je da se definiše mera sličnosti koja će težiti da što bolje aproksimativno opise sličnost, odnosno različitosti te dve raspodele. U [35] autori pokušavaju da oslobode ovu meru od individualnih osobina i konstruišu je tako da ona upoređuje cele raspodele. U tu svrhu uvode *ground distance* o kojoj će biti reči u ovom poglavlju.

Mere sličnosti između histograma

Sistemi za prikupljanje slika (eng. *image retrieval*) obično reprezentuju osobine slike pomoću više-dimenzionalnog histograma. Za takva istraživanja je neophodno definisati meru sličnosti između histograma. U ovom odeljku će biti navedene najčešće korišćene mere

sličnosti između histograma.

Histogram h_i je mapiranje iz d-dimenzionalnog skupa vektora celih brojeva i u skup nenegativnih realnih brojeva. Ovi vektori obično reprezentuju binove (ili njihove centre) u fiksnim delovima relevantnog regiona odgovarajućeg prostora osobina, a odgovarajući realni brojevi predstavljaju frekvenciju koja odgovara datom binu.

Prikazaćemo neke mere sličnosti između dva histograma $H = \{h_i\}$ i $K = \{k_i\}$. Ove mere sličnosti delimo na dve kategorije. *Bin po bin* mere sličnosti porede samo sadržaj odgovarajućih binova koji se porede. Ove mere porede h_i i k_i za svako i, ali ne i h_i i k_j kada je $i \neq j$. *Unakrsno poređenje binova* (eng. *cross-bin*) poredi i binove koji nisu korespodentni. Ovo poređenje koristi *ground distance* d_{ij} koje je definisano kao rastojanje između reprezentativnih vektora osobina za bin i i bin j.

Bin po bin mere. U ovoj kategoriji se porede samo binovi koji su istog indeksa. Mera sličnosti između dva histograma je samo kombinacija mera sličnosti svih ovih parova. Ove mere koriste *ground distance* samo implicitno: osobine koji upadaju u isti bin su dovoljno blizu jedan drugom da bi se razmatrali na isti način, i nisu jako daleko da bi se razmatrala njihova sličnost, odnosno različitost. U ovom smislu bin po bin mere sadrže binarnu *ground distance* sa pragom koji zavisi od veličine bina.

- **Minkowski rastojanje**

$$d_{L_r}(H, K) = \left(\sum_i |h_i - k_i|^r\right)^{1/r}$$

  L_1 rastojanje se često koristi za merenje sličnosti između slika u boji (videti [35]). Druge dve najčešće korišćene mere su L_2 i L_∞.

- **Presek histograma**

$$d_\cap(H, K) = 1 - \frac{\sum_i \min(h_i, k_i)}{\sum_i k_i}$$

  Ova mera je dobra zbog njene mogućnosti da radi parcijalno poređenje kada su oblasti dva histograma različite. Pokazuje se

da kad su oblasti dva histograma iste ova mera je ekvivalentna (normalizovanom) L_1 rastojanju.

- **Kullback-Leiber divergenca i Jeffrey divergenca**
  Kullback-Leiber (KL) divergenca je definisana sa

$$d_{KL}(H,K) = \sum_i h_i \log \frac{h_i}{k_i}.$$

  KL divergenca je nesimetrična i osetljiva na "binovanje" histograma. Jeffrey divergenca je modifikacija KL divergence koja je numerički stabilna, simetrična i robusna na šum i veličinu binova [35]. Definisana je sa:

$$d_J(H,K) = \sum_i (h_i \log \frac{h_i}{m_i} + k_i \log \frac{k_i}{m_i}),$$

  gde je $m_i = \frac{h_i + k_i}{2}$.

- χ^2 **statistika**

$$d_{\chi^2}(H,K) = \sum_i \frac{(h_i - m_i)^2}{m_i},$$

  gde je $m_i = \frac{h_i + k_i}{2}$. Ova mera meri koliko jedinstveno se može prikazati neka populacija nekom raspodelom.

Problem bin po bin mera sličnosti je to što su osetljive na veličinu binova. "Binovanje" koje je previše grubo neće dati dovoljnu diskriminativnost, dok "binovanje" koje je previše fino će staviti slične osobine u različite binove koji se nikad neće ni porediti. Za razliku od njih mere koje unakrsno porede binove uvek daju bolje rezultate na manjim binovima.

Unakrsno poređenje binova. Kada se dodatno koristi i *ground distance* pri merenju sličnosti za svaki *feature*, onda to daje dodatnu informaciju o sličnosti i značajniju meru sličnosti.

- **Kvadratna distanca**

$$d_A(H,K) = \sqrt{(h-k)^T A (h-k)},$$

gde su h i k vektori koji sadrže sve podatke iz H i K. Unakrsno upoređivanje histograma je ovde uključeno matricom sličnosti $A = [a_{ij}]$ gde je a_{ij} sličnost između binova i i j. Ovde su i i j indeksi binova.

- **Rastojanje poređenja** (eng. *match distance*)

$$d_M = \sum_i |\hat{h}_i - \hat{k}_i|,$$

gde je $\hat{h}_i = \sum_{j \leq i} h_j$ kumulativni histogram od h_i, slično za k_i. Ovo rastojanje između dva jednodimenzionalna histograma je definisana kao L_1 distanca između odgovarajućih kumulativnih histograma.

- **Kolmogorov-Smirnov rastojanje**

$$d_{KS}(H, K) = \max_i(|\hat{h}_i - \hat{k}_i|).$$

$\hat{h}_i$ i $\hat{k}_i$ kumulativni histogrami.

- **Parametarski bazirana rastojanje**
Ova metoda prvo, ili implicitno ili eksplicitno, određuje mali skup parametara iz histograma, a zatim poredi te parametre. Tekture se npr. mogu porediti pomoću mere koja je bazirana na periodičnosti, orijentaciji i slučajnosti teksture (videti [35]).

Histogrami ili obeležja

U prethodnom delu smo definisali histogram i popisali neke mere sličnosti između histograma. Histogram je opisan kao rezultat fiksne podele domena raspodele. Naravno, iako su veličine binova fiksne, one su različite u različitim delovima odgovarajućeg prostora vektora osobina. Nažalost, za slike samo mali deo binova sadrži značajnije informacije. Fina kvantizacija histograma onda nije efikana. S' druge strane, kod slika koje sadrže veliku količinu informacija gruba podela histograma nije adekvatna. Slični problemi se javljaju i kada se koristi adaptivni histogram. Dakle, kako je histogram struktura koja je fiksne veličine teško se postiže balans između efikasnosti i izražajnosti (opisa osobina koje bi histogram trebalo da prikaže).

Obeležje $s_j = (m_j, w_j)$ reprezentuje skup osobina klastera (videti [35]). Svaki klaster je prikazan preko svog centroida (ili moda) m_j, i sa delom piksela w_j koji pripadaju tom klasteru. Ceo broj j je indeks koji uzima vrednosti od 1 do broja koji varira u zavisnosti od kompleksnosti slike. Kako je j ceo broj, m_j je d-dimenzionalni vektor. Veličina klastera bi trebalo da je ograničena i da ne prekoračuje obim koji ima sličan ili veoma sličan vektor osobina.

Kako je definicija klastera dosta široka, histogram h_i se može videti kao obeležje $s_j = (m_j, w_j)$ u kojem indeks i vektora skupa klastera definisan fiksnom *a priori* podelom odgovarajućeg prostora. Ako se vektor i preslikava u klaster j, tačka m_j je centralna vrednost u binu i histograma, i w_j je jednako sa h_i (videti [35]).

EMD

Autori u [35] pokušavaju da oslobode rastojanje od individualnih karakterisitika na cele raspodele. Drugim rečima konstrušu konzistentno rastojanje ili sličnost između dve raspodele. Za boje bi to značilo nalazenje distance između raspodela boja. Novo rastojanje između dva obeležja je predloženo u [35] pod nazivom *Earth Mover's distance* (EMD). Ovo je korisno i fleksibilno rastojanje koje je metrika, zasnovano na minimalnim troškovima koji nastaju prilikom transformacije obeležja jednog u drugo. EMD je zasnovano na rešavanju transportnog problema, pomoću linearne optimizacije.

Konstrukicija EMD se može formalizovati kao linearni transportni problem: Neka je $P = \{(p_1, w_{p_1}), \ldots, (p_m, w_{p_m})\}$ prvo obeležje sa m klastera, gde je p_i reprezent klastera i w_{p_i} je težina klastera; $Q = \{(q_1, w_{q_1}), \ldots, (q_n, w_{q_n})\}$ je drugo obeležje sa n klastera; i $D = [d_{ij}]$ je *ground distance* matrica gde su d_{ij} *ground distance* između klastera p_i i q_j. Želimo da nađemo prelaz (eng. *flow*) $\mathbf{F} = [f_{ij}]$, gde je f_{ij} prelaz između p_i i q_j, koje minimizira ukupan trošak

$$WORK(P, Q, \mathbf{F}) = \sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} f_{ij},$$

pri sledećim uslovima

$$f_{ij} \geq 0, \quad 1 \leq i \leq m, 1 \leq j \leq n \tag{2.23}$$

$$\sum_{j=1}^{n} f_{ij} \leq w_{p_i}, \quad 1 \leq i \leq m \tag{2.24}$$

$$\sum_{i=1}^{m} f_{ij} \leq w_{q_j}, \quad 1 \leq j \leq n \tag{2.25}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} = \min(\sum_{i=1}^{m} w_{p_i}, \sum_{j=1}^{n} w_{q_j}). \tag{2.26}$$

Uslov 2.23 dozvoljava prelaz "sadržaja" sa P na Q,a obratno ne. Uslov 2.24 ograničava količinu prelaza, po težinama, koja se prenosi sa klastera iz P. Uslov 2.25 ograničava klastere u Q da ne primaju više "sadržaja" od njihove težine; i uslov 2.26 forsira da se izvrši maksimalni mogući prelaz. Ovaj poslednji uslov zovemo *ukupan prelaz* (eng. *total flow*). Kada se reši transportni problem i kad nađemo optimalni prelaz $\mathbf{F}$, *earth mover's distance* je definisana kao trošak (WORK) koji je normalizovan *ukupnim prelazom*:

$$EMD(P,Q) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} d_{ij} f_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}}.$$

Faktor normalizacije je ukupna težina manjeg obeležja, zbog uslova 2.26. Ovaj faktor je nephodan zbog slučaja u kojem dva obeležja imaju različitu ukupnu težinu, s' ciljem da se izbegne favorizovanje manjih obeležja. U opštem slučaju, *ground distance* d_{ij} može biti bilo koja distanca i bira se u skladu sa problemom koji rešavamo.

Dakle, EMD prirodno proširuje notaciju distance između pojedinačnih elemenata na distancu između skupova, ili raspodela, elemenata. Prednosti EMD u odnosu na ranije definicije distanci raspodela su višestruke. Prvo, EMD se primenjuje na obeležjima. Bolja kompaktnost i fleksibilnost obeležja je njihova prednost, i postojanje distance koja meri rastojanje između ovakvih struktura promenljive veličine je veoma bitno. Drugo, troškovi koje daje *moving*

"earth" odražava se na pojam blizine na odgovarajući način, bez problema kvantizacije koja se javlja kod većine mera, iako histogrami daju iste troškove ako poredimo stavke iz susednih binova . Treće, EMD dozvoljava delimično poređenje na veoma prirodan način. Ovo je veoma bitno u problemima koji se bave okluzijom i šumom u problemima prikupljanje slika (eng. *image retrieval*), gde se porede samo delovi slike. Četvrto, *ground distance* je metrika i ukupne težine dva obeležja su jednake. EMD je prava metrika, što dozvoljava da prostor slike bude metričke strukture.

Pokažimo sada da je EMD metrika (videti [35]). Dakle, želimo da pokažemo da kada obeležja imaju jednake težine i kada je *ground distance* $d(\cdot,\cdot)$ matrika, onda je EMD metrika. Nenegativnost i simetričnost trivijalno važi u svim slučajevima, tako da treba da se dokaže samo nejednakost trougla. Bez umanjenja opštosti, pretpostavimo da je ukupna suma prelaza jednaka sa 1. Neka je f_{ij} optimalni prelaz iz P u Q i g_{ij} je optimalni prelaz iz Q u P. Razmatramo prelaze $P \longmapsto Q \longmapsto R$. Pokažimo sada kako se konstruiše ostvarljiv prelaz iz P u R koji ne daje mnogo više troškova (WORK) nego što je potrebno za optimalno kretanje mase iz P u R kroz Q. Kako je EMD najmanja količina uloženog rada, ova konstrukcija će dovesti do toga da važi nejednakost trougla.

Najveća težina koja se pomera kao jedna jedinica iz p_i u q_j i iz q_j u r_k definiše prelaz koji označavamo sa b_{ijk} gde i, j, k odgovaraju p_i, q_j, r_k, redom. Jasno je da je $\sum_k b_{ijk} = f_{ij}$ prelaz iz P u Q, i $\sum_i b_{ijk} = g_{ij}$ je prelaz iz Q u R. Definišimo sada

$$h_{ik} \doteq \sum_j b_{ijk}$$

prelaz iz p_i u r_k. Ovaj prelaz je ostvarljiv i zadovoljava uslove 2.23-2.26. Uslov 2.23 je zadovoljen zbog konstrukcije koja daje da je $b_{ijk} > 0$. Uslovi 2.24 i 2.25 važe, jer

$$\sum_k h_{ik} = \sum_{j,k} b_{ijk} = \sum_j f_{ij} = w_{p_i}$$

i

$$\sum h_{ik} = \sum_{i,j} b_{ijk} = \sum_{j} g_{jk} = w_{r_k}.$$

Uslov 2.26 važi jer su sva obeležja istih težina. Kako je EMD(P,R) minimalni prelaz iz P u R, i h_{ik} je ispravan prelaz iz P u R i važi

$$\begin{aligned} EMD(P,R) &\leq \sum_{i,k} h_{ik} dp_i, r_k \\ &= \sum_{i,j,k} b_{ijk} dp_i, r_k \\ &\leq \sum_{i,j,k} b_{ijk} dp_i, q_j + \sum_{i,j,k} b_{ijk} dq_j, r_k \qquad (2.27) \\ &= \sum_{i,j} f_{ij} dp_i, q_j + \sum_{j,k} g_{jk} dq_j, r_k \\ &= EMD(P,Q) + EMD(Q,R), \end{aligned}$$

gde 2.27 važi, jer je $d(\cdot,\cdot)$ metrika.

2.4.3 Novo EMD rastojanje

Mere sličnosti slika (odnosno distanca) igraju bitnu ulogu u problemima kao što su klasifikacija, segmentacija, prikupljanje slika (eng. *image retrieval*). Histogram se često koristio za modelovanje slike, dok je merenje sličnosti između njih proučavano decenijama. Ove funkcije sličnosti uglavnom uključuju informacije zasnovane na KL-divergenci ili *Jesson-Shannon* divergenci, χ^2-rastojanju ili l_p normi. EMD može da poredi binove histograma ili obeležja, što su njene prednosti u odnosu na konvencionalne metode zasnovane na merama sličnosti koje upoređuju bin po bin.

Alternativan i široko rasprostranjen metod modelovanja slika je parametrizovan model Gausovih smeša (GMM)(videti [11], [12], [18]). KL-divergenca između GMM-ova se često adaptira za poređenje raspodela. Kako ne postoji zatvorena forma za KL-divergencu, Monte-Carlo procedura je prirodan izbor za njenu aproksimaciju, ali ona ima veliku računsku složenost. U [11] autori se fokusiraju na aproksimaciju KL-divregence. Više o aproksimaciji KL-divergence i merama sličnosti basnovane na njoj se mogu naći koje su u sekciji 2.4.1.

Adaptiranje EMD kao mere sličnosti između GMM-ova je prvi put urađeno za klasifikaciju muzike i kasnije u oblasti vizije na problemu klasifikacijie teksture i vizualnog praćenja. Iako su velike prednosti EMD u poređenju histograma ili obeležja već istražene [35], potencijal EMD-a u merenju sličnosti između GMM-ova ostaje nejasan i slabo istražen. U tu svrhu autori u [27] predlažu novu EMD (*Novel Earth Mover's Distance*) metodologiju baziranu na GMM-ove za poređenje slika. Doprinosi nove EMD su trostruki:

- Autori u [27] uvode SR-EMD koja je *sparse* reprezentacija bazirana na EMD. Upoređujući SR-EMD sa konvencionalnom EMD, SR-EMD je mnogo robustnija na šum slike i mnogo efikasnija, posebno kad se radi o problemima velike dimenzije.
- Pogodna *ground distance* igra bitnu ulogu u efikasnosti EMD-a. Uleganjem prostora Gausijana u Liovu grupu ili proizvod Liovih grupa autori u [27] uvode dve nove *ground distance* između komponenti Gausijana. Za razliku od tradicionalnog rastojanja, nove

ground distances mogu da okarakterišu unutrašnje rastojanje u osnovnoj Rimanovoj površi prostora Gausijana.

- U [27] je predložena i jednostavna još efektivnija EMD metoda nadgledanog učenja za GMM-ove, u cilju adaptacije distance (koja je i metrika) između GMM-ova za specifične aplikacije.

SR-EMD za GMM-ove

U ovoj sekciji ćemo opisati način uvođenja SR-EMD u [27]. Estimacija GMM-ova je postignuta pomoću *Expectation-Maximization* (EM) algoritma (videti 2.2.3). Broj Gausovih komponenti koji najbolje odgovaraju podacima se može estimirati pomoću kriterijuma MDL (eng. *minimum description length*). Estimirani GMM koji karakteriše raspodelu verovatnoće deskriptora slike f se može zapisati

$$G(\mathbf{f}) = \sum_{i=}^{n} w_i N(\mathbf{f}|\mu_i, \Sigma_i)$$

gde je $N(\mathbf{f}|\mu_i, \Sigma_i)$ Gausova komponenta sa *a priori* verovatnoćom w_i, centroidom μ_i i kovarijansnom matricom Σ_i.

Pretpostavimo da imamo dva GMM-a $G_p = \{(g_i^p, w_i^p)\}_{i=1,\ldots,n_p}$ i $G_q = \{(g_i^q, w_i^q)\}_{i=1,\ldots,n_q}$. Označimo sa $d_{i,j}^{pq}$ *ground distance* između g_i^p i g_j^q. EMD je prelaz "svojine" sa minimalnim troškovima sa GMM G_p na GMM G_q sa jediničnim transportnim troškovima d_{ij} (videti [27]). Ovo je modelovano pomoću klasičnog transportnog problema (videti [27]), koji se može zapisati u standardnoj matričnoj formi

$$\begin{aligned} EMD(G_p, G_q) &= \min_{z_{pq}} c_{pq}^T z_{pq}, \\ uz\ uslov\ \ A_{pq} z_{pq} &= y_{pq}, z_{pq} \succeq 0. \end{aligned} \tag{2.28}$$

gde su c_{pq} $n_p n_q$-dimenzionalni vektori koji su vektorizacija *ground distance* matrice $\{d_{ij}^{pq}\}_{n_p \times n_q}$, A_{pq} je $(n_p + n_q) \times (n_p n_q)$ matrica nula i jedinica, i $y_{pq} = [w_1^p, \ldots, w_{n_p}, w_1^q, \ldots, w_{n_q}^q]$ je težinski vektor. Jednačina 2.28 se može rešiti pomoću simpleks algoritma ili algoritma unutrašnje tačke. Kako je $\sum_i^{n_p} w_i = \sum_j^{n_q} w_j = 1$, EMD iz 2.28 sigurno konvergira optimalnom rešenju z_{pq} u kojem su ne-nula vrednosti manji od $(n_p + n_q) \times (n_p n_q)$ (videti [27]). Prema tome z_{pq} je retko (eng.

sparse), npr. ima manje od 20% ne-nula vrednosti ako je $n_p = n_q = 10$.

Generalniji je slučaj kada su podaci kontaminirani šumom, tada je $A_{pq}z_{pq} = y_{pq} + v_{pq}$, gde je v_{pq} Gausov šum sa nulom kao centroidom. Ako se upotrebi osobina 2.28 dobija se $C_{pq}z_{pq} = x_{pq}$, gde je C_{pq} dijagonalna matrica koja na dijagonalama ima c_{pq}. Sada 2.28 možemo zapisati na sledeći način

$$SR - EMD(G_p, G_q) = \tag{2.29}$$

$$\min_{x_{pq}} \frac{1}{2}\|y_{pq} - A_{pq}C_{pq}^{-1}x_{pq}\|_2^2 + \lambda\|x_{pq}\|_1 \tag{2.30}$$

$$uz\ uslov\ \ x_{pq} \succeq 0.$$

gde je $\lambda > 0$ konstanta koja je regularizator, $\|\cdot\|_1$ i $\|\cdot\|_2$ označavaju l_1 i l_2 normu, redom. Kako *ground distance* može biti nula, dodaje se veoma mali realan broj reda 10^{-15}, dijagonalnim elementima od C_{pq} tako da je svaki element veći od nule. Ovde je distanca suma dva izraza: prvi izraz meri grešku sistema lineranih uslova, dok drugi izraz meri transportne troškove. Kako EMD 2.28 uvek ima optimalno rešenje, vrednost prvog izraza je ustvari veoma mali i zanemarljiv u odnosu na drugi izraz. Dodatna osobina 2.29 je ta da je uslov prilagođen i da je funkcija cilja diferencijal u odnosu na uključene parametre.

Formulacija konvencionalne EMD iz 2.28 kao retke reprezentacije problema 2.29 daje dve prednosti: robustnost na šum i efikasnost. Poznato je sa stanovišta Bajesa da ciljna funkcija iz 2.29 vodi ka optimalnom rešenju ako je v_{pq} Gausijan i *prior* raspodela od x_{pq} Laplasijan. Prema tome, retka reprezentacija zasnovana na EMD (SR-EMD) je prirodno više tolerantna na šum nego klasična EMD.

LARS *Homotopy* algoritam je pogodan za rešavanje SR-EMD. Kako algoritam konvergira u ne više od oko $n_p + n_q$ iteracija, i u svakoj iteraciji je računska složenost upoređujuća sa najmanjim kvadratom računska složenost SR-EMD je $O((n_p + n_q)^3 n_p n_q)$. U konvencionalnoj EMD, kako svaka iteracija uključuje operaciju Gausove eleiminiacije sistema uslova, računska složenost je $O(n_l(n_p + n_q)^2 n_p n_q)$, gde n_l označava broj uključenih iteracija. Generalno, imamo $n_l \gg (n_p + n_q)$

i n_l raste multinominalno, ili čak eksponencijalno sa $(n_p + n_q)$ (videti [27]). Dakle, efikasnost SR-EMD je znatno veća od konvencionale EMD, posebno za probleme sa velikom dimenzijom.

Ground distance između komponenti Gausijana

U [27] autori predlažu novu *ground distance* između komponenti Gausijana zasnovanu na teoriji informacija, koja koristi znanje iz statistike i verovatnoće s' tačke gledišta Rimanove geometrije. Uleganjem prostora Gausijana u Liovu grupu ili u proizvod Liovih grupa možemo meriti unutrašnje rastojanje između Gausijana na posmatranoj Rimanovoj mnogostrukosti.

Ground distance bazirana na Liovoj grupi. Prostor višedimenzionalnih Gausijana je Rimanova mnogostrukost i može se ulegnuti u prostor SPD matrica. Neka $\mathcal{N}(\mathbf{0}, \mathbf{I})$ označava k-dimenzionalnu Gausovu raspodelu gde je centroida vektor $\mathbf{0}$ i $\mathbf{I}$ je kovarijansna matrica (jedinična matrica). Sa $|\cdot|$ označavamo determinantu matrice. Poznatao je da ako slučajan vektor $\mathbf{x}$ iz $\mathcal{N}(\mathbf{0}, \mathbf{I})$, tada njegova afina transformacija $\mathbf{Qx} + \mu$ je iz $\mathcal{N}(\mu, \mathbf{\Sigma})$, gde je $\mathbf{\Sigma}$ dekompozicija $\mathbf{\Sigma} = \mathbf{Q^T Q}$, $|\mathbf{Q}| > 0$, i obratno. Dakle, Gausijan $\mathcal{N}(\mu, \mathbf{\Sigma})$ se može okarakterisati afinom transformacijom $(\mu, \mathbf{Q})$. Neka je τ_1 preslikavanje iz afine grupe $\mathcal{AFF}_k^+ = \{(\mu, \mathbf{Q}) | \mu \in \mathbb{R}^k, \mathbf{Q} \in \mathbb{R}^{k \times k}, |\mathbf{Q}| > 0\}$ u generalnu linearnu grupu $\mathcal{SL}_{k+1} = \{\mathbf{A} | \mathbf{A} \in \mathbb{R}^{(k+1)\times(k+1)}, |\mathbf{A}| > 0\}$, i τ_2 je preslikavanje iz $\mathcal{SL}_{k+1}$ u prostor SPD matrica $\mathcal{SPD}_{k+1}^+ = \{\mathbf{P} | \mathbf{P} \in \mathbb{R}^{(k+1)\times(k+1)}, \mathbf{P} = \mathbf{P}^T, |\mathbf{P}| > 0\}$, tj.

$$\tau_1 : \mathcal{AFF}_k^+ \longmapsto \mathcal{SL}_{k+1}$$
$$\tau_2 : \mathcal{SL}_{k+1} \longmapsto \mathcal{SPD}_{k+1}^+$$
$$(\mu, \Sigma) \longmapsto \mathcal{C}_Q \begin{bmatrix} Q & \mu \\ 0^T & 1 \end{bmatrix}$$
$$S \longmapsto SS^T$$

gde je $\mathcal{C}_Q = |Q|^{-1/(k+1)}$. Ova dva preslikavanja mogu da ulegnu k-dimenzionalni Gausijan $\mathcal{N}(\mu, \Sigma)$ u $\mathcal{SPD}_{k+1}^+$. Na taj način jedinstveno

reprezentuju $(k+1) \times (k+1)$ SPD matricu $\mathbf{P}$ (videti [27]). Prema tome važi

$$\mathcal{N}(\mu, \Sigma) \sim P = |\Sigma|^{-\frac{1}{k+1}} \begin{bmatrix} \Sigma + \mu\mu^T & \mu \\ \mu^T & 1 \end{bmatrix} \tag{2.31}$$

SPD matrice čine Liovu grupu koja formira Rimanovu mnogostrukost. U [27] koriste Log-Euklidsku metriku da bi izmerili unutrašnje distance u ovom prostoru. U primeni Log-Euklidske metrike logaritamsko množenje $\odot$ i skalarno logaritamsko množenje $\otimes$ je definisano tako da $\mathcal{SPD}_n^+$ ima strukturu linearnog prostora. Liova lagebra $\mathcal{S}_n$ je $\mathcal{SPD}_n^+$ linearni prostor sa matričnim sabiranjem $+$ i skalarnim matričnim množenjem $\cdot$. Matrično eksponencijalno preslikavanje $exp : \mathcal{S}_n \longmapsto \mathcal{SPD}_n^+$ je injektivno, sirjektivno, glatko preslikavanje i izometrija. Ovo nam dozvoljava da operacije u $\mathcal{S}_n$ budu ekvivalnetne sa ovim operacijama u $\mathcal{SPD}_n^+$. Geodezijsko rastojanje između SPD matrica $\mathbf{P}_1$ i $\mathbf{P}_2$ je $d(\mathbf{P}_1, \mathbf{P}_2) = \| \log(\mathbf{P}_1) - \log(\mathbf{P}_2) \|_F$, gde log označava logaritam matrice, a $\| \cdot \|_F$ označava Frobeniusovu normu.

Neka su $\mathbf{P}_i$ i $\mathbf{P}_j$ ulegnute SPD matrice koje odgovaraju Gausijanima g_i i g_j, redom. U [27] *ground distance* između njih je definisano sa

$$d_{g_i,g_j}(M) = tr((\log(P_i) - \log(P_j))^T M(\log(P_i) - \log(P_j))) \tag{2.32}$$

gde je M SPD matrica. Ako je M jedinična matrica, 2.32 postaje geodezijsko rastojanje između $\mathbf{P}_i$ i $\mathbf{P}_j$. Kako je $d(\mathbf{P}_1, \mathbf{P}_2)$ metrika, $d_{g_i,g_j}(M)$ je takođe matrika. Neka je $M = AA^T$. Tada 2.32 postaje

$$d_{g_i,g_j}(A) = \|A(\log(P_i) - \log(P_j))\|_F. \tag{2.33}$$

Sada se *ground distance* može posmatrati kao linearna transformacija matrice u logaritamskom domenu. Matrica M se može naučiti tako da se prilagodi određenoj primeni ili bazi.

Ground distance bazirana na proizvodu Liovih grupa. n-dimenzionalni Gausijan je određen centroidom $\mu \in \mathbb{R}^n$ i kovarijansnom matricom $\Sigma \in \mathcal{SPD}_n^+$. $\mathbb{R}^n$ je Liova grupa sa operacijom vektorskog sabiranja i $\mathcal{SPD}_n^+$ je Liova grupa sa operacijom logaritamskog množenja $\odot$: $\Sigma_1 \odot \Sigma_2 = \exp(\log(\Sigma_1) + \log(\Sigma_2))$ gde $\Sigma_1, \Sigma_2 \in \mathcal{SPD}_n^+$.

Neka je dat proizvod grupa $\mathbb{R}^n \times \mathcal{SPD}_n^+$ i definisana operacija $\oslash$ (videti [27]) :

$$(\mu_1, \Sigma_1) \oslash (\mu_2, \Sigma_2) = (\mu_1 + \mu_2, \Sigma_1 \odot \Sigma_2). \tag{2.34}$$

Može se pokazati da je ovaj proizvod grupa takođe Liova grupa i njena Liova algebra je $\mathbb{R}^n \times \mathcal{S}_n$.

Rastojanje u Liovoj grupi $\mathbb{R}^n$ između dve centoride se može adekvatno ponderisti sa kovarijansnom matricom. Ovo je opravdano uraditi, jer na taj način se postiže balans između distance u $\mathbb{R}^n$ i u $\mathcal{SPD}_n^+$. Sa ovim obrazloženjem autori u [27] uvode novu *ground distance* na sledeći način:

$$d_{g_i,g_j}(\theta) = (1-\theta)a_{g_i,g_j} + \theta b_{g_i,g_j}, \tag{2.35}$$

gde je

$$a_{g_i,g_j} = \left((\mu_i - \mu_j)^T ({\Sigma_i}^{-1} + {\Sigma_j}^{-1})(\mu_i - \mu_j)\right)^{1/2}$$
$$b_{g_i,g_j} = \| \log(\Sigma_i) - \log(\Sigma_j) \|_F.$$

U 2.35 $\theta \in [0,1]$ je konstanta koja balansira između dva izraza; a_{g_i,g_j} meri razliku između dva centorida i postaje *Mahalanobis* rastojanje ako je $\Sigma_i = \Sigma_j$; b_{g_i,g_j} meri Log-Euklidsku rastojanje između dve kovarijansne matrice, koje je geodezijsko rastojanje u Rimanovom prostoru. d_{g_i,g_j} je metrika (zadovoljava sve aksiome metrike), jer su a_{g_i,g_j} i b_{g_i,g_j} metrike.

2.5 Kovarijansni deskriptori

Lokalni deskriptori se mogu koristiti ili za detekciju objekta na slici (ključne tačke ili tačke od interesa) ili za globalni opis slike (eng. *scene analysis*) kao npr. za opis njene teksture. U ovom radu u sekciji sa eksperimentima 4, oni se koriste za globalni opis teksture.

Željena osobina ekstrahovanih deskriptora je da budu robustni, tj. neosetljivi na promene kao što su skaliranje, šum, promena osvetljaja. To je takođe bitno i za globalni opis slike, samo što se tada lokalni deskriptori uzimaju gusto po slici.

U ovoj tezi koristimo **kovarijansne deskriptore** (videti [36]), koje ćemo ovde ukratko objasniti. Neka je data slika I i neka je ϕ mapiranje koje ekstrahuje n-dimenzionalni vektor osobina z_i za svaki piksel $i \in I$ tako da je

$$\phi(I, x_i, y_i) = z_i$$

gde $z_i \in \mathbb{R}^n$ i (x_i, y_i) je pozicija i-tog piksela. Data regija slike R je reprezentovana sa $n \times n$ kovarijansnom matricom C_R vektora osobina $z_{i i=1}^{|R|}$ piksela iz regije $|R|$. Prema tome, kovarijansni deskriptori regije su dati sa

$$C_R = \frac{1}{|R| - 1} \sum_{i=1}^{|R|} (z_i - \mu_R)(z_i - \mu_R)^T,$$

gde je μ_R centroida,

$$\mu_R = \frac{1}{|R|} \sum_{i=1}^{|R|} z_i.$$

Vektor osobina z obično sadrži informacije o boji.

Iako kovarijansna matrica u opštem slučaju moze biti semidefinitna, sami kovarijansni deskriptori se regulišu dodavanjem male konstante koja je pomnožena sa jediničnom matricom. Na ovaj način kovarijansni deskriptori postaju strogo pozitivno definitni.

Prema tome kovarijansni deskriptori pripadaju $\mathbf{S}_{++}^n$ prostoru $n \times n$ pozitivno definitnih matrica koje formiraju Rimanovu površ. Neka su date dve kovarijansne matrice C_i i C_j. Rimanova distanca je metrika

$d_{geo}(C_i, C_j)$, koja daje geodezijsko rastojanje između dve tačke na ovoj površi. Ona je data sa

$$d_{geo}(C_i, C_j) = \| \log(C_i^{-1/2} C_j C_i^{-1/2} \|_F$$

gde $\log(\cdot)$ označava matrični logaritam, a $\| \cdot \|_F$ Frobeniusovu normu.

Mnogi postojeći algoritmi za klasifikaciju, sa kovarijansnim deskriptorima regije, koriste geodezijsku distancu u okviru kNN-a. Geodezijska distanca se može koristiti u okviru modifikovanog K-*means* algoritma.

Glava 3

Mera sličnosti između modela Gausovih smeša zasnovana na transformaciji prostora parametara

3.1 Pregled postojećih metoda

U ovom poglavlju su izložene poznate mere sličnosti između GMM-ova sa kojima poredimo meru sličnosti uvedenu u ovoj tezi. Ove mere sličnosti su zasnovane na različitim aproksimacijama KL divergence. Takođe, dajemo ponovan kratak osvrt na LPP metodu za redukciju dimenzionalnosti obeležja, jer ćemo je modifikovati radi konstrukcije nove mere sličnosti između GMM-ova ([23]).

3.1.1 Mere sličnosti između GMM-ova

Kao što je već ranije istaknuto, mere sličnosti između GMM-ova igraju veliku ulogu u mnogim primenama prepoznavanja oblika i mašinskog učenja, kao npr. u upoređivanju slika na osnovu sadržaja, prepoznavanju i verifikaciji govornika, prepoznavanju teksture, itd. Kao najprirodnija mera izmađu gustina raspodele p i q je KL divergenca, definisana kao

$$KL(p||q) = \int_{\mathbb{R}^d} p(x) \log \frac{p(x)}{q(x)} dx.$$

Međutim, ne postoji zatvorena forma za njen analitički izraz.

Direktno računanje, moguće je preko standardne Monte-Carlo metode (videti [18]). Ideja je da se uzimaju uzorci gustine f tako da

$$E_f\left[\ln\frac{f(x)}{g(x)}\right] = KL(f||g).$$

Ako za odabir koristimo podjednako raspodeljene nezavisne opservacije x_i,$i = 1,\ldots,N$ Monte-Carlo aproksimacija je data sa:

$$KL_{MC}(f||g) \approx \frac{1}{N}\sum_{i=1}^{N}\ln\frac{f(x_i)}{g(x_i)}. \tag{3.1}$$

Ona je najpreciznija, ali računski najskuplja za primene u realnim problemima. U literaturi autori su predložili razne načine za aproksimaciju KL-divergenci između GMM-ova (videti [11]-[18]).

Najgrublja aproksimacija KL divergence između dva GMM-a je dobijena na osnovu osobine konveksnosti KL-divergence [7]. Naime, neka su data dva GMM-a sa $f = \sum_{i=1}^{n}\alpha_i f_i$ i $g = \sum_{j=1}^{m}\beta_j g_j$, gde su $f_i = \mathcal{N}(\Sigma_i,\mu_i)$ i $f_j = \mathcal{N}(\Sigma_j,\mu_j)$ Gausove komponenete koje pripadaju smešama f i g respektivno, $\alpha_i > 0$, $\beta_j > 0$ odgovarajuće težine i $\sum_i \alpha_i = 1$, $\sum_j \beta_j = 1$ su odgovarajuće težine. Važi

$$KL(f||g) \leq \sum_{i,j}\alpha_i\beta_j KL(f_i||g_j). \tag{3.2}$$

Gornja granica data ocenom (3.2) (predložena u [7]) je gruba, naročito ako su Gausove komponente koje pripadaju različitim GMM-ovima jako udaljene jedna od druge. KL-divergence $KL(f_i||g_j)$ između odgovarajućih Gausovih komponenti postoje u zatvorenoj formi

$$KL(f_i||g_j) = \ln\frac{|\Sigma_{f_i}|}{|\Sigma_{g_j}|} + Tr\left[\Sigma_{g_j}^{-1}\Sigma_{f_i}\right] + (\mu_{f_i}-\mu_{g_j})^T\Sigma_{g_j}^{-1}(\mu_{f_i}-\mu_{g_j}) - d, \tag{3.3}$$

gde je d dimenzija obeležja. Gornja granica (3.2) indukuje težinsku aproksimaciju datu sa

$$KL_{WE}(f||g) \approx \sum_{i,j} \alpha_i \beta_j KL(f_i||g_j). \tag{3.4}$$

U [11], [9] i [12], autori su predložili razne aproksimacije KL divergenci između GMM-ova i efikasno ih primenili na razne realne probleme, kao što su prikupljanje slika *image retrieval*, identifikacija i verifikacija govornika. U [11], autori su predložili aproksimaciju zasnovanu na upoređivanju datu sa

$$KL_{MB}(f||g) \approx \sum_i \alpha_i \left[\min_j KL(f_i||g_j) + \log\left(\frac{\alpha_i}{\beta_j}\right) \right] \tag{3.5}$$

koja je zasnovanu na pretpostavci da element u sumi $g = \sum_j \beta_j g_j$ koji je najbliži f_i dominira u integralu $\int f_i \log g \quad dx$. Efikasnija aproksimacija motivisana sa (3.5) je data sa

$$KL_{MBS}(f||g) \approx \sum_i \alpha_i \min_j KL(f_i||g_j). \tag{3.6}$$

Metod baziran na upoređivanju daje dobru aproksimaciju KL-divergence između GMM-ova ako su komponente f i g uglavnom udaljene među sobom. Međutim, ako to nije slučaj , tj. ako postoji značajno preklapanje između komponenti f i g onda aproksimacija data sa (3.6) nije adekvatna. Da bi se to razrešilo autori predlažu aproksimaciju baziranu na *Unscented Transform* (videti [12]). Ta metoda računa statistike slučajnih promenljivih koja podleže nelinearnoj transformaciji (videti [12]). Kako važi

$$KL(f||g) = \int_{\mathbb{R}^d} f \log f \quad dx - \int_{\mathbb{R}^d} f \log g \quad dx,$$

primenjuje se *unscented transform* i aproksimira se integral

$$\int_{\mathbb{R}^d} f \log g \quad dx$$

kao

$$\int_{\mathbb{R}^d} f \log g \;\; dx \approx \frac{1}{2d} \sum_{i=1}^{n} \alpha_i \sum_{k=1}^{2d} \log g(x_{i,k}), \tag{3.7}$$
$$x_{i,k} = \mu_i + \left(\sqrt{d\Sigma_i}\right)_k, \;\; k = 1, \ldots, d,$$
$$x_{i,d+k} = \mu_i - \left(\sqrt{d\Sigma_i}\right)_k, \;\; k = 1, \ldots, d,$$

i slično za drugi integral. Ovde $\left(\sqrt{d\Sigma_i}\right)_k$ označava k-tu kolonu kvadratnog korena matrice $d\Sigma$ (videti [12]), dok $\int_{\mathbb{R}^d} \;\; dx$ označava da su u pitanju Lebegovi integrali na $\mathbb{R}^d$. Tako se $KL_{UC}(f||g)$ dobija na način opisan u (3.7).

Druga interesantna i efikasna mera sličnosti između GMM-ova predložena u [18] (videti i [9]), je data sa

$$KL_{VAR}(f||g) = \sum_i \alpha_i \frac{\sum_{\hat{i}} \alpha_{\hat{i}} e^{-KL(f_i||f_{\hat{i}})}}{\sum_j \beta_j e^{-KL(f_i||g_j)}} \tag{3.8}$$

U [27] autori predlažu meru sličnosti zasnovanu na geometrijskim osobinama, pokazujući da (μ_i, Σ_i) obrazuju Rimanovu površ ulegnutu u visoko dimenzionalni Euklidski prostor (videti [31]).

3.1.2 LPP redukcija dimenzionalnosti prostora obeležja

Dok PCA teži da očuva globalnu strukturu podataka, LPP teži da sačuva lokalnu strukturu podataka. Naime, LPP čuva više informacija o diskriminativnosti obeležja nego PCA, pretpostavljajući da su uzorci iz iste klase uglavnom bliže jedni drugima unutar klasa nego među klasama (u smislu geodezije na nisko-dimenzionalnoj površi na kojoj leže).

Poenta LPP analize je da posmatra tačke podataka $x_i \in \mathbb{R}^d$, $i = 1, \ldots, N$ iz originalnog visoko dimenzionalnog prostora, kao čvorove neorijentisanog grafa, sa težinama predstavljenim preko težinske matrice $[p_{ij}]_{N\times N}$, date sa

$$p_{ij} = \begin{cases} e^{-\|x_i - x_j\|^2/t} & \text{ako } x_i \in kNN(x_j) \text{ or } x_j \in kNN(x_i), \\ 0 & \text{inače} \end{cases} \tag{3.9}$$

gde je $kNN(x_j)$ okolina koja sadrži k najbližih euklidskih suseda do x_j, dok je $t > 0$ faktor skaliranja.

U jednodimenzionalnom slučaju koji će biti izložen radi jednostavnosti, cilj je da se minimizuje sledeća ciljna funkcija

$$f(w) = \sum_{i,j} (y_i - y_j)^2 p_{ij}, \tag{3.10}$$

gde su $y_i = w^T x_i$ transformisani niže-dimenzionalni podaci. Tako se podstiče da y_i i y_j budu blizu jedan drugom u transformisanom prostoru, kao što su x_i i x_j blizu u originalnom prostoru.

Optimalno w dobija se (videti [16], [33]) rešavanjem uopštenog spektralnog problema (tj. problema sopstvenih vrednosti)(videti Prilog B)

$$XLX^T w = \lambda XDX^T w, \tag{3.11}$$

gde je $X = [x_1|\cdots|x_N]$ matrica podataka, D je dijagonalna matrica čije je dijagonala dobijena sumiranjem vrsta (kolona) matrice P date sa (3.9), tj. $d_{ii} = \sum_j p_{ij}$, dok je $L = D - P$ Laplasijan matrica pomenutog grafa.

U cilju transformisanja prostora u l-dimenzionalni prostor niže dimenzije, generalizacija jednodimenzionalnog slučaja je direktna, i sastoji se u uzimanju l karakterističnih vektora w_i koji odgovaraju l najvećih karakterističnih vrednosti prethodno pomenutog problema (3.11), i njihovim stavljanjem u vrste projektivne matrice W koja projektuje visoko-dimenzionalni prostor $\mathbb{R}^d$ na nisko-dimenzionalni euklidski prostor $\mathbb{R}^l$ ($l \ll d$).

3.2 GMM mere sličnosti zasnovane na projektovanju LPP tipa parametarskog prostora

Da bismo iskoristili mogućnost da parametri Gausovih komponenti GMM-ova aproksimativno leže na niže-dimenzionalnoj površi, uvodimo novu, mnogo efikasniju meru sličnosti između proizvoljnih GMM-ova (videti [23]).

3.2.1 Mera sličnosti između GMM-ova zasnovana na redukciji dimenzionalnosti parametarskog prostora

Mera sličnosti, koju uvodimo u [23], koristi LPP baziranu tehniku da bi obučila projektivnu (transfomacionu) matricu W, koja projektuje parametre proizvoljnih GMM-ova, tj. vektorizovane parametre pripadajućih Gausovih komponenti, u niže-dimenzionalni prostor parametara, dok u isto vreme čuva informaciju o lokalnom susedstvu koje postoji u konfiguracionom prostoru, tj. originalnom prostoru parametara. Plan je da izvedemo dve vrste takve mere: nesimetričnu i simetričnu. Nesimetrična koristi jednostranu KL-divergencu između pripadajućih Gausovih komponenti. Simetirčna koristi simetričnu KL-divergencu između pripadajućih Gausovih komponenti.

Metodologija se svodi na sledeće: Za sve date GMM-ove koji se koriste u nekom konkretnom problemu prepoznavanja, skupljamo sve pripadajuće Gausove komponente i posmatramo ih kao čvorove neorijentisanog težinskog grafa. Za svaki i-ti čvor grafa, parametri odgovarajućeg Gausijana, tj. njegov centroid μ_i i kovarijansa Σ_i su vektorizovani u jedan vektor, koji pripada visoko-dimenzionalnom euklidskom prostoru dimenzije $d(d+1)/2+d$, koji zovemo konfiguracioni prostor parametara.

Za formiranja težina grafa p_{ij}, korisitimo određenu meru sličnosti između multivarijabilnih Gausijana i primenjujemo je na Gausijane $\mathcal{N}(\mu_i, \Sigma_i)$, $\mathcal{N}(\mu_j, \Sigma_j)$ koji odgovaraju i-tom i j-tom čvoru grafa, redom. Zatim, plan je da minimizacijom ciljne funkcije date sa (3.10)

sačuvamo osobinu lokalnosti. Naime, za one Gausijane koji su "blizu" jedan drugom u smislu pomenute mere sličnosti, želimo da su odgovarajući transformisani nisko-dimenzionalni euklidski vektori takođe "blizu" jedan drugom u smislu euklidske norme u transformisanom prostoru.

Osobina lokalnosti u euklidskom prostoru obeležja koju obezbeđuje LPP, sada treba da važi i u slučaju koji je ovde predložen , ali u slučaju nenegativno definitnih predstavnika iz $SPD^{+}_{(d+1)}$ kome pripadaju parametri Gausovih komponenti. Tada zajedno sa težinama α_i koje odgovaraju i-toj komponenti nekog razmatranog GMM-a, ti transformisani nisko-dimenzionalni euklidski vektori potpuno predstavljaju dati GMM u transformisanom prostoru parametara.

Za dobijanje težina grafa p_{ij} koje se koriste u funkciji cilja (3.10), koristi se činjenica da se prostor multivarijabilnih Gausijana koji čine Rimanovu površ može umetnuti u konus pozitivno definitnih matrica (SPD) (videti [27] i [31]).

Tako bi se distanca između dva GMM-a računala ne koristeći aproksimacije KL divergenci pripadajućih Gausovih komponenti u originalnom konfiguracionom prostoru parametara, već koristeći određene operatore agregacije distance nad objektima, koji reprezentuju pojedinačne date GMM-ove u transformisanom prostoru parametara, tipa $F = (v_1, \ldots, v_m, \alpha_1, \ldots, \alpha_m)$.

3.2.2 Konstrukcija težinskog grafa i matrica projekcije LPP tipa

Za konstruisanje težinskog grafa težine p_{ij} se moraju ubaciti u funkciju cilja (3.10). Koristimo činjenicu da je prostor multivarijabilnih Gausijana Rimanova površ i da se može ulegnuti u konus pozitivno definitnih matrica (videti [27] i [31]). Ustvari, d dimenzionalni multivarijabilni Gausijan $\mathcal{N}(\mu, \Sigma)$, se može ulegnuti u $\mathcal{SPD}^+{}_{(d+1)}$, što je konus u $n = d(d+1)/2 + d$ euklidskom dimenzionalnom prostoru i takođe Rimanova površ. Ovo ćemo uraditi na sledeći način [27], [31]:

$$\mathcal{N}(\mu, \Sigma) \hookrightarrow P = |\Sigma|^{-\frac{1}{d+1}} \begin{bmatrix} \Sigma + \mu\mu^T & \mu \\ \mu^T & 1 \end{bmatrix}, \tag{3.12}$$

gde $|\Sigma| > 0$ označava determinantu kovarijansne matrice komponente Gausijana $\mathcal{N}(\mu, \Sigma)$.[1]

Prema tome, informacije o određenoj komponenti Gausijana $\mathcal{N}(\mu, \Sigma)$ date GMM su sadržani u jednoj pozitivno definitnoj matrici P, elementa od $\mathcal{SPD}^+{}_{(d+1)}$. Sada konstruišemo težinski graf težina p_{ij} uključujući unutrašnju *ground distance*, tj. mere sličnosti između P_i and P_j koji su u konusu $\mathcal{SPD}^+{}_{(d+1)}$, u izraz (3.9), umesto euklidske distance $\|\cdot\|$ primenjene na visoko-dimenzionalni euklidski vektor.

Ground distances koje koristimo između P_i i P_j odgovara i-tom i j-tom čvoru grafa na sledeći način: Prva je jednostrana, tj. u pitanju je klasična KL-divergenca D između centriranih multivarijabilnih Gausijana $\mathcal{N}(0, P_i)$ i $\mathcal{N}(0, P_j)$ koja je data u zatvorenoj formi:

$$\begin{aligned} D_{ord}(P_i, P_j) &= KL\left(\mathcal{N}(0, P_i), \mathcal{N}(0, P_j)\right) \\ &= \frac{1}{2} \ln \frac{|P_j|}{|P_i|} + \frac{1}{2} Tr\left(P_i^{-1} P_j\right), \end{aligned} \tag{3.13}$$

gde je $Tr(\cdot)$ trag matrice. Druga je simetrična KL-divergenca D_{sym} koja je data u zatvorenoj formi

[1] Za detaljniju matematičku teoriju koja leži iza uleganja (3.12) pogledati [31].

$$D_{sym}(P_i, P_j) = KL\left(\mathcal{N}(0, P_i), \mathcal{N}(0, P_j)\right) + KL\left(\mathcal{N}(0, P_j), \mathcal{N}(0, P_i)\right)$$
$$= \frac{1}{2}\left(Tr\left(P_i^{-1}P_j\right) + Tr\left(P_j^{-1}P_i\right)\right). \tag{3.14}$$

Kao što je ranije objašnjeno, koristeći (3.14) dobijamo željene težine

$$p_{ij} = \begin{cases} e^{-D^2(P_i,P_j)/t}, & \text{ako je } P_i \in kNN(P_j) \text{ ili } P_j \in kNN(P_i), \\ 0, & \text{inače,} \end{cases} \tag{3.15}$$

gde je $kNN(P_j)$ okolina koja sadrži k najbližih suseda u P_j, u odnosu na *ground distance* D date sa (3.13) ili (3.14). Kao u slučaju klasičnog LPP baziranog na euklidskoj distanci (pri budućoj primeni) svrha parametra t je da skalira distance u standardnom odnosu. Iako ima smisla da se primene različiti agregatori unutar matrice $\left[D^2(f_i, g_j)\right]_{m\times n}$ u cilju da se dobije pozitivan skalar t, korisitimo centroidu (eng. *mean*), tj.

$$t = \frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m} D^2(f_i, g_j). \tag{3.16}$$

Sada, minimizovanjem ciljne funkcije (3.10), dobijamo projektivnu matricu W, koja projektuje originalni euklidski visoko-dimenzionalni prostor parametara $\mathbb{R}^n$, $n = d(d+1)/2 + d$, koji sadrži vektorizovane parove (μ_i, Σ_i), $i = 1, \ldots, N$, u niže-dimanzionalni euklidski prostor parametara dimenzije $l << n$. Napomenimo, da će l biti dato *a priori*, zavisno od date primene. Ustvari, ono zavisi od dimanzionalnosti niže-dimenzionalne potpovrši ulegnute u $\mathcal{SPD}^+{}_{(d+1)}$, gde očekujemo da uzorci podataka leže blizu, u datoj primeni.

Napomenimo da W (slično kao kod klasičnog LPP-a) je, u stvari, dobijeno kao $l \times n$ matrica, koja sadrži karateristične vektore koji odgovaraju najvećim l karateističnim vrednostima dobijenim rešavanjem spektralnog problema datog sa (3.11), pri čemu p_{ij} je dato sa (3.15). Takođe, l je izabrano tako da bude malo, čak iako prethodne pretpostavke nisu potpuno zadovoljene, pri čemu je niska računska složenost prioritet. Prema tome, mi možemo dobiti bolji odnos (eng. *trade-off*) između tačnosti prepoznavanja i računske složenosti u određenom zadataku prepoznavanja.

Iako su eksperimenti rađeni sa klasičnom KL-divergencom definisanom sa 3.13, umesto nje u fomulu za težine date sa 3.15 može se umetnuti i Rimanova metrika na $\mathcal{SPD}^+{}_{(d+1)}$ date sa 2.29. Tako da za dva predstavnika $p_i, p_j \in \mathcal{SPD}^+{}_{(d+1)}$ koji su po, na taj način, definisanoj geodezijskoj liniji bliski njihove slike pri transformaciji W koja je opisna u ovoj sekciji takođe bliske i u euklidskom transformisanom prostoru. Time bi održavanje osobina lokalnosti bilo potpuno zadovoljeno u formalnom smislu. Ovako je osobina loklanosti zadovoljena u smislu KL-divergence.

3.2.3 Konstrukcija GMM-LPP mere sličnosti

Konstruišemo meru sličnosti između dva data GMM-a zasnovano jedino na informaciji dobijenoj iz transformisanog prostora (videti [23]): svaka komponenta Gausijana $\mathcal{N}(\mu, \Sigma)$, od bilo kojeg GMM-a f koji nas interesuje, je sada prikazana preko jednog niskodimenzionalnog vektora. Prema tome, određena m komponenta GMM data sa $f = \sum_{i=1}^{m} \alpha_i f_i$, sa $f_i = \mathcal{N}(\mu_i, \Sigma_i)$ je prikazana $2m$-torkom $(v_1, \ldots, v_m, \alpha_1, \ldots, \alpha_m)$, gde je $v_i \in \mathbb{R}^l$. S' prethodnom reprezentacijom, koristimo meru sličnosti između dva GMM-a datu sa $f = \sum_{i=1}^{m} \alpha_i f_i$ i $g = \sum_{i=1}^{n} \beta_j g_j$, jednostavnim upoređivanjem odgovarajućih reprezenata $F = (v_1, \ldots, v_m, \alpha_1, \ldots, \alpha_m)$, $G = (u_1, \ldots, u_n, \beta_1, \ldots, \beta_n)$, u transformisanom prostoru, kao težinski euklidski vektori. Srednja vrednost sakupljenih informacija iz F i G određuje formu mere sličnosti. Npr., prirodno je da mera bude funkcija ponderisna euklidskim distancama $\|v_i - u_j\|_2$ između reprezenata v_i i u_j, koji odgovaraju f_i i g_j redom.

Postoji više načina da ovo uzmemo u obzir. To znači da postoji više načina da dobijemo jedan pozitivan skalar koji predstavlja "meru" između F i G, i u najopštijem slučaju možemo koristiti proizvoljnu *fuzzy* uniju ili presek, date sa npr. različitim operatorima agregacije (mnogi od njih su prikazani u [22]). U ovoj tezi koristimo dva tipa agregacionih operatora nad $\|v_i - u_j\|_2$.

Prvi je max-min agregacioni operator koji je primenjen na prethodno pomenute reprezente kao što sledi:

$$\begin{aligned} p_i &= \min\{\beta_j \|v_i - u_j\|_2 \mid j = 1, \ldots, n\}, \\ a &= \max\{\alpha_i p_i \mid i = 1, \ldots, m\}, \\ q_j &= \min\{\alpha_i \|v_i - u_j\|_2 \mid i = 1, \ldots, m\}, \\ b &= \max\{\beta_j q_j \mid j = 1, \ldots, n\}, \\ D_1(F, G) &= \frac{1}{2}(a + b). \end{aligned} \tag{3.17}$$

Drugi je maksimum težinskih suma

$$a = \max\{\alpha_i \sum_{j=1}^{n} \beta_j \|v_i - u_j\|_2 \ | i = 1, \ldots, m\} \tag{3.18}$$

$$b = \max\{\beta_j \sum_{i=1}^{m} \alpha_i \|v_i - u_j\|_2 \ | j = 1, \ldots, n\}$$

$$D_2(F, G) = \frac{1}{2}(a + b).$$

Napomenimo da su i D_1 i D_2 date sa (3.17) i (3.18) simetrične funkcije po F i G. Izbor i osobina distance D (ne-simetrična D_{ord} ili simetrična D_{sym}), koja se koristi za formiranje težinske matrice $[p_{ij}]$ date sa (3.15) ne utiče na simetričnost.

Umesto prethodno pomenutog fazi agregacionog operatora može se korisititi i *Earth Mover's Distance* metoda koja je prikazana u [35], ali to ostavljamo za neko od budućih istraživanja. Napomenimo da KL-divergenca zadovoljava sledeće tri osobine za proizvoljne raspodele f i g:

- $KL(f||f) = 0$,
- $D(f||g) = 0$ ako i samo ako $f = g$ (skoro svuda),
- $D(f||g) \geq 0$.

Štaviše, kako je KL-divergenca simetrična važi i $D(f||g) = D(g||f)$.

Iz prethodnog, jasno je da mere D_1 i D_2 između GMM-ova koje predlažemo, date sa (3.17) i (3.18) redom, zadovoljavaju prvu i treću osobinu, za proizvoljne GMM-ove f i g, u oba slučaja i za nesimetričan (kada koristimo nesimetričnu KL-divergencu) i za simetričan slučaj (kada koristimo simetričnu KL-divergencu). Druga osobina nije zadovoljena ni u jednom slučaju. Dodatno se dobija da je i za D_1 i za D_2 zadovoljena osobina simetričnosti.

3.2.4 Nadgledana GMM-LPP mera sličnosti

Nadgledana verzija GMM-LPP mere je zasnovana na primeni nadgledane verzije LPP.

Naime, uključujemo informacije o labelama, u odnosu na pripadanje komponentama Gausijana nekog GMM-a koji se korisiti za određenu klasu, u izraz težine grafa p_{ij}. Tako da umesto da se težine grafa formiraju kao u (3.15), sada ih formiramo na sledeći način

$$p_{ij} = \begin{cases} e^{-D^2(P_i,P_j)/t}, & \text{ako } P_i,\ P_j \text{ pripadaju istoj klasi,} \\ 0, & \text{inače.} \end{cases} \tag{3.19}$$

gde je D dato sa 3.13 ili 3.14.

Svi ostali elementi aplikacije LPP pri dobijanju težinske matrice W uvedeni su i objašnjene u Sekciji 3.1.2.

Konstruisanje mere sličnosti između dva data GMM-a zasnovane na informacijama dobijenim iz transformisanog prostora koje je prikazano u Sekciji 3.2.3 je isto i za slučaj nenadgledane verzije GMM-LPP mere.

Korišćenjem (3.19), moguće je dobiti veću diskriminativost u etapi prepoznavanja. Demonstriraćemo konstrukciju na realnim podacima u zadatku prepoznavanja teksture u Sekciji 4.0.7.

3.2.5 Računska složenost

Računska složenost za mere sličnosti (3.4)-(3.7) (videti [9]) predstavlja broj računskih operacija potrebnih za poređenje dva GMM-a. Ona je ekvivalentna i grubo data sa $O(m^2d^2)$ (u slučaju *full covariance*), gde pretpostavljamo da GMM-ovi f and g imaju isti broj komponenti, tj. $n = m$, pri čemu je d dimenzija originalnog prostora obeležja.

Naravno, Monte-Carlo aproksimacija KL_{MC}, data sa (3.1) je računski najsloženija i njena računska složenost je data sa $O(Nmd^2)$ (za slučaj *full covariance*), gde je sa N dat broj uzoraka koji se koriste u Monte-Carlo metodi i gde pretpostavljamo da GMM-ovi f i g imaju isti broj komponenti, tj. $n = m$. Broj N mora biti velik, tj. $N >> m$, da bi se dobila efikasnija aproksimacija.

Mere sličnosti između GMM-ova sa kojima se poredimo su zasnovane na EMD. Za slučaj $m = n$, dobija se da je računska složenost $O(k_{iter}m^4d^3)$ i $O(8m^5d^3)$, za EMD-KL i SR-EMD-M, redom (videti [27]). k_{iter} predstavlja broj iteracija u numeričkom algoritmu koji se koristi za nalaženje rešenja problema optimizacije koji je uključen u izračunavanju EMD-KL. Napominjemo da LARS/Homotopy algoritam koji se koristi za nalaženje numeričkog rešenja problema optimizacije uključenog u SR-EMD-M, konvergira u ne više od $2m$ iteracija (videti [14]). Obično je $m << k_{iter}$ (videti [27]).

Za meru sličnosti (kako simetričnu, tako i nesimetričnu), koja će biti predložena u ovom istraživanju, računska složenost je data sa $O(m^2l)$ (za slučaj *full covariance*), gde je $l << d(d+1)/2+d$. Podsetimo se, l je dimenzija transformisanog niže-dimenzionalnog prostora. Jasno je da je predložena mera sličnosti računski mnogo efikasnija od bilo koje ranije predložene mere koje su opisane u odeljku 3.1.1.

Treba napomenuti, da računska efikasnost za bilo koju meru sličnosti je data (i definisana) za etapu testiranja, a ne za etapu učenja.

Glava 4

Eksperimentalni rezultati

U ovoj glavi prikazani su eksperimentalni rezultati u kojima poredimo GMM-LPP meru sličnosti (simetrični slučaj) koja je predložena u tezi sa *state-of-the-art* merama sličnosti, koje su prikazane u prethodnim poglavljima.

Eksperimenti su izvedeni na specijalno konstruisanim sintetičkim podacima, kao i na realnim podacima, u problemu prepoznavanja teksture. Kao mere za poređenje sa predloženom merom, korišćene su sledeće GMM mere sličnosti: KL_{WE}, KL_{MB} i KL_{VAR} , definisane sa (3.4), (3.5) i (3.8), redom.

U slučaju specijalno konstruisanih sintetičkih podataka, dobijeno je da predložena GMM-LPP mera ostvaruje značajno manju računsku složenost u svim izvedenim eksperimentima. U isto vreme, ostvarena je i veća tačnost prepoznavanja u poređenju sa svim merama sličnosti koje su koriš ćene.

Na realnim podacima, predložena GMM-LPP mera sličnosti ostvarila je značajno bolji *trade-off* između računske složenosti i tačnosti prepoznavanja, u poređenju sa svim drugim korišćenim GMM merama sličnosti u skoro svim esksperimentima.

Napomenimo da su u eksperimentima na sintetičkim podacima testirane dve različite GMM-LPP mere sličnosti, koje obe koriste simetričnu D_{sym} meru sličnosti između Gausovih komponenti, datu sa (3.14). Jedna od njih je označena sa $\mathrm{GMM\text{-}LPP}_1$ i odgovara meri u

kojoj koristimo D_1 definisanu sa (3.17). Druga je označena sa GMM-LPP$_2$ i odgovara meri D_2 definisanoj sa (3.18).

U eksperimentima na realnim podacima, pošto su dobijeni skoro identični rezultati korišćenjem D_1 i D_2, ali značajno različiti kada se koristila verzija algoritma sa supervizorom u odnosu na onu bez supervizora, označena je verzija GMM-LPP mere bez supervizora sa GMM-LPP$_1$, dok je ona sa supervizorom označena sa GMM-LPP$_2$. Obe koriste simetričnu D_{sym} distancu.

4.0.6 Eksperimenti na sintetičkim podacima

U eksperimentima na sintetičkim podacima, izvedene su dve klase eksperimenata u dva različita scenarija. U prvom scenariju, forsirano je da parametri Gausijana leže na niskodimenzionalnoj podmnogostrukosti ulegnutoj u konus $\mathcal{SPD}^+{}_{(d+1)}$, (koji leži u $\mathbb{R}^n$, $n = d(d+1)/2+d$ euklidskom prostoru), gde je $d\times d$ dimenzija kovarijansnih matrica. U eksperimentima su isprobane različite vrednosti d . Podmnogostrukosti koje su formirane su dimenzija $l = 1$ i $l = 2$. U eksperimentima prikazanim u tabelama 4.1-4.4 i 4.8-4.11 sve centroide Gausijana, koji su generisani sa pomenutih niskodimenzionalnih mnogostrukosti, su jednake nula vektorima. U tabelama 4.5-4.7 i 4.12-4.14 prikazani su eksperimenti gde su centroide Gausijana različite od nula vektora. U tim eksperimentima, centroide Gausijana su formirane kao d-dimenzioni euklidski vektori, gde je prvih $l_\mu = \lceil d/2 \rceil$ vrednosti odabirano po uniformnoj raspodeli kao $z = hx$, sa $x \sim \mathcal{U}([0, 1])$ i $h = 100$, u svim eksperimentima. Ostalih $d - l_\mu$ vrednosti držano je na nuli. U pomenutim eksperimentima, za prethodno opisane $l = 1$ i $l = 2$ slučajeve, rezultujuće podmnogostrukosti su, u stvari, direktni proizvodi podmnogostrukosti na kojima leže kovarijanse (dimenzije l) i podmnogostrukosti na kojima leže centroide (dimenzije l_μ). Konačne dimenzije tih podmnogostukosti su $l + l_\mu$.

U slučaju $l = 1$, uniformno se generišu pozitivno definitne matrice A_1, A_2 formata $d \times d$, na sledeći način: Neka je $l = 1$ (na isti način se radi i za slučaj $l = 2$). Uniformno se generiše matrica tako da su elementi nezavisno i podjednako raspoređeni, tj. sa raspodelom $\mathcal{U}([0, 1])$. Zatim se vrši simetrizacija. Dobija $A_1 = \tilde{A}_1 + \epsilon I$, gde je $\tilde{A}_1$ matrica dobijena matrice B_1^{sym} samo zamenom dijagonalnih elemenata matrice $B_1^{sym} = [b_{ij}]_{d\times d}$ sumom vandijagonalnih elemenata matrice B_1^{sym}, tj. $b_{ii} \leftarrow \sum_{\substack{j=1 \\ j\neq i}}^{d} b_{ij}$ (važi $b_{ij} > 0$). Kako je $\epsilon > 0$ u eksperimentima je birano $\epsilon = 0.00001$. Pošto je rezultujuća A_1 simetrična i dijagonalno dominantna matrica, ona je pozitivno definitna (videti [6]). Isto važi za matricu A_2. Zatim se formira $l = 1$ dimenzionalna mnogostrukost u formi paraboličke krive, u konusu $\mathcal{SPD}^+{}_{(d+1)}$ pozitivno definitnih matrica, kao

$$F(t) = at^2 \frac{A_2}{\|A_2\|} + bt \frac{A_1}{\|A_1\|} + c \in \mathcal{SPD}^+{}_{(d+1)}, \qquad (4.1)$$
$$t \in [r_1, r_2],\ r_1, r_2 \in \mathbb{R},\ r_1 < r_2,$$

gde važi $a, b, c \in \mathbb{R}_+ \cup \{0\}$, $a \neq 0$. Birano je $a = 1$ i bez umanjenja opštosti i $b, c = 0$, u svim eksperimentima. Zatim se uniformno uzima predefinisan N broj Gausovih komponenti, tako da leže na krivoj datoj sa (4.1), sa centroidama koje sadrže l_μ ne-nula vrednosti, na način prethodno opisan. Formira se skup od $\tilde{N}$ GMM-ova, predefinisane dužine K, uniformnim uzimanjem Gausovih komponenti iz prethodno dobijenog skupa, pri čemu su sve težine u smešama postavljene da budu $1/K$. Za tako određen skup GMM-ova, za svaki eksperiment, se vrši kros-validacija: uniformno se bira 10 procenata u test skup, ostavljajući ostalih 90 skupu za obuku, usrednjavajući konačan rezultat brojem kros-validacija, koji postavljamo na 10 u svim eksperimentima. Korišćene su različite vrednosti za K, r_1 i r_2 u eksperimentima: $K = 1$ i $K = 5$, kao i sledeći intervali za $[r_1, r_2]$: $[0, 5]$, $[-3, 5]$, $[-4, 5]$ i $[-5, 5]$.

Kao što se može videti iz tabela 4.1 do 4.7, gde je predstavljen slučaj $l = 1$, predložene GMM-LPP$_{1,2}$ mere sličnosti ostvarile su mnogo bolji *trade-off* između tačnosti prepoznavanja i računske složenosti, u poređenju sa svim ostalim korišćenim GMM merama sličnosti. Takođe, dobijena je veća tačnost prepoznavanja u slučaju predložene GMM-LPP$_{1,2}$ mere nego u slučaju svih ostalih korišćenih GMM mera, u skoro svim eksperimentima.

U isto vreme, računska složenost je mnogo puta manja u slučaju predloženih GMM-LPP$_{1,2}$ mera sličnosti. Naime, u eksperimentima predstavljenim u tabelama 4.1 do 4.4, procenjen broj nenula karakterističnih vrednosti, u smislu da su veće od praga $T = 10^{-6}$, bio je jednak $\hat{l} = 1$. Takođe, taj broj je pripadao intervalu $[1, 2]$, što pokazuje da uvedena nova procedura procenjuje dimenziju $l = 1$ ulegnute mnogostrukosti (4.1), sa potpunom tačnošću (tj. $\hat{l} = l$). U slučaju $l = 1$, izraz za GMM-LPP računsku složenost dat je sa $O(\tilde{N}K^2\hat{l}) = O(\tilde{N}K^2)$. Računska složenost koja je izražena je, u stvari, računska složenost pri jednom poređenju između GMM-ova.

Term $\tilde{N}$ predstavlja ukupan broj GMM-ova u trening skupu. U isto vreme, izraz za računsku složenost (definisanu na isti način), dat je aproksimativno sa $O(\tilde{N}K^2d^3)$ (videti 3.2.5). Zato je odnos računske složenosti između predložene metode i metoda sa kojima se poredimo, dat sa $\hat{l}/d^3 = 1/d^3$. Za d izabrano u tabelama 4.1 do 4.4, da bude $d \in \{10, 20, 30, 50\}$, jasno je da je računska složenost predložene GMM-LPP$_{1,2}$ značajno manja (nekoliko puta) od GMM mera sličnosti sa kojima se poredimo. U tabelama 4.5 do 4.7, gde je predstavljen slučaj $l_\mu = \lceil d/2 \rceil \neq 0$, algoritam iz sekcije 3.2.2, procenjuje sa potpunom tačnošću dimenziju ulegnute površi $l + l_\mu$, kao broj nenula karakterističnih vektora $\hat{l}$, tj. dobija se $\hat{l} = l + l_\mu$. Može se videti da obe GMM-LPP$_1$ i GMM-LPP$_2$ ostvaruju bolju tačnost prepoznavanja. U isto vreme, obe GMM-LPP$_1$ i GMM-LPP$_2$ ostvaruju znatno manju računsku složenost, koja je aproksimativno data sa $O(\tilde{N}K^2\hat{l}) = O\left(\left(\lceil \frac{d}{2} \rceil + 1\right)\tilde{N}K^2\right)$, u poređenju sa ostalim merama sličnosti, čija je složenost aproksimativno data sa $O(\tilde{N}K^2d^3)$. Naime, odnos između računske složenosti između predložene mere i mera sa kojima se poredimo dat je aproksimativno sa $\hat{l}/d^2 = \left(\lceil \frac{d}{2} \rceil + 1\right)/d^3$.

Tabela 4.1: Rezultati tačnosti prepoznavanja dobijeni na sintetičkim podacima: $l = 1$, $t \in [-3, 5]$, $N = 800$, $\tilde{N} = 200$, $l_\mu = 0$

tip	$K = 1$				$K = 5$			
mere	$d = 10$	$d = 20$	$d = 30$	$d = 50$	$d = 10$	$d = 20$	$d = 30$	$d = 50$
$GMM - LPP_1$	0.68	0.73	0.81	0.74	0.94	0.90	0.94	0.96
$GMM - LPP_2$	0.68	0.73	0.82	0.74	0.95	0.91	0.94	0.95
KL_{WE}	0.62	0.68	0.79	0.73	0.92	0.86	0.90	0.94
KL_{MB}	0.62	0.68	0.79	0.73	0.91	0.87	0.92	0.93
KL_{VAR}	0.62	0.68	0.79	0.73	0.91	0.87	0.92	0.91

Za slučaj $l = 2$, formira se $l = 2$ dimenzionalna mnogostrukost kao

$$F(t_1, t_2) = a\left(t_1^2\frac{A_2}{\|A_2\|} + t_2^2\frac{A_2}{\|A_2\|}\right) + b\left(t_1\frac{A_1}{\|A_1\|} + t_2\frac{A_1}{\|A_1\|}\right) + c \in \mathcal{SPD}^+{}_{(d+1)}, \quad t_1, t_2 \in [r_1, r_2],\ r_1, r_2 \in \mathbb{R},\ r_1 < r_2, \tag{4.2}$$

Tabela 4.2: Rezultati tačnosti prepoznavanja dobijeni na sintetičkim podacima: $l = 1$, $t \in [-4, 5]$, $N = 800$, $\tilde{N} = 200$, $l_\mu = 0$

tip	$K = 1$				$K = 5$			
mere	$d = 10$	$d = 20$	$d = 30$	$d = 50$	$d = 10$	$d = 20$	$d = 30$	$d = 50$
$GMM - LPP_1$	0.55	0.56	0.70	0.69	0.82	0.77	0.80	0.81
$GMM - LPP_2$	0.56	0.56	0.72	0.70	0.83	0.77	0.80	0.83
KL_{WE}	0.55	0.54	0.64	0.54	0.57	0.74	0.67	0.66
KL_{MB}	0.55	0.54	0.64	0.54	0.80	0.75	0.74	0.77
KL_{VAR}	0.55	0.54	0.64	0.54	0.85	0.74	0.75	0.78

Tabela 4.3: Rezultati tačnosti prepoznavanja dobijeni na sintetičkim podacima: $l = 1$, $t \in [0, 5]$, $N = 800$, $\tilde{N} = 200$, $l_\mu = 0$

tip	$K = 1$				$K = 5$			
mere	$d = 10$	$d = 20$	$d = 30$	$d = 50$	$d = 10$	$d = 20$	$d = 30$	$d = 50$
$GMM - LPP_1$	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
$GMM - LPP_2$	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
KL_{WE}	0.97	0.97	0.99	0.98	0.99	1.0	0.98	0.99
KL_{MB}	0.97	0.97	0.99	0.98	1.0	0.98	1.0	0.97
KL_{VAR}	0.97	0.97	0.99	0.98	1.0	1.0	0.98	0.97

Tabela 4.4: Rezultati tačnosti prepoznavanja dobijeni na sintetičkim podacima: $l = 1$, $t \in [-5, 5]$, $N = 800$, $\tilde{N} = 200$, $l_\mu = 0$

tip	$K = 1$				$K = 5$			
mere	$d = 10$	$d = 20$	$d = 30$	$d = 50$	$d = 10$	$d = 20$	$d = 30$	$d = 50$
$GMM - LPP_1$	0.32	0.54	0.46	0.51	0.50	0.46	0.42	0.53
$GMM - LPP_2$	0.31	0.53	0.46	0.49	0.48	0.46	0.40	0.51
KL_{WE}	0.33	0.52	0.45	0.48	0.50	0.43	0.46	0.55
KL_{MB}	0.33	0.52	0.45	0.48	0.51	0.42	0.47	0.57
KL_{VAR}	0.33	0.52	0.45	0.48	0.50	0.42	0.48	0.57

gde $a, b, c \in \mathbb{R}_+ \cup \{0\}$, $a \neq 0$. Ona je ulegnuta u $\mathbb{R}^n$, gde kao i u slučaju $l = 1$, biramo $a = 1$ i zbog jednostavnosti, bez umanjenja opštosti, biramo $b, c = 0$ u svim eksperimentima. Takođe, pozitivno definitne matrice A_1 i A_2, generisane su na isti način kao i u slučaju $l = 1$. GMM-ovi se formiraju na isti način kao u slučaju $l = 1$ i sa istim vrednostima za K, r_1 i r_2. Isti zaključci mogu biti doneseni

Tabela 4.5: Rezultati tačnosti prepoznavanja dobijeni na sintetičkim podacima: $l = 1$, $t \in [-3, 5]$, $N = 800$, $\tilde{N} = 200$, $l_\mu = \lceil d/2 \rceil$

tip	$K = 1$				$K = 5$			
mere	$d = 10$	$d = 20$	$d = 30$	$d = 50$	$d = 10$	$d = 20$	$d = 30$	$d = 50$
$GMM - LPP_1$	0.82	0.78	0.78	0.72	0.72	0.75	0.75	0.74
$GMM - LPP_2$	0.82	0.76	0.75	0.72	0.71	0.74	0.75	0.73
KL_{WE}	0.62	0.63	0.68	0.58	0.52	0.52	0.55	0.54
KL_{MB}	0.62	0.63	0.68	0.58	0.53	0.53	0.55	0.53
KL_{VAR}	0.62	0.63	0.68	0.58	0.53	0.54	0.57	0.54

Tabela 4.6: Rezultati tačnosti prepoznavanja dobijeni na sintetičkim podacima: $l = 1$, $t \in [0, 5]$, $N = 800$, $\tilde{N} = 200$, $l_\mu = \lceil d/2 \rceil$

tip	$K = 1$				$K = 5$			
mere	$d = 10$	$d = 20$	$d = 30$	$d = 50$	$d = 10$	$d = 20$	$d = 30$	$d = 50$
$GMM - LPP_1$	0.80	1.0	1.0	0.99	0.86	0.99	0.99	1.0
$GMM - LPP_2$	0.79	1.0	1.0	0.98	0.86	0.98	0.98	1.0
KL_{WE}	0.72	0.80	0.72	0.82	0.57	0.65	0.61	0.60
KL_{MB}	0.72	0.80	0.72	0.82	0.57	0.66	0.61	0.60
KL_{VAR}	0.72	0.80	0.72	0.82	0.58	0.66	0.60	0.59

Tabela 4.7: Rezultati tačnosti prepoznavanja dobijeni na sintetičkim podacima: $l = 1$, $t \in [-4, 5]$, $N = 800$, $\tilde{N} = 200$, $l_\mu = \lceil d/2 \rceil$

tip	$K = 1$				$K = 5$			
mere	$d = 10$	$d = 20$	$d = 30$	$d = 50$	$d = 10$	$d = 20$	$d = 30$	$d = 50$
$GMM - LPP_1$	0.62	0.59	0.67	0.63	0.61	0.69	0.69	0.64
$GMM - LPP_2$	0.61	0.59	0.65	0.62	0.59	0.67	0.68	0.62
KL_{WE}	0.54	0.65	0.66	0.54	0.48	0.53	0.50	0.50
KL_{MB}	0.54	0.65	0.66	0.54	0.49	0.53	0.52	0.51
KL_{VAR}	0.54	0.65	0.66	0.54	0.49	0.54	0.53	0.52

za $l = 2$ kao i za slučaj $l = 1$. Naime, kako se može videti iz tabela 4.8-4.14, gde je predstavljen slučaj $l = 2$, predložene GMM-$\text{LPP}_{1,2}$ mere sličnosti ostvaruju mnogo bolji *trade-off* između tačnosti prepoznavanja i računske složenosti, u poređenju sa ostalim korišćenim GMM merama sličnosti. U svim eksperimentima koji su prezentovani, u tabelama 4.8-4.14, procenjen broj nenula karakterističnih vrednosti,

u smislu da su iznad $T = 10^{-6}$, sa $\hat{l} = 2$, pri čemu su vrednosti u intervalu $[0.5, 1]$. To pokazuje da predloženi algoritam procenjuje stvarnu dimenziju $l = 2$ ulegnute podmnogostrukosti sa potpunom tačnošću, tj. $\hat{l} = l$. U slučaju $l = 2$, izraz za računsku složenost GMM-LPP aproksimativno dat sa $O(\tilde{N}K^2\hat{l}) = O(2\tilde{N}K^2)$, dok je za ostale algoritme sa kojima se poredimo dat sa $O(\tilde{N}K^2d^3)$ (pogledati 3.2.5). Takođe, odnos između računske složenosti uvedene GMM-LPP i ostalih metoda je ugrubo dat sa $\hat{l}/d^3 = 2/d^3$. U tabelama 4.8-4.14 je izabrano da $d \in \{10, 20, 30, 50\}$. Vidi se da je računska složenost mnogo manja za predloženi GMM-LPP$_{1,2}$ u poređenju sa računskom složenošću GMM mera sličnosti sa kojima se poredimo. Za tabele 4.12-4.14, gde je prezentovan slučaj $l_\mu = \lceil d/2 \rceil \neq 0$, mogu da se izvedu slični zaključci kao i za slučaj $l = 1$. Naime, GMM-LPP$_{1,2}$ ostvarila je mnogo bolju tačnost prepoznavanja u poređenju sa ostalim merama. U isto vreme, računska složenost od GMM-LPP$_{1,2}$ zadata grubo sa $O(\tilde{N}K^2\hat{l}) = O\left((2 + \lceil \frac{d}{2} \rceil)\tilde{N}K^2\right)$, mnogo je manja od računske složenosti ostalih metoda, date grubo sa $O(\tilde{N}K^2d^3)$. Odnos računske složenosti predložene mere i mera sa kojima se poredimo data je grubo sa $\hat{l}/d^3 = \left(\lceil \frac{d}{2} \rceil + 2\right)/d^3$.

Kao što se može videti iz prethodno prikazanih tabela, jedini slučajevi gde predložene GMM-LPP$_{1,2}$ mere ne ostvaruju veću tačnost prepoznavanja u poređenju sa ostalim GMM merama, su oni koji odgovaraju intervalu $[r_1, r_2] = [-5, 5]$. U tim slučajevima, transformacija W koja je procenjena kao što je to objašnjeno u sekcijama 3.1.2 i 3.2.2, nije bila u stanju da na dovoljno zadovoljavajući način "uhvati" podatke koji pripadaju ciljnoj mnogostrukosti. U tim slučajevima podaci su postavljeni simetrično na i oko temena krive date sa (4.1), ili površi date sa (4.2). Međutim, u tim slučajevima sve testirane mere postigle su loš rezultat. U svim drugim slučajevima u prethodno prikazanim tabelama, predložena GMM-LPP ostvaruje najveću tačnost prepoznavanja. U tim slučajevima osobina očuvanja lokalnosti GMM-LPP metode dolazi do izražaja, ostvarujući najveću distancu (D_1 i D_2 date sa (3.17) i (3.18)) između GMM-ova koji pripadaju različitim klasama i u isto vreme manju distancu između GMM-ova koji pripadaju istim klasama.

Tabela 4.8: Rezultati tačnosti prepoznavanja dobijeni na sintetičkim podacima: $l = 2$, $t_1, t_2 \in [-3, 5]$, $N = 800$, $\tilde{N} = 200$, $l_\mu = 0$

tip	$K = 1$				$K = 5$			
mere	$d = 10$	$d = 20$	$d = 30$	$d = 50$	$d = 10$	$d = 20$	$d = 30$	$d = 50$
$GMM - LPP_1$	0.79	0.84	0.78	0.83	1.0	0.97	0.99	1.0
$GMM - LPP_2$	0.78	0.82	0.78	0.83	0.98	0.97	0.99	0.99
KL_{WE}	0.78	0.76	0.75	0.95	0.97	0.94	0.94	0.93
KL_{MB}	0.78	0.76	0.75	0.83	0.94	0.97	0.95	0.94
KL_{VAR}	0.78	0.76	0.75	0.83	0.95	0.97	0.95	0.96

Tabela 4.9: Rezultati tačnosti prepoznavanja dobijeni na sintetičkim podacima: $l = 2$, $t_1, t_2 \in [-4, 5]$, $N = 800$, $\tilde{N} = 200$, $l_\mu = 0$

tip	$K = 1$				$K = 5$			
mere	$d = 10$	$d = 20$	$d = 30$	$d = 50$	$d = 10$	$d = 20$	$d = 30$	$d = 50$
$GMM - LPP_1$	0.61	0.58	0.64	0.66	0.82	0.78	0.71	0.74
$GMM - LPP_2$	0.59	0.56	0.64	0.65	0.80	0.78	0.71	0.73
KL_{WE}	0.54	0.57	0.56	0.64	0.79	0.69	0.74	0.69
KL_{MB}	0.54	0.57	0.56	0.64	0.78	0.70	0.73	0.69
KL_{VAR}	0.54	0.57	0.56	0.64	0.79	0.71	0.74	0.69

Tabela 4.10: Rezultati tačnosti prepoznavanja dobijeni na sintetičkim podacima: $l = 2$, $t_1, t_2 \in [0, 5]$, $N = 800$, $\tilde{N} = 200$, $l_\mu = 0$

tip	$K = 1$				$K = 5$			
mere	$d = 10$	$d = 20$	$d = 30$	$d = 50$	$d = 10$	$d = 20$	$d = 30$	$d = 50$
$GMM - LPP_1$	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
$GMM - LPP_2$	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
KL_{WE}	1.0	0.98	0.97	0.99	0.98	1.0	1.0	1.0
KL_{MB}	1.0	0.98	0.97	0.99	0.97	0.99	1.0	1.0
KL_{VAR}	1.0	0.98	0.97	0.99	0.99	0.98	1.0	1.0

U drugom scenariju, odabira se N pozitivno definitnih matrica A_i, gde je svaka A_i formirano na isti način kao i A_1 u prethodnom tekstu. Tako dobijamo pozitivno definitne matrice, uniformno raspodeljene u konusu $\mathcal{SPD}^+{}_{(d+1)}$, tj. takve da ne leže na nekoj niskodimenzionoj mnogostrukosti ulegnutoj u pomenuti konus. Formira se skup Gausijana, iz prethodno pomenutog skupa pozitivno definitnih matrica, gde

Tabela 4.11: Rezultati tačnosti prepoznavanja dobijeni na sintetičkim podacima: $l = 2$, $t_1, t_2 \in [-5, 5]$, $N = 800$, $\tilde{N} = 200$, $l_\mu = 0$

tip	$K = 1$				$K = 5$			
mere	$d = 10$	$d = 20$	$d = 30$	$d = 50$	$d = 10$	$d = 20$	$d = 30$	$d = 50$
$GMM - LPP_1$	0.51	0.47	0.44	0.41	0.50	0.60	0.54	0.55
$GMM - LPP_2$	0.51	0.47	0.44	0.41	0.50	0.60	0.54	0.55
KL_{WE}	0.46	0.48	0.34	0.43	0.61	0.54	0.54	0.55
KL_{MB}	0.46	0.48	0.34	0.43	0.59	0.53	0.54	0.53
KL_{VAR}	0.46	0.48	0.34	0.43	0.58	0.54	0.54	0.53

Tabela 4.12: Rezultati tačnosti prepoznavanja dobijeni na sintetičkim podacima: $l = 2$, $t_1, t_2 \in [-3, 5]$, $N = 800$, $\tilde{N} = 200$, $l_\mu = \lceil d/2 \rceil$

tip	$K = 1$				$K = 5$			
mere	$d = 10$	$d = 20$	$d = 30$	$d = 50$	$d = 10$	$d = 20$	$d = 30$	$d = 50$
$GMM - LPP_1$	0.65	0.85	0.87	0.82	0.85	0.91	0.98	1.0
$GMM - LPP_2$	0.64	0.85	0.86	0.83	0.84	0.91	0.98	0.98
KL_{WE}	0.68	0.70	0.63	0.73	0.52	0.53	0.52	0.63
KL_{MB}	0.68	0.70	0.63	0.73	0.50	0.51	0.54	0.63
KL_{VAR}	0.68	0.70	0.63	0.73	0.52	0.53	0.54	0.62

Tabela 4.13: Rezultati tačnosti prepoznavanja dobijeni na sintetičkim podacima: $l = 2$, $t_1, t_2 \in [0, 5]$, $N = 800$, $\tilde{N} = 200$, $l_\mu = \lceil d/2 \rceil$

tip	$K = 1$				$K = 5$			
mere	$d = 10$	$d = 20$	$d = 30$	$d = 50$	$d = 10$	$d = 20$	$d = 30$	$d = 50$
$GMM - LPP_1$	1.0	1.0	0.98	1.0	1.0	1.0	1.0	1.0
$GMM - LPP_2$	1.0	1.0	0.98	1.0	1.0	1.0	1.0	1.0
KL_{WE}	0.77	0.70	0.79	0.84	0.51	0.50	0.55	0.80
KL_{MB}	0.77	0.70	0.79	0.84	0.51	0.52	0.57	0.81
KL_{VAR}	0.77	0.70	0.79	0.84	0.52	0.51	0.58	0.82

su (bez umanjenja opštosti) sve centroide postavljene na nula-vektore. Zatim se formira $\tilde{N}$ različitih GMM-ova dužine K, sa komponentama uzetim uniformno iz prethodno pomenutog skupa Gausijana. U eksperimentima je korišćeno da je $N = 800$, $\tilde{N} = 200$ i $K = 5$. Dobijeno je da se predložena GMM-LPP, sa stanovišta tačnosti prepoznavanja, kao i računske efikanosti, ponaša isto kao i ostale GMM

Tabela 4.14: Rezultati tačnosti prepoznavanja dobijeni na sintetičkim podacima: $l = 2$, $t_1, t_2 \in [-4, 5]$, $N = 800$, $\tilde{N} = 200$, $l_\mu = \lceil d/2 \rceil$

tip	$K = 1$				$K = 5$			
mere	$d = 10$	$d = 20$	$d = 30$	$d = 50$	$d = 10$	$d = 20$	$d = 30$	$d = 50$
$GMM - LPP_1$	0.58	0.64	0.64	0.63	0.62	0.77	0.79	0.78
$GMM - LPP_2$	0.57	0.63	0.64	0.61	0.61	0.77	0.78	0.76
KL_{WE}	0.58	0.62	0.52	0.62	0.60	0.52	0.53	0.55
KL_{MB}	0.58	0.62	0.52	0.62	0.60	0.54	0.54	0.56
KL_{VAR}	0.58	0.62	0.52	0.62	0.59	0.55	0.54	0.57

mere. Takođe, procenjeni broj značajnih karakterističnih vrednosti bio je jednak dimenziji punog prostora, tj. $l = n = d(d + 1)/2$. Time da smo potvrdili da ako podaci ne zadovoljavaju osobinu da leže na niskodimenzionalnoj mnogostrukosti ulegnutoj u konus $\mathcal{SPD}^+_{(d+1)}$, nema koristi od predložene metode.

4.0.7 Eksperimenti na realnim podacima

U ovoj sekciji procenjujemo mogućnosti predložene metode u odnosu na pomenute metode sa kojima se poredimo u zadatku prepoznavanja na realnim podacima. Pored metoda koje su bazirane na KL-divergenci KL_{WE}, KL_{MB} i KL_{VAR} koje su opisane u 3.1, poredimo predložene GMM-LPP mere sa još tri druge *state-of-the-art* mere koje su EMD tipa (videti 3.1). Jedna od tih mera je pomenuta u [41] i takođe prezentovana i iskorišćena kao reper u [27]. Tu meru zovemo EMD-KL mera. Druga je nenadgledana *sparse* mera zasnovana na EMD, tj. EMD mera zasnovana na pretpostavci *sparse* osobine. Ona je predložena u [27] i zovemo je SR-EMD-M mera. Treća je nadgledana *sparse* mera zasnovana na EMD takođe predložena u [27], koju zovemo SR-EMD-M-L mera.

U zadacima prepoznavanja teksture eksperimenti su vršeni na tri različite baze teksture. To su: UMD (videti [42]), koja sadrži 25 klasa, sa ukupno 1000 slika; CUReT (videti [8]), koja ima 61 klasu, sa ukupno 5612 slika; KTH-TIPS (videti [15]), koja ima 10 klasa, sa ukupno 8010 slika. Za sve eksperimente, kao obeležja, tj. deskriptore tekstura u eksperimentima, koriste se kovarijansni deskriptori koji se smatraju jednim od najefikasnijih u zadacima prepoznavanja teksture (videti [44], [36], [19], sekciju 2.5). Oni su formirani kao što sledi: Za određenu sliku teksture, vrsta obeležja je uzeta u obliku vektora čija je dimenzija $\tilde{d} = 5$ oblika $[I, |I_x|, |I_y|, |I_{xx}|, |I_{yy}|](x, y)$. Zatim se formiraju kovarijansni deskriptori određivanjem kovarijansne matrice $R \times R$ uzorka centriranog u (x, y) i vektorizovanjem njenog gornjeg trougla u vektor obaležja dimenzije $d = \tilde{d}(\tilde{d}+1)/2 = 15$. Dobijamo ih za svaki piksel sa kordinatom (x, y). U svim eksperimentima uzeto je da je $R = 30$. Zatim, za tu neku određenu sliku teksture, kovarijansni deskriptori vektora obeležja dobijeni kao što je prethodno objašnjeno, smeštaju se u jedan skup, i parametri GMM-ova se dobijaju korišćenjem EM procedure (videti npr. [40], sekciju 2.2.2).

U tabeli 4.15 su prikazani rezultati eksperimenata sa procesorskim CPU vremenom (u sekundama) za predložene GMM-LPP mere kao i za GMM mere sličnosti koje su ranije navedene (videti [23]). Za svaku posebnu meru, za koju se procenjuje procesorsko CPU vreme, potrebno je oceniti distancu između dva data GMM-a. Izvršeno je

usrednjavanje na 100 pokušaja. Trenutni GMM-ovi su obučavani iz primera teksture slika, slučajnim odabirom iz KTH-TIPS baze. Broj komponenti Gausijana K u GMM-ovima varira od $K = 5$ do $K = 20$, sa korakom 5. Može se videti da sve predložene GMM-LPP mere, u svim slučajevima, daju mnogo manje (i do više od 10 puta manje) procesorsko CPU vreme u poređenju sa ostalim metodama koje su prikazane. Ovi rezultati su potpuno kompatibilni sa računskim granicama dobijenim za predlozene mere i one sa kojima se porede, date u sekciji 3.2.5. Napominjemo da je rađeno na radnoj stanici koja ima 2.3GHz CPU i 6G RAM.

Tabela 4.15: Srednja vrednost procesorskog CPU vremena za predložene GMM-LPP mere, u poređenju sa osnovnim merama, kao funkcija od K broja GMM komponenti, kao i dimenzije redukovanog prostora l_{max} (merna jedinica: [ms])

K	5			10			15			20		
KL_{WE}	17.6			70.5			159.2			282.3		
KL_{MB}	14.7			80.1			187.3			323.4		
KL_{VAR}	32.9			128.0			297.5			528.3		
EMD-KL	49.3			1987			15102			61123		
l_{max}	30	60	100	30	60	100	30	60	100	30	60	100
GMM-LPP	1.9	4.1	6.5	7.7	15.4	25.6	17.6	35.2	58.6	31.2	62.4	103.8

Na grafikonima 4.16-4.18, je prikazana tačnost prepoznavanja za predložene GMM-LPP mere u poređenju sa osnovnim merama koje su ranije prikazane. Eksperimenti su izvršeni na UMD, CUReT i KTH-TIPS bazama tekstura, redom, gde su tačnosti za SR-EMD-M i SR-EMD-M-L uzete iz rada [27]. Mi koristimo iste parametre za eksperimente kao što je navedeno u [27]. Naime, za sve eksperimente stavljamo da je broj komponenti Gausijana u trening i/ili u skupu za testiranje fiksan i jednak sa $K = 5$. Za svaku klasu selektujemo slučajnim odabirom (po uniformnoj raspodeli) fiksan broj n slika teksture da bude u trenažnom skupu. Ostale slike idu u skup za testiranje. Za svaku bazu slika čiji su rezultati prezentovani, variramo N pomenuti broj slika teksture za trening. Za predložene GMM-LPP mere, treniramo projektivnu matricu W na celom skupu GMM-ova dobijenih iz primera slika teksture za trening (koristeći EM

proceduru). Ovo je urađeno kao što je prikazano u 3.2.2 i 3.1.2 za nenadgledanu GMM-LPP$_1$ meru, kao i za opisanu sa 3.2.4 i 3.1.2 nadgledanu GMM-LPP$_2$ meru. Napomenimo da dobijamo skoro slične rezultate na svim bazama u slučajevima težinske min-max mere date sa (3.17) i težinske sume date sa (3.18). Mi prikazujemo samo min-max slučaj. U fazi testiranja, svaku posebnu sliku teksture uniformno delimo na četiri pod-slike. Za svaku sliku procenjujemo GMM sa $K = 5$ komponenti Gausijana. Tako je svaka slika reprezentovana sa četiri GMM-a. Svaki od ova četiri GMM-a se upoređuje sa svim GMM-ovima iz trenažnog skupa i njegova labela je determinisana kNN algoritmom. Konačna labela klase je determinisana "glasanjem". Fiksiramo $k = 3$ u kNN proceduri za sve eksperimente. Konačni rezultati su zatim dobijeni usrednjavanjem od preko 20 pokušaja. U svim eksperimentima, broj svih značajnih karakterističnih vrednosti, tj. onih čija je vrednost veća od $T = 10^{-6}$, je ograničena na fiksan maksimalan broj l_{max}, i taj određeni maksimalan broj je, u stvari, dosegnut u svim pokušajima. Prema tome, za sve korišćene baze slika, variramo od $l_{max} = 30$ do $l_{max} = 100$, sa ciljem da analiziramo *trade-off* između tačnosti prepoznavanja i računske složenosti.

Za sve baze podataka koje koristimo (videti grafikone 4.16-4.18), se može videti da za slučaj $l_{max} = 30$ preciznost prepoznavanja za nenadgledano GMM-LPP$_1$ je malo niža u poređenju sa svim merama baziranim na KL i EMD-KL meri. Među tim u slučaju $l_{max} = 70$ i $l_{max} = 100$ tačnost prepoznavanja je na nivou mera baziranih na KL i EMD-KL mere. Štaviše, SR-EMD-M i SR-EMD-M-L daju najbolju tačnost prepoznavanja. Za nadgledano GMM-LPP$_2$, tačnost prepoznavanja, za slučaj $l_{max} = 70$ i $l_{max} = 100$ je veća u poređenju sa merama zasnovanim na KL kao i EMD-KL merom. Ona je na nivou sa SR-EMD-M merom, ali nižeg nivoa prepoznavanja nego SR-EMD-M-L mera, za sve baze podataka.

S' druge strane, računska složenost u velikoj meri ide u korist predloženim GMM-LPP merama u poređenju sa ostalim izabranim merama. Naime, kao što je to prikazano u Sekciji 3.2.5, dobijamo da je računska složenost (po jednom poređenju GMM-ova, tj. računanju distance) za sve predložene GMM-LPP mere približno $O(K^2 l_{max})$, gde je K broj GMM komponenti (veličina GMM) u određenom GMM-

u [1]. Kao što je prikazano u Sekciji (3.2.5), za sve algoritme sa kojima se poredimo zasnovane na KL, tj. KL_{WE}, KL_{MB} i KL_{VAR}, dobijeno je da je računska složenost približno $O(K^2d^3)$. Takođe (videti Sekciju 3.2.5), za algoritme sa kojima se poredimo bazirane na EMD, tj. EMD-KL i SR-EMD-M, dobijeno je da je računska složenost približno $O(8K^5d^3)$ i $O(k_{iter}K^4d^3)$, redom, pri čemu je $k_{iter} >> K$ (videti [27]). Prema tome, dobijeno je da je odnos između računske složenosti za sve predložene GMM-LPP metode i metode bazirane na KL približno l_{max}/d^3. Odnos između računske složenosti svih predloženih GMM-LPP i mera baziranih na EMD je $\frac{l_{max}}{k_{iter}K^2d^3}$ i $\frac{l_{max}}{8K^3d^3}$, redom. Iz prethodnog, jasno je da računska složenost u velikoj meri ide u korist predloženih GMM-LPP mera u poređenju sa merama zasnovanim na KL, kao i zasnovanim na EMD. Ovo je potvrđeno i u tabeli 4.15, slučaj $K = 5$, gde je dato CPU vreme procesiranja za KTH-TIPS bazu podataka. Stavljeno je da je $k_{iter} = 20$, $d = 15$ i $l_{max} = 30$, $l_{max} = 70$ ili $l_{max} = 100$ za sve eksperimente. Kao što se može videti iz rezultata prikazanih na grafikonu 4.18 i u tabeli 4.15, u slučaju KTH-TIPS baze tekstura, najbolji *"trade-off"* između preciznosti prepoznavanja i CPU vremena procesiranja za sve GMM-LPP mere je dobijen za slučaj $l_{max} = 70$. Takođe je prikazan i najbolji *"trade-off"* između preciznosti prepoznavanja i CPU vremena procesiranja za sve mere sa kojima se poredimo. Naravno, srednja vrednost CPU vremena procesiranja za testirane mere ne zavisi od baze podataka (razlike su male), već zavisi samo od veličine prostora obležja (koju smo fiksirali). Za veličinu GMM smo koristili K, a u slučaju GMM-LPP mera zavisi od veličine transformisanog prostora parametara l_{max}. Dakle, na osnovu tabele 4.15 i grafikona 4.16 i 4.17, prethodni zaključi za *"trade-off"* između preciznosti prepoznavanja i CPU vremena procesiranja važe za sve baze podataka.

Napominjemo da CPU vreme procesiranja za SR-EMD-M meru (SR-EMD-M-L ima istu složenost), nije dato u tabeli 4.15, jer to nije implementirano. Napominjemo da njena računska složenost, data u Sekciji (3.2.5) je najmanje nekoliko puta veća od računske složenosti EMD-KL čije je CPU vreme procesiranja dato u tabeli 4.15 i ima mnogo veće CPU vreme procesiranja nego mere zasnovane na KL. Prema tome, zaključujemo da predložene GMM-LPP mere imaju

[1]Pretpostavljamo da dva GMM-a imaju istu veličinu.

najbolji *"trade-off"* u odnosu na sve ostale mere sa kojima se poredi.

U cilju sumiranja eksperimentalnih rezultata, a i da bismo uvideli prednosti i mane predloženih GMM-LPP mera, još jednom napominjemo sledeće: mere sličnosti između GMM-ova sa kojima se poredimo zasnovane na KL-divergenci su KL_{WE}, KL_{MB} i KL_{VAR} date sa (3.4), (3.5) i (3.8), kao sa i merama datim sa (3.1), (3.7) koje su zasnovane na aproksimaciji KL divergence između GMM-ova. Sve takve aproksimacije uključuju dobijanje *ground distances* u obliku KL-a između pripadajućih komponenti Gausijana. Ove *ground distances* dobijamo u fazi eksploatacije. One su date u zatvorenoj formi sa (3.3), čija je računska složenost (videti Sekciju 3.2.5) reda $O(K^2d^3)$ gde K označava veličinu GMM-ova (svi su iste veličine) i gde kovarijanse koje odgovaraju komponentama Gausijana formata $d \times d$. Mere bazirane na EMD meri, EMD-KL mera predložena u [41] i SR-EMD-M, SR-EMD-M-L predložene u [27] nisu direktno bazirane na aproksimaciji KL divergence između određenih GMM-ova. Međutim, one između Gausovih komponenti koriste kao *ground distance* KL-divergencu u zatvorenoj formi ili 3.12. Prema tome, ove mere su još kompleksnije, tj. njihova računska složenost je najmanje reda $O(K^5d^3)$ (videti Sekciju 3.2.5). GMM-LPP mere koje su ovde predložene takođe nisu direktno zasnovane na aproksimaciji KL divergence između GMM-ova. Štaviše, kako ove mere koriste projektovanje LPP-tipa originalnog parametarskog prostora na niže dimenzionalni prostor parametara, one koriste *ground distances* između pripadajućih komponenti Gausijana u originalnom prostoru parametara samo u fazi treninga. *Ground distances* se računaju *off-line*. Prema tome, kako je prezentovano u Sekciji 3.2.3, ovo transformiše problem nalaženja distanci između dva GMM-a u aproksimativan problem nalaženja distanci između dva ponderisana "oblaka" niže-dimenzionalnih euklidskih vektora u $\mathbb{R}^l$, sa $l << d^2$, svaki odgovara određenom GMM-u. Stoga, one koriste euklidsku distancu u fazi eksploatacije za koju je rađena računska složenost (videti Sekciju 3.2.5). Ta računska složenost za GMM-LPP mere ja data sa $O(K^2l)$ i mnogo je manja u poređenju sa prethodno pomenutim merama. Nasuprot merama zasnovanim na KL koje su prethodno pomenute i merama sličnim SR-EMD-M, predložena GMM-LPP može biti zadata u svojoj nadgledanoj verziji (učenje u kojem svaka komponenta Gausijana u trenažnom skupu ima labelu klase), koja je prezentovana u Sekciji 3.2.4. Kao što se može

videti iz eksperimenata na realnim podacima u zadatku prepoznavanja teksture, prednosti nadgledanog pristupa su višestruki. Ponovimo da je naša pretpostavka da parametri Gausijana koji pripadaju GMM-ovima koji se koriste u nekom određenom problemu, leže približno u nekoj niže-dimenzionalnoj površi ulegnutoj u više-dimenzionalni konfiguracijski prostor parametara uz neka ograničenja efikasnosti aplikacije predložene GMM-LPP. Ovo je od velike važnosti za dobijanje veće tačnosti prepoznavnja sa višom računskom efikasnošću. Naravno, ako to nije slučaj, GMM-LPP može u najgorem slučaju dati bolji *" trade-off"* između tačnosti prepoznavanja i računske složenosti, što je često prihvatljivo u Veštačkoj inteligenciji i Ekspertskim sistemima koji rade sa velikim podacima. Demonstrirano je na različitim bazama tekstura, da je ovo slučaj sa zadacima prepoznavanja tekstura. U stvari, korišćenjem nenadgledane verzije predložene GMM-LPP, dobijamo veću tačnost prepoznavnja u poređenju sa svim merama baziranim na KL sa kojima se poredimo, za sve baze tekstura, pri čemu je zadržano da je računska složenost mnogo manja (videti Sekciju 4.0.7). Tačnost prepoznavanja dobijena za nadgledane GMM-LPP je blizu tačnosti prepoznavnaja *sparse* reprezentacije bazirane na SR-EMD-M i SR-EMD-M-L, ali ima mnogo manju računsku složenost. Napominjemo, da *sparse* reprezentacija zasnovana na SR-EMD-M i SR-EMD-M-L koja svuda daje bolje rezultate tačnosti prepoznavnja, takođe koristi neku vrstu niže-dimenzionalnog uleganja, kao i to da se *sparse* reprezentacija može videti kao reprezentacija podataka na uniji niže-dimenzionalnih potprostora.

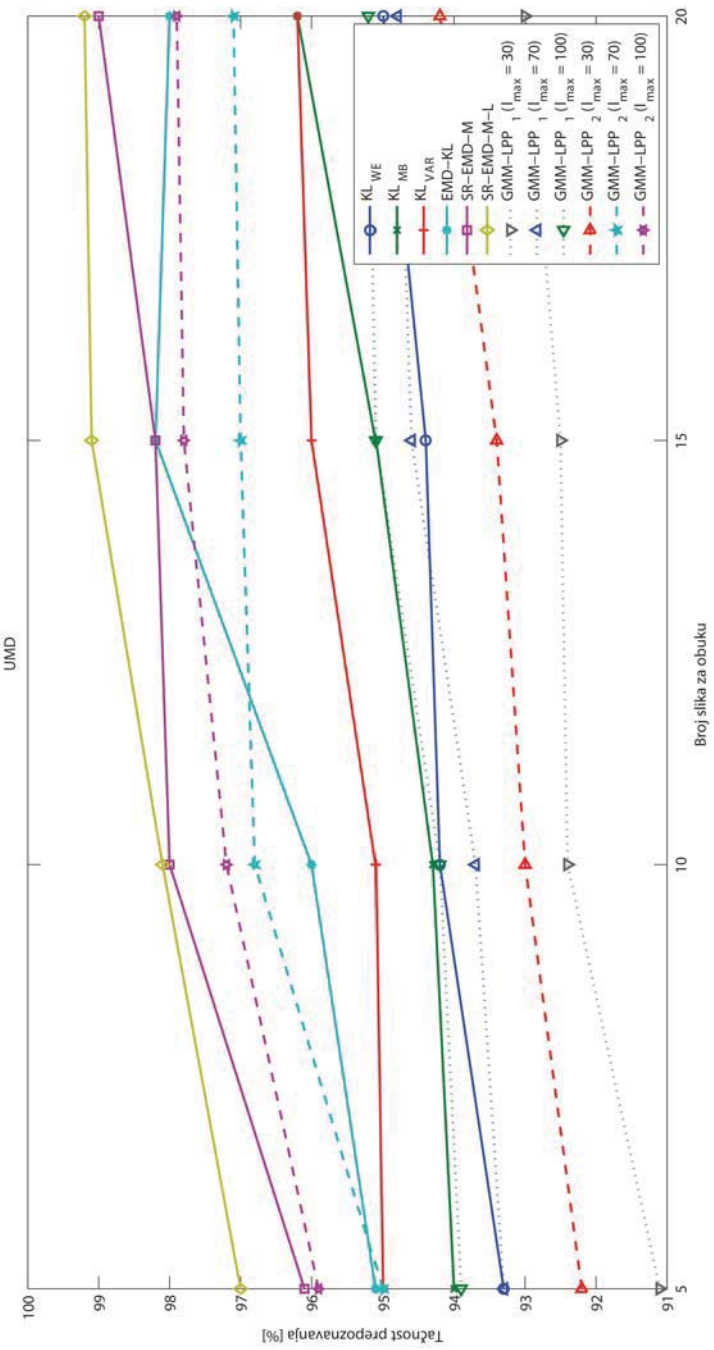


Tabela 4.16: Tačnost prepoznavanja za UMD bazu tekstura u odnosu na broj primera u skupu za obuku

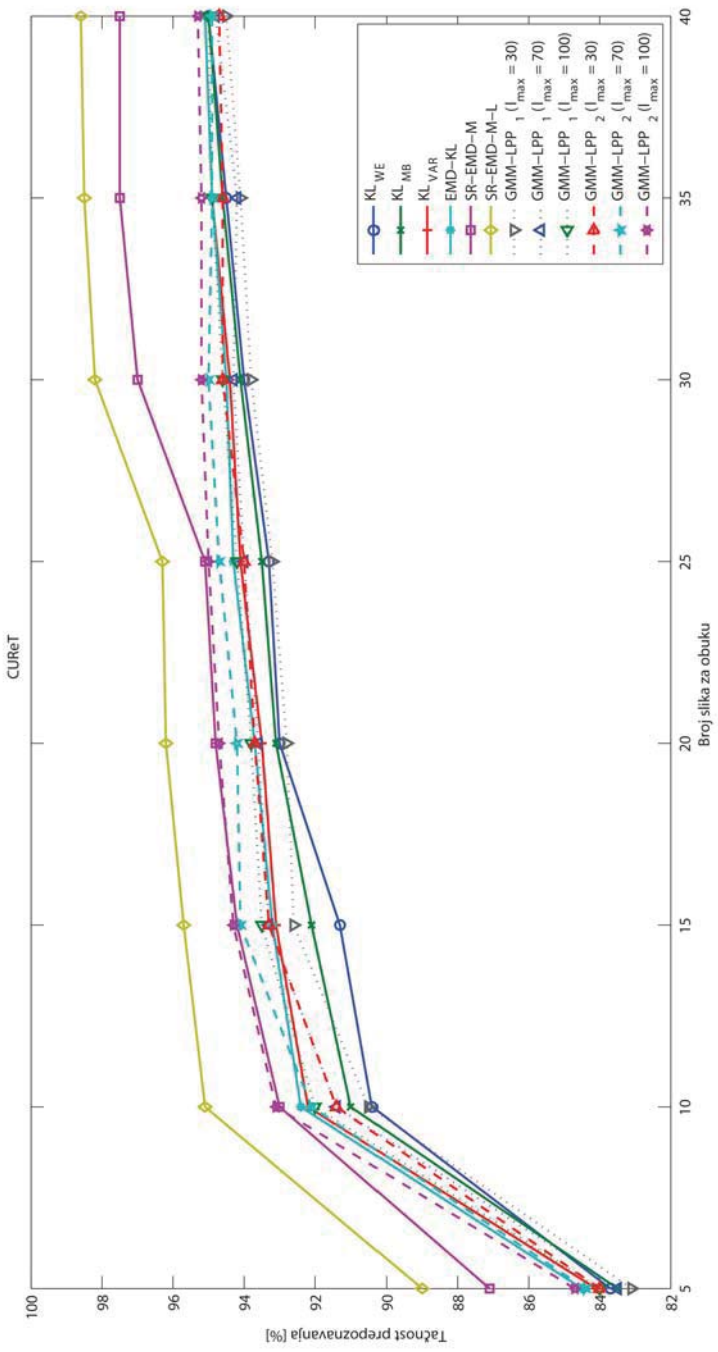


Tabela 4.17: Tačnost prepoznavanja za CUReT bazu tekstura u odnosu na broj primera u skupu za obuku

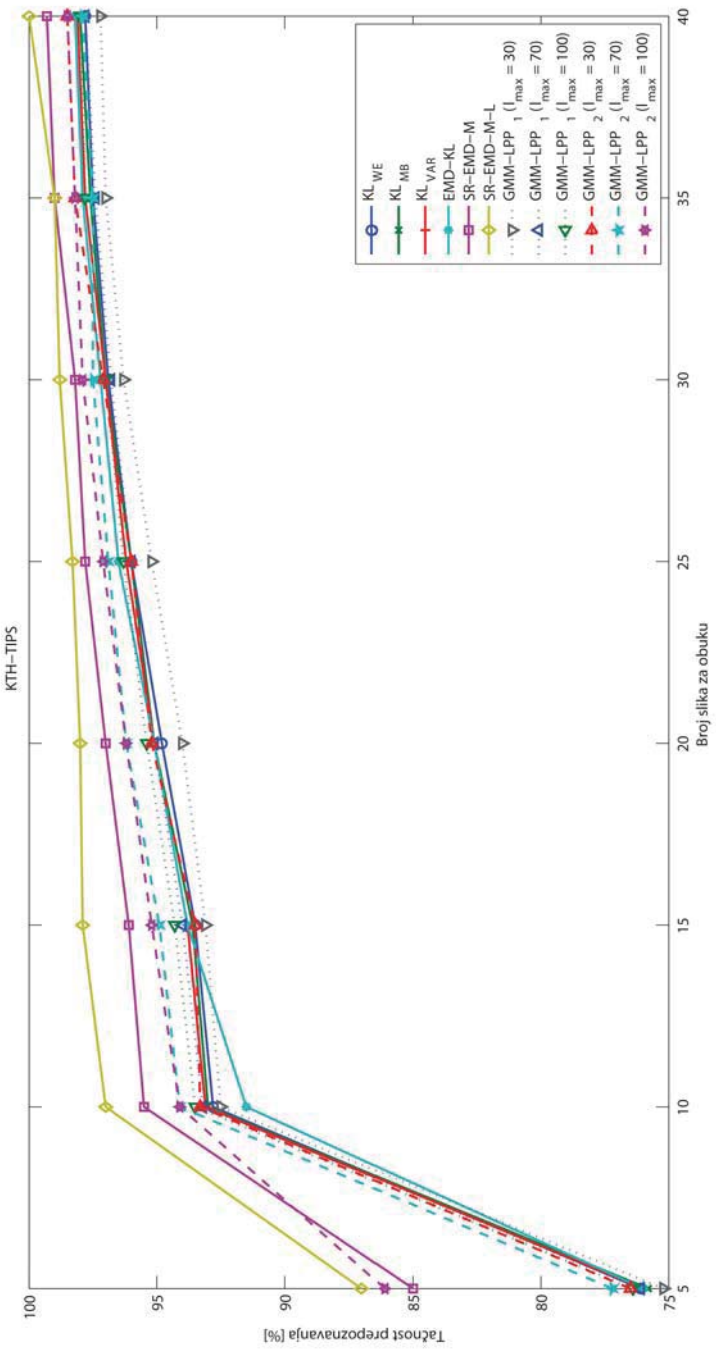


Tabela 4.18: Tačnost prepoznavanja za KTH-TIPS bazu tekstura u odnosu na broj primera u skupu za obuku

Konačno, upoređujemo predložene GMM-LPP mere ($l_{max} = 70$ slučaj) sa šest metoda klasifikacija teksture. Ovi rezultati u smislu tačnosti prepoznavanja su prikazani u tabeli (4.19). Rezultati su uzeti iz relevantnih radova koji su referencirani u prvoj koloni tabele. Simbol "-" znači da je rezultat nedostupan. Zaključujemo da za KTH-TIPS bazu podataka u slučajevima $N = 10$ i $N = 40$ predloženen GMM-LPP mere daju veću tačnost prepoznavanja u odnosu na metode sa kojima se porede, osim [30], koja daje najbolje rezultate za $N = 40$. Za UMD i CUReT bazu podataka, nadgledana GMM-LPP$_2$ daje tačnost prepoznavanja koja je uporediva sa najboljom od prikazanih metoda, dok GMM-LPP$_1$ daje manju tačnost (iako ne znatno manju) osim za CUReT bazu podataka u slučaju $N = 26$, gde metoda koja je predložena u [15] i [30] se pokazuje kao mnogo bolja.

Tabela 4.19: Tačnost prepoznavanja predložene nenadgledane GMM-LPP$_1$ i nadgledane GMM-LPP$_2$ u poređenju sa *state-of-the-art* metodama za prepoznavanje teksture. U aplikacijama za $GMM - LPP$, postavljeno je da je $K = 10$ i $l_{max} = 70$.

	KTH-TIPS			UMD			CUReT	
metoda	5	10	40	5	10	20	10	26
Zhang[43]	80.1	90.0	96.1	-	-	-	80.0	91.1
Hayman[15]	78.3	85.3	94.8	-	-	-	91.0	97.6
VZ-joint [39]	72.9	80.5	92.1	-	-	-	83.4	93.1
WMFS [20]	-	-	96.5	-	-	98.7	-	-
Liu [30]	80.5	87.8	99.3	95.0	97.5	99.3	91.5	98.3
CLBP [13]	76.1	85.5	96.8	92.4	96.0	98.0	93.6	92.9
GMM-LPP$_1$	76.1	93.2	97.8	93.3	93.7	94.9	91.4	94.1
GMM-LPP$_2$	77.2	94.1	97.9	95.0	96.8	97.1	92.1	94.8

Glava 5

Baze podataka korišćene u eksperimentima

U eksperimentima na realnim podacima korišćene su tri baze tekstura. To su: UMD (videti [42]), CUReT (videti [8]) i KTH-TIPS (videti [15]).

KTH-TIPS je baza koja je nastala kao dodatak CUReT baze podataka. Obe se koriste za algoritme klasifikacije radi zadataka prepoznavanja u realnom svetu. CUReT baza podataka sadrži 61 vrstu materijala, tj. teksture (koji variraju u položaju i osvetljenju, ali koji su slikani sa iste distance). KTH-TIPS ima za cilj da uvede:
- varijacije u skaliranju ili varijacije položaja i osvetljenja. Ovo omogućava istraživanja na temu kako nepoznato rastojanje utiče na klasifikaciju teksture, kao i na robustnost algoritma;
- druge uzorke (eng. *sample*) podskupova CUReT tekstura, koji su uzeti na drugačiji način. Cilj je da se vidi da li je zaista moguće klasifikovati teksture u realnom svetu.

U nastavku su prikazane slike predstavnika tekstura iz KTH-TIPS baze. Dakle, ova baza ima 10 klasa i ukupno 8010 slika. Ime svake klase je napisano iznad svakog predstavnika klase.

U radu [8] autori uvode CUReT bazu. Rad [8] rezultuje sa tri baze podataka:
- BRDF (eng. *bidirectional reflectance distribution function*) baza koja sadrži i mere refleksije za preko 60 uzoraka, svaki uzorak je posmatran

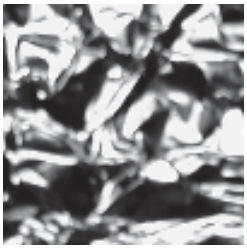
u preko 200 kombinacija pogleda i svetlosnih usmerenja,
- BRDF parametarska baza sa parametrima iz dva razlicita BRDF modela: Oren-Nayar model i Koenderink reprezentacija (videti [8]),
- BTF (eng. *bidirectional texture function*) sa slikama tektura sa preko 60 uzoraka, svaki uzorak je posmatran u preko 200 kombinacija pogleda i svetlosnih usmerenja.

U nastavku je prikazana 61 slika teksture iz realnog sveta koje su korisćene za potrebe rada [8] i na osnovu kojih je nastala CUReT baza.

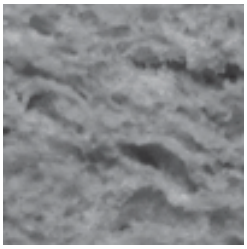
Takođe su prikazane i slike predstavnika za svaku klasu UMD baze. Klase su označene brojevima, iznad svake slike se nalazi broj klase kojoj pripada. Ova baza ima 25 klasa i ukupno 1000 slika. U pitanju su realne 3D slike koje su slikane u različitim svetlosnim uslovima i pozicijama kamere.

KTH-TIPS

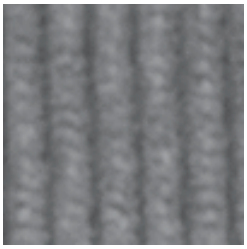
aluminijumska folija



crni hleb

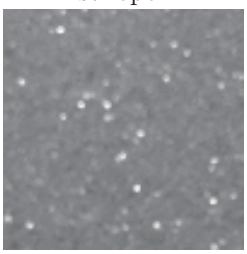


somot

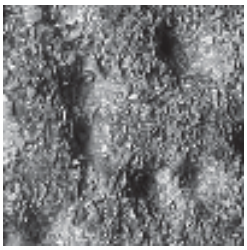


pamuk

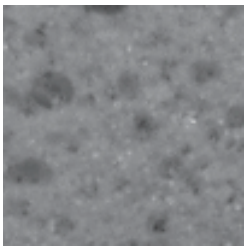
stiropor



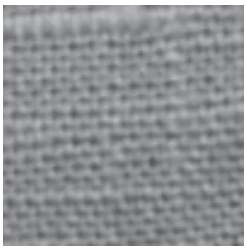
kreker



sunđer

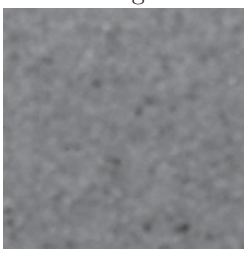


lan



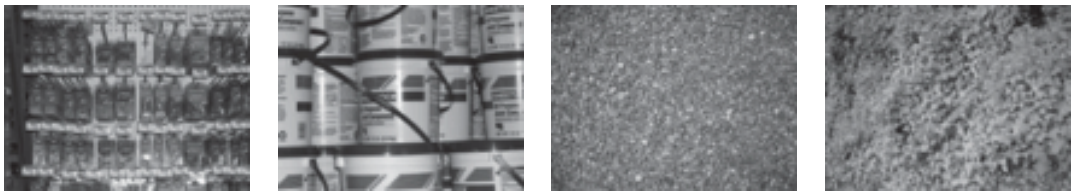
pomarandžina kora

šmirgla

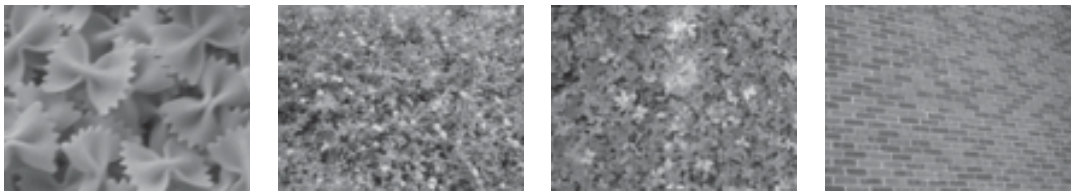


UMD

1 2 3 4

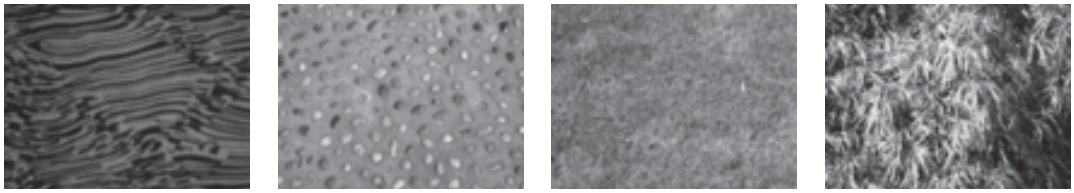


5 6 7 8

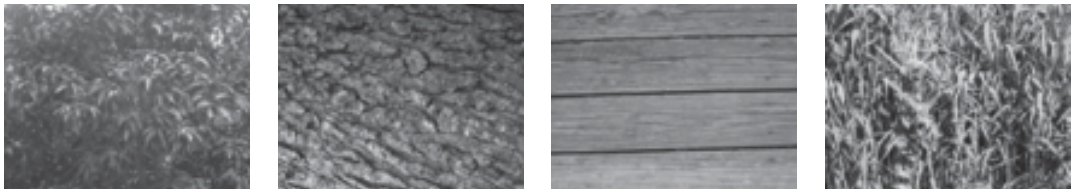


9 10 11 12

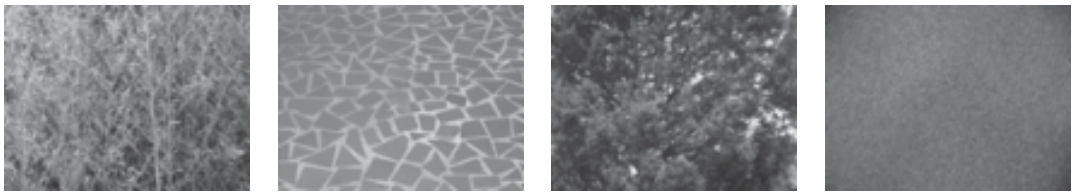
13 14 15 16



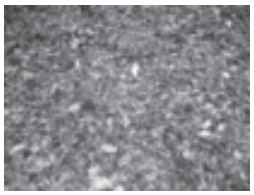
17 18 19 20



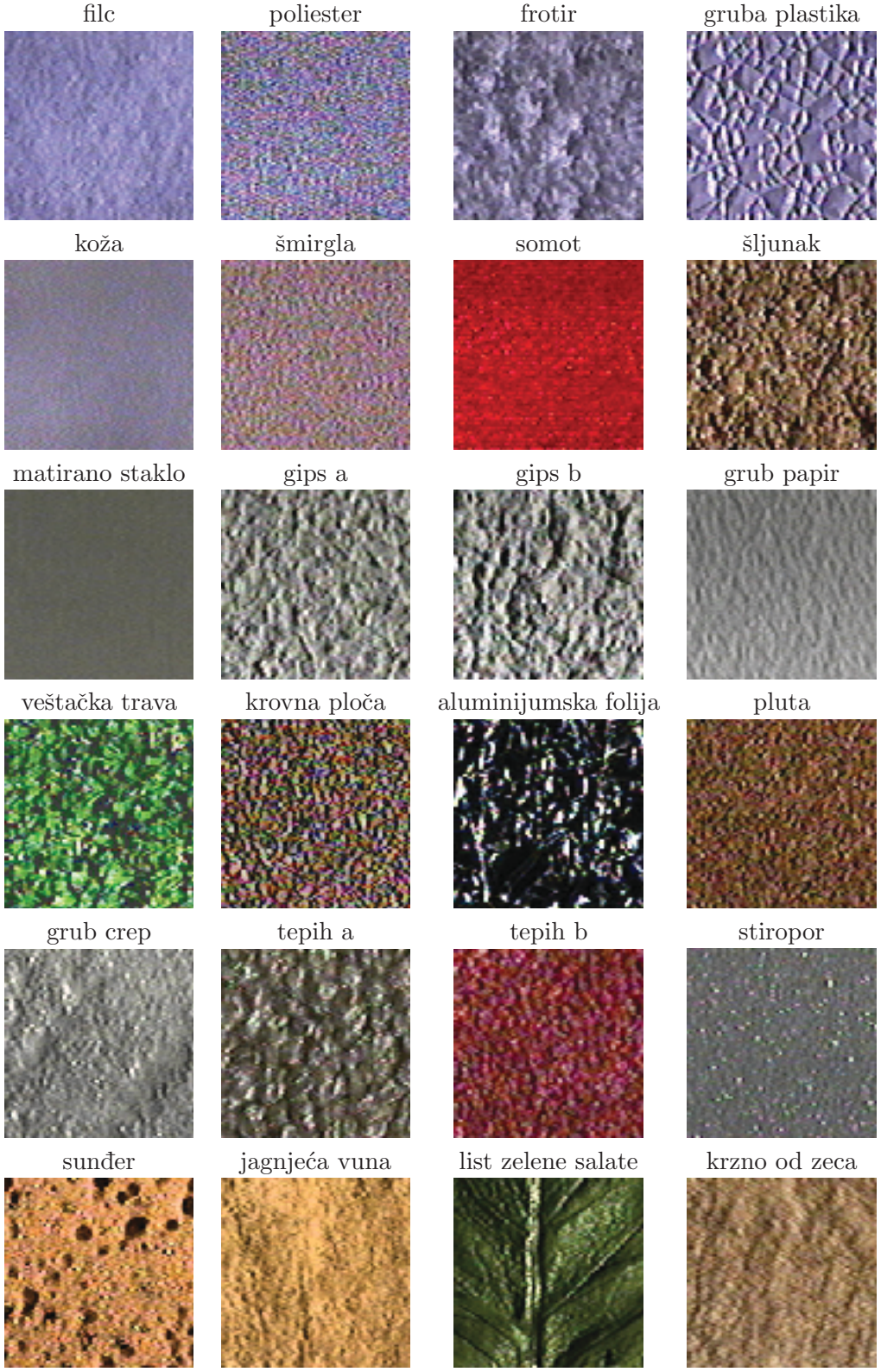
21 22 23 24



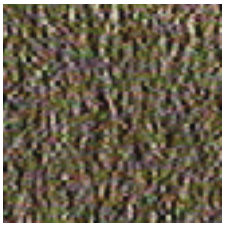
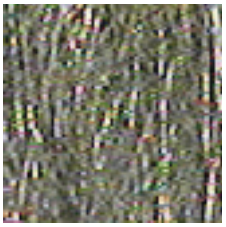
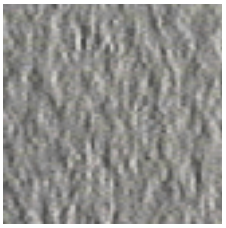
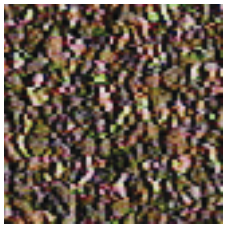
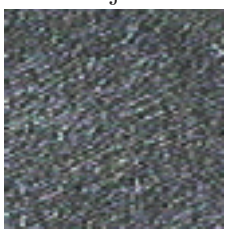
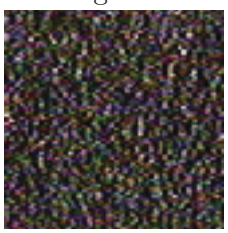
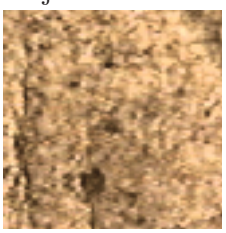
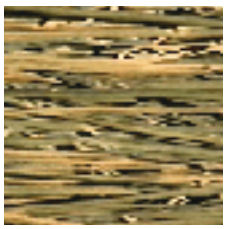
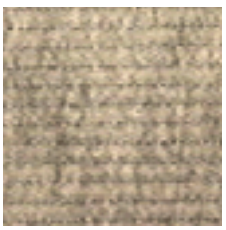
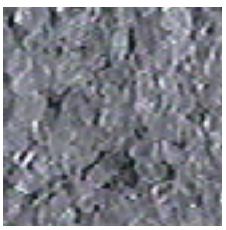
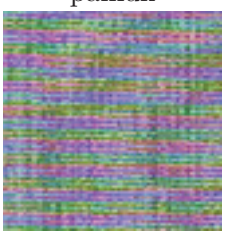
25

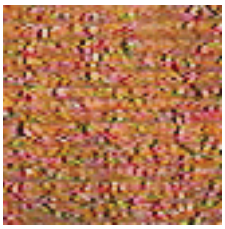
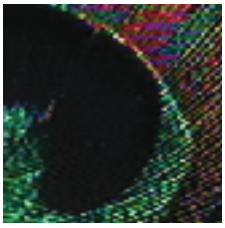
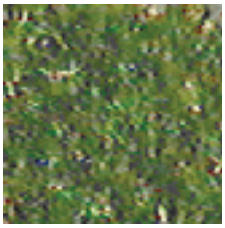


CuRET


filc
poliester
frotir
gruba plastika
koža
šmirgla
somot
šljunak
matirano staklo
gips a
gips b
grub papir
veštačka trava
krovna ploča
aluminijumska folija
pluta
grub crep
tepih a
tepih b
stiropor
sunđer
jagnjeća vuna
list zelene salate
krzno od zeca

CuRET

kamena ploča	sunđer	izolacioni materijal	tamni papir
			
uveličan poliester	uveličan gips b	uveličan grub papir	uveličana krovna ploča
			
škriljac a	škriljac b	obojene sfere	krečnjak
			
cigla a	rebrasti papir	ljudska koža	slama
			
opeka b	somot	kristali soli	lan
			
beton a	pamuk	kamenje	crni hleb
			

CuRET			
beton b	beton c	kukuruzna ljuska	beli hleb
			
Soleirolia biljka	drvo a	pomarandžina kora	drvo b
			
paunovo pero	kora drveta	kreker a	kreker b
			
mahovina			
			

Glava 6

Budući pravci istraživanja

6.0.8 NPE

Autori u [17] inspirisani geometrijski motivisanim pristupima za analizu podataka u visoko dimenzionalnom prostoru razmatraju slučaj kada ti podaci leže na podmnogostrukosti euklidskog prostora. Predlažu novi algoritam učenja podmnogostrukosti koji zovu *Neighborhood Preserving Embedding* (NPE). Navedimo nekoliko aspekata ovog pristupa:

- NPE ima nekoliko sličnih osobina kao LPP algoritam (videti 2.3.2). Oba algoritma posmatraju lokalne strukture podataka date mnogostrukosti. Međutim, njihove ciljne funkcije su različite.

- NPE je linearan. Zato je brži i pogodniji za praktičnu primenu. Može se primeniti na originalan prostor ili na Hilbertov prostor (eng. *reproducing kernel Hilbert space*)(RKHS) u koji se podaci mapiraju.

- NPE se može primeniti i kod nadgledanog i kod nenadgledanog učenja. Kada su informacije o klasama dostupne može se upotrebiti da se dobije bolja težinska matrica.

U [17] NPE algoritam je primenjen na bazu lica.

Linearne tehnike za redukciju dimenzionalnosti

Dve najpopularnije tehnike za učenje su PCA (videti 2.3.1) i Linearna diskriminativna analiza (eng. *Linear Discriminant Analysis*) (LDA). PCA je nenadgledano, a LDA nadgledano učenje.

Osnovna ideja PCA je da se podaci projektuju na pravac koji ima maksimalnu varijansu. Na taj način se greška minimizuje. Neka je dat skup podataka $x_1, \ldots, x_n$ i neka je a vektor koji vrši projektovanje, tj. $y_i = a^T x_i$. Ciljna funkcija je

$$\begin{aligned} a_{opt} &= \arg\max_a \sum_{i=1}^{n} (y_i - \bar{y})^2 \\ &= a^T C a \end{aligned}$$

gde je $\bar{y} = \frac{1}{n}\sum y_i$, a C je kovarijansna matrica. Karakteristični vektori kovarijansne matrice podataka odgovaraju najvećim karakterisitčnim vrednostima.

Dok PCA zahteva pravac da bi se dobila sto bolja reprezentacija podataka u niže dimenzionalnom prostoru, LDA zahteva pravac da bi se dobila što veća diskriminativnost podataka. Pretpostavimo da podaci leže u l klasa. Ciljna funkcija je

$$\begin{aligned} a_{opt} &= \arg\max_a \frac{a^T S_B a}{a^T S_W a} \\ S_B &= \sum_{i=1}^{l} n_i (m_i - m)(m_i - m)^T \\ S_W &= \sum_{i=1}^{l} \left(\sum_{j=1}^{n_i} (x_j^{(i)} - m^{(i)})(x_j^{(i)} - m^{(i)})^T \right) \end{aligned}$$

gde je m centorid, n_i broj uzoraka u i-toj klasi, $m^{(i)}$ vektor srednjih vrednosti i-te klase $x_j^{(i)}$ j-ti uzorak i-te klase.

NPE

Neka je dat skup tačaka $x_1, \ldots, x_m$ u $\mathbb{R}^n$. Da bi se izvršila redukcija dimenzionalnosti potrebno je naći matricu A koja preslikava

ovih m tačaka u skup tačaka $y_1, \ldots, y_m$ iz $\mathbb{R}^d$ ($d \ll n$), pri čemu y_i "reprezentuju" x_i na sledeći način $y_i = A^T x_i$.

Procedura algoritma je data sa (videti [17]):

- **Konstruisanje grafa suseda.** Označimo sa G graf suseda koji konstruišemo. Neka i-ti čvor odgovara tački x_i. Dva načina za konstruisanje grafa su:
  - *K najbližih suseda (KNN):* Postavljanjem usmerene grane od čvora i ka čvoru j, za svaki čvor j koji spada u K najbližih suseda čvora i
  - ϵ *susedstvo:* Postavljanjem grane između čvorova i i j ako je $\|x_j - x_i\| \leq \epsilon$
- **Dobijanje težina.** Neka je W težinska matrica koju konstruišemo. Označimo sa W_{ij} težinu grane koja spaja čvor i i čvor j, a sa 0 ako ne postoji grana koja povezuje ta dva čvora. Težine grana se mogu dobiti minimizovanjem sledeće ciljne funkcije:

$$min \sum_i \|x_i - \sum_j W_{ij} x_j\|$$

  uz uslov

$$\sum_j W_{ij} = 1, j = 1, \ldots, m.$$

- **Dobijanje linearnih projekcija.** Prvo treba rešiti karakterističnu jednačinu

$$XMX^T a = \lambda XX^T a, \tag{6.1}$$

  gde su

$$X = (x_1, \ldots, x_m),$$
$$M = (I - W)^T (I - W),$$
$$I = diag(1, \ldots, 1).$$

  M je simetrična i semi-pozitivno definitna matrica. Neka su vektori kolone $(a_0, \ldots, a_{d-1})$ rešenja jednačine 6.1 koji su poređani

u skladu sa poretkom njihovih odgovarajućih karakterističnih vrednosti $\lambda_0, \ldots, \lambda_{d-1}$. Dakle, traženo preslikavanje je

$$x_i \to y_i = A^T X_i$$

$$A = (a_0, \ldots, a_{d-1})$$

gde je y_i d-dimenzionalni vektor, a A je $n \times d$ matrica.

Sada ćemo teorijski objasniti NPE algoritam (eng. *Neighborhood Preserving Embedding*), kao što je to urađeno u [17]. Prvo konstruišemo graf suseda za zadati skup tačaka, odnosno podataka. Za svaku tačku nalazaimo K najbližih suseda. U mnogim slučajevima skup tačaka možda leži na nelinearnoj mnogostrukosti, ali je razumljivo pretpostaviti da je svako lokalno susedstvo linearno. Dakle, možemo okarakterisati lokalnu geometriju ovih delova (eng. *patches*) linearnim koeficijentima koji rekonstruišu svaku tačku sa njenim susedstvom. Greška rekonstrukcije se meri funkcijom troškova

$$\phi(W) = \sum_i \|x_i - \sum_j W_{ij} x_j\|^2,$$

$\phi(W)$ sabira kvadrate rastojanja između svih tačaka i njihovih "rekonstrukcija". Razmotrimo sada problem preslikavanja originalnih tačaka podataka na liniju tako da se svaka tačka na liniji može reprezentovati kao linearna kombinacija svog susedstva sa koeficijentima W_{ij}. Neka je $y = (y_1, \ldots, y_m)^T$ takvo preslikavanje. Da bismo odredili da li je preslikavanje "dobro" minimiziramo sledeću funkciju troškova

$$\phi(y) = \sum_i (y_i - \sum_j W_{ij} y_j)^2,$$

uz odgovarajuće uslove.

Pretpostavimo da je transformacija $y^T = a^T X$ linearna (videti [17]), gde je i-ta kolona vektora X x_i. Neka je

$$z_i = y_i - \sum_j W_{ij} y_j,$$

odnosno

$$\begin{aligned} z &= y - Wy \\ &= (I - W)y. \end{aligned}$$

Sada se funkcija troškova može zapisati u sledećem obliku

$$\begin{aligned}
\phi(y) &= \sum_i (y_i - \sum_j W_{ij} y_j)^2 \\
&= \sum_i (z_i)^2 \\
&= z^T z \\
&= y^T (I - W)^T (I - W) y \\
&= a^T X (I - W)^T (I - W) X^T a \\
&= a^T X M X^T a
\end{aligned}$$

gde je $M = (I - W)^T (I - W)$. Matrica XMX^T je simetrična i pozitivno semidefinitna. Da bi se izbegao proizvoljan faktor skaliranja pri projektovanju, uvodi se uslov

$$y^T y = 1 \Rightarrow a^T X X^T a = 1.$$

Dakle, problem minimizacije sveden je na

$$\arg\min_a a^T X M X^T a, \tag{6.2}$$

uz uslov $a^T X X^T a = 1$.

XMX^T i XX^T su simetrične i pozitivno semidefinitne. Neka je l rang matrice X. U tom slučaju se korišćenjem SVD (eng. *Singular Value Decomposition*) tačke podataka mogu projektovati u l-dimenzionalni podprostor u kojem matrica X postaje nesingularna

$$X = USV^T$$

$$\tilde{X} = U^T X = SV^T$$

gde je $U = (u_1, \ldots, u_l)$, u_i je karakteristični vektor matrice XX^T, $V = (v_1, \ldots, v_l)$, v_i je karakteristični vektor matrice $X^T X$, S je $l \times l$ dijagonalna matrica koja na dijagonali ima ne-nula singularne vrednosti X. S i V su potpunog ranga, pa je i $\tilde{X}$ potpunog ranga. Na ovaj način optimalno projektovanje čine karakteristični vektori matrice $(\tilde{X}X^T)^{-1}(\tilde{X}M\tilde{X}^T)$.

6.0.9 Budući pravci istraživanja

Kao pravac budućeg istraživanja predlaže se model (i samim tim GMM mera sličnosti) koji unosi princip očuvanja lokalnosti prilikom odgovarajuće linearne transformacije A, prostora Gausovih komponenti prisutnih u odgovarajućim GMM-ovima, procenjenim iz podataka, u nižedimenzionalni prostor euklidskih predstavnika.

Cilj je da se iskoristi pristup prisutan u *NPE* (videti 6.0.8) da bi se dobila transformaciona matrica A koja transformiše vektorizovane predstavnike $P_i \in Sym_+(n)$ pomenutih Gausovih komponenti, u niskodimenzioni euklidski vektorski prostor. Zatim bi se koristile iste mere sličnosti između GMM-ova koje porede ponderisane "oblake" niskodimenzionalnih predstavnika odgovarajućih Gausovih komponenti, a koje su predstavljene u prethodnim poglavljima.

Neka je $Sym_+(n)$ konus nenegativno definitnih matrica ulegnut u $\mathbb{R}^n$. Kao i u slučaju *NPE*, za svaki i-ti Gausijan $\mathcal{N}(\mu_i, \Sigma_i)$, uvodi se pretpostavka da se odgovarajući predstavnik

$$P_i = |\Sigma_i|^{\frac{-1}{n+1}} \begin{bmatrix} \Sigma_i + \mu_i \mu_i^T & \mu_i \\ \mu_i^T & 1 \end{bmatrix} \in Sym_+(n)$$

može aproksimirati nenegativnom sumom na sledeći način

$$\begin{aligned} P_i &\approx \sum_{j=1}^{K} W_{ij} P_j, \\ W_{ij} &\geq 0 \\ \|W\|_0 &\leq T, \end{aligned} \tag{6.3}$$

gde je K ukupan broj Gausovih komponenti prisutan u GMM-ovima dobijenim pri obuci. Ovde je, za razliku od *NPE*, uvedena pretpostavka da malo predstavnika P_j utiče na dato P_i. Naime, koristi se retka reprezentaciju, koja se oslikava u uslovu $\|W\|_0 < T$, gde je sa $\|\cdot\|_0$ data l_0 "norma" (ne zadovoljava osobinu nejednakosti trouglova) koja, za neku matricu $M \in \mathcal{M}(\mathbb{R}, n)$, predstavlja broj nenultih elemenata u M. Takođe, u (6.3) term $T > 0$ predstavlja neki predefinisani prag, koji diktira retkoću reprezentacije koeficijenata u W. Napomena je, da za razliku od *NPE*, gde figurišu euklidski vektori $x_i \in \mathbb{R}^N$, u (6.3) figurišu elementi $P_i \in Sym_+^(n)$. Napomenimo

da uslov $W_{ij} \geq 0$, $i, j \in \{1, \ldots, K\}$ obezbeđuju da aproksimacija $\sum_{j=1}^{K} W_{ij} P_j$ u (6.3) bude nenegativno definitna.

Dobijanje transformacione matrice A sastoji se iz dva koraka:

- Dobijanje težina W_{ij} minimizacijom greške aproksimacije, tj. rešavanjem jednog od dva problema minimizacije:

  problem **A**

$$\min_{W_{ij}} \sum_{i=1}^{K} D_{ld} \left(\sum_{j=1}^{K} W_{ij} P_j, P_i \right),$$

$$\text{pri uslovima} \tag{6.4}$$

$$W_{ij} \geq 0$$

$$\|W\|_0 \leq T. \tag{6.5}$$

  problem **B**

$$\min_{W_{ij}} D_{ld} \left(\sum_{j=1}^{K} W_{ij} P_j, P_i \right), \ i = 1, \ldots, K$$

$$\text{pri uslovima} \tag{6.6}$$

$$W_{ij} \geq 0$$

$$\|W\|_0 \leq T. \tag{6.7}$$

  U **A** i **B** je za $X, Y \in Sym_+(n)$, "distanca" D_{ld} data sa

$$D_{ld}(X, Y) = \operatorname{tr}(XY^{-1}) - \log\det(XY^{-1}) - n, \tag{6.8}$$

  gde $\operatorname{tr}(\cdot)$ označava trag matrice. Pošto su $\operatorname{tr}(\cdot)$ i $\log\det(\cdot)$ invarijantni u odnosu na transformaciju sličnosti, važi sledeće:

$$D_{ld}\left(\sum_{j=1}^{K} W_{ij}P_j, P_i\right) = \text{tr}\left(P_i^{-\frac{1}{2}}\left(\sum_{j=1}^{K} W_{ij}P_j\right)P_i^{\frac{1}{2}}\right)$$
$$-\log\det\left(P_i^{-\frac{1}{2}}\left(\sum_{j=1}^{K} W_{ij}P_j\right)P_i^{\frac{1}{2}}\right) - n$$
$$= \sum_{j=1}^{K} W_{ij}\text{tr}\,(\hat{P}_j^{(i)}) - \log\det\left(\sum_{j=1}^{K} W_{ij}\hat{P}_j^{(i)}\right) - n, \qquad (6.9)$$

gde je $\hat{P}_j^{(i)} = P_i^{-\frac{1}{2}}P_jP_i^{\frac{1}{2}}$, za svako $i, j \in \{1, \ldots, K\}$.

Uvodimo dodatno ograničenje, dato sa $\sum_{j=1}^{K} W_{ij}P_j \preceq P_i$, za svako $i, j \in \{1, \ldots, K\}$, koje je ekvivalentno sa $\sum_{j=1}^{K} W_{ij}\hat{P}_j \preceq I_n$. Time se obezbeđuje da rezidual $P_i - \sum_{j=1}^{K} W_{ij}P_j$ bude nenegativno definitna matrica za svako $i \in \{1, \ldots, K\}$. Takođe, vrši se konveksna relaksacija uslova $\|W\|_0 \leq T$, prelaskom na l_1 normu $\|\cdot\|_1$ koja je konveksna obvojnica od $\|\cdot\|_0$. Dobijaju se sledeći optimizacioni problemi

problem $\mathbf{A}'$

$$\min_{W_{ij}} \sum_{i=1}^{K}\left(\sum_{j=1}^{K} W_{ij}\text{tr}\,(\hat{P}_j^{(i)}) - \log\det\left(\sum_{j=1}^{K} W_{ij}\hat{P}_j^{(i)}\right)\right)$$
pri uslovima (6.10)
$$W_{ij} \geq 0$$
$$\sum_{i=1}^{K}\sum_{j=1}^{K} W_{ij} < T, \qquad (6.11)$$

problem $\mathbf{B}'$

$$\min_{W_{ij}} \sum_{j=1}^{K} W_{ij} \mathrm{tr}\,(\hat{P}_j^{(i)}) - \log\det\left(\sum_{j=1}^{K} W_{ij}\hat{P}_j^{(i)}\right),\ i = 1, \ldots, K$$

pri uslovima (6.12)

$$W_{ij} \geq 0$$

$$\sum_{i=1}^{K}\sum_{j=1}^{K} W_{ij} < T. \quad (6.13)$$

Problemi $\mathbf{A}'$ i $\mathbf{B}'$, spadaju u klasu MAXDET problema (videti [5]). To su konveksni problemi optimizacije, sa ograničenjima, kao što je dokazano u [5].

- Dobijanje transformacione matrice A rešavanjem istog problema minimizacije kao i u *NPE* pristupu, predstavljenom u poglavlju 6.0.8 (videti 6.2). Ako se rešava problem $\mathbf{B}'$, pošto rezultujuća težinska matrica W nije simetrična, pri dobijanju transformacione matrice A, kao u 6.0.8 se vrši simetrizacija W, tj. $W \leftarrow \frac{1}{2}(W + W^T)$.

Pri samoj eksploataciji, tj. fazi prepoznavanja, metodologija bi bila kao u već opisanim poglavljima.

Glava 7

Zaključak

U ovom istraživanju uvodimo novu meru sličnosti između GMM-va.

Ova mera sličnosti koristi LPP tehniku, u cilju učenja linearne projektivne matrice, koja će projektovati parametre GMM-ova na niže-dimenzionalni prostor parametara. Pri tom, lokalne osobine susedstva koje postoje u originalnom prostoru parametara GMM-ova treba da se očuvaju u transformisanom parametarskom prostoru.

Ovo značajno smanjuje računsku složenost rezultujuće metode, u odnosu na metode sa kojima se poredi. U isto vreme, dobijena je veća diskriminativnost između klasa, ako parametri GMM-ova leže na aproksimativno niže-dimenzionalnoj površi ulegnutoj u konusu pozitivno definitnih matrica. U eksperimentima koji su izvršeni na sintetičkim, kao i na realnim podacima, dobijen je mnogo veći *"trade-off"* između tačnosti prepoznavanja i računske složenosti u odnosu na GMM mere sličnosti sa kojima se poredimo. Veća tačnost prepoznavanja u odnosu na mere sa kojima se poredimo, takođe je dobijena i u eksperimentima sa sintetičkim podacima.

Kako je predložena metoda u potpunosti nenadgledana, u daljem istraživanju ćemo težiti primeni sličnih polu-nadgledanih metoda, sposobnim da dodatno povećaju diskriminativnost između klasa. Takođe, težiće se da se istraži efikasnost mera sličnosti izraženih pomoću različitih formi operatora agregacije.

Prilog A

Teorija verovatnoće

U ovom prilogu ćemo prezentovati osnovne pojmove iz teorije verovatnoće koji su neophodni za razumevanje ove teze.

Polazni pojam u teoriji verovatnoće je neprazan skup Ω koji predstavlja skup svih mogućih ishoda jednog eksperimenta. Skup Ω zove se prostor **elementarnih događaja** ili ishoda. Skup svih ishoda se obično označava sa F. **Slučajni događaj** ili jednostavno događaj A se definiše kao neki podskup skupa Ω.

Mera verovatnoće

Verovatnoća ili mera verovatnoće (eng. *probability measure*) je funkcija $P(A)$ čiji je argument skup A. Klasična definicija verovatnoće glasi: Verovatnoća $P(A)$ događaja $A \subseteq \Omega$ jednaka je količniku pozitivnih ishoda eksperimenta, koji doprinose realizaciji događaja A, i broja svih ishoda. Verovatnoća je, dakle funkcija $P : F \to \mathbb{R}$ koja zadovoljava sledeće osobine:

- $0 \leq P(A) \leq 1$
- $P(\Omega) = 1$
- Ako su A i B disjunktni događaji važi

  $$P(A \cup B) = P(A) + P(B).$$

  Ako A i B nisu disjunktni važi

  $$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Ove tri osobine zovu se aksiome verovatnoće.

Iako je argument P skup, često se korisiti opis događaja kao argument. Dakle, umesto $P(\omega : X(\omega) > Y(\omega))$ kraće pišemo $P(X > Y)$.

Slučajna promenljiva

Razmotrimo sada primer bacanja novčića. Neka se novčić baca 10 puta. Tada je element iz Ω niz pismo/glava dužine 10 npr. $\omega_0 = \langle P, P, P, G, P, G, G, P, G, G\rangle \in \Omega$. U praksi, obično nas ne interesuje verovatnoća dobijanja određenog niza pismo/glava. Obično nas interesuje neka realna funkcija ishoda, kao što je broj palih glava ili pisama. Ova funkcija, pri određenim uslovima, zove se **slučajna promenljiva** (videti [40], [32]).

Formalno, slučajna promenljiva X je funkcija $X : \Omega \to \mathbb{R}$. Korišćenje slučajne promenljive X označavamo tako što je pišemo u indeksu X.

Ako $X(\omega)$ prima naviše prebrojivo mnogo vrednosti $x_1, x_2, \ldots, x_k, \ldots$ zovemo je **diskretna slučajna promenljiva**. Ona je određena zakonom raspodele:

$$X : \begin{pmatrix} x_1 & x_2 & \ldots & x_k & \ldots \\ p_1 & p_2 & \ldots & p_k & \ldots \end{pmatrix},$$

gde je $p_k = P(X = k) := P(\omega : X(\omega) = k)$.

Ako $X(\omega)$ prima beskonačno mnogo vrednosti zovemo je **neprekidna slučajna promenljiva**. Verovatnoću da X prima vrednosti između dva realna broja a i b (gde je $a < b$) označavamo sa

$$P(a \leq X \leq b) := P(\omega : a \leq X(\omega) \leq b).$$

Kumulativna funkcija raspodele

Kumulativna funkcija raspodele (eng. *cumulative distribution function*) (CDF) je funkcija $F_X : \mathbb{R} \to [0, 1]$, koja je određena merom verovatnoće

$$F_X(x) \doteq P(X \leq x). \tag{A.1}$$

Osobine CDF su:

- $0 \leq F_X(x) \leq 1$.
- $\lim_{x \to -\infty} F_X(x) = 0$.
- $\lim_{x \to \infty} F_X(x) = 1$.
- $x \leq y \Rightarrow F_X(x) \leq F_X(y)$.

Funkcija mase verovatnoće

Kada je X diskretna slučajna promenljiva, najjednostavniji način da se reperezentuje mera verovatnoće koja odgovara slučajnoj promenljivi je da se direktno odredi verovatnoća za svaku vrednost koju slučajna promenljiva može da primi. Dakle, **funkcija mase verovatnoće** (eng. *probability mass function*) (PMF) je funkcija $p_x : \Omega \to \mathbb{R}$ takva da je (videti [32])

$$p_X \doteq P(X = x).$$

U slučaju diskretne slučajne promenljive sa $Val(X)$ označavamo skup mogućih vrednosti koje može X da primi.

Osobine PMF su:

- $0 \leq p_X(x) \leq 1$.
- $\sum_{x \in Val(X)} p_X(x) = 1$.
- $\sum_{x \in A} p_X(x) = P(X \in A)$.

Funkcija gustine verovatnoće

Za neke neprekidne slučajne promenljive, kumulativna funkcija raspodele $F_X(x)$ je diferencijabilna svuda. Definišemo **funkciju gustine verovatnoće** (eng. *Probability Density Function*) (PDF) kao izvod CDF

$$f_X(x) \doteq \frac{dF_X(x)}{dx}. \tag{A.2}$$

Napomenimo da za neke neprekidne slučajne promenljive PDF možda ne postoji, jer $F_X(x)$ nije diferencijabilna svuda.

Osobine PDF su:

- $f_X(x) \geq 0$.
- $\int_{-\infty}^{\infty} f_X(x) = 1$.
- $\int_{x \in A} f_X(x)dx = P(X \in A)$.

Očekivanje

Pretpostavimo da je X diskretna slučajna promenljiva za koju je PDF $p_X(x)$ i $g : \mathbb{R} \to \mathbb{R}$ je neka funkcija. U ovom slučaju se $g(x)$ može posmatrati kao slučajna promenljiva. Definišemo **očekivanje** ili **matematičko očekivanje** $g(X)$ kao (videti [32])

$$E[g(X)] \doteq \sum_{x \in Val(X)} g(x)p_X(x).$$

Očekivanje $E[X]$ se nalazi kada se $g(x)$ zameni sa x. Ovo očekivanje je poznato i kao centroida (eng. *mean*) slučajne promenljive X.

Ako je X neprekidna slučajna promenljiva za koju je PDF $f_X(x)$, onda očekivanje $g(X)$ je definisano sa

$$E[g(X)] \doteq \int_{-\infty}^{\infty} g(X)f_X(x)dx.$$

Osobine očekivanja su

- $E[a] = a$ za svako $a \in \mathbb{R}$.
- $E[af(X)] = aE[f(X)]$ za svako $a \in \mathbb{R}$.
- $E[f(X) + g(X)] = E[f(X)] + E[g(X)]$.
- Za diskretnu slučajnu promenljivu X važi $E[1\{X = k\}] = P(X = k)$.

Varijansa

Varijansa ili **disperzija** slučajne promenljive X je mera "koncentrisanosti" raspodele slučajne promenljive oko svoje centroide. Formalno, varijansa slučajne promenljive X definisana je sa

$$Var[x] \doteq E[(X - E(X))^2].$$

Koristeći osobine matematičkog očekivanja dobijamo

$$\begin{aligned} E[(X - E[X])^2] &= E[X^2 - 2E[X]X + E[X]^2] \\ & E[X^2] - 2E[X]E[X] + E[X]^2 \\ & E[X^2] - E[X]^2. \end{aligned}$$

Osobine varijanse su:

- $Var[a] = 0$ za svako $a \in \mathbb{R}$.
- $Var[af[X]] = a^2 Var[f(X)]$ za svako $a \in \mathbb{R}$.

Uslovna verovatnoća

Neka je B događaj za koji je $P(B) > 0$. **Uslovna verovatnoća** za bilo koja dva događaja A i B data je sa

$$P(A|B) \doteq \frac{P(A \cap B)}{P(B)}$$

ili

$$P(A \cap B) = P(A|B)P(B). \tag{A.3}$$

$P(A|B)$ je verovatnoća događaja A posle opservacije događaja B. Dva događaja su nezavisna ako i samo ako važi $P(A \cap B) = P(A)P(B)$ (ili, što je ekvivalentno, $P(A|B) = P(A)$). Dakle, nezavisni su ako događaj B nema nikakvog uticaja na događaj A.

Bajesova teorema

Kako je $P(A \cap B) = P(B \cap A)$, iz A.3 sledi

$$P(A|B)P(B) = P(B|A)P(A)$$

ili

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

Ovo je **Bajesova teorema** (videti npr. [40]).

Ako su $A_1, \ldots, A_N$ događaji koji čine particiju (razbijanje) prostora događaja (tj. oni su uzajamno disjunktni i njihova unija je Ω) tada

$$\begin{aligned} P(B) &= \sum_{i=1}^{N} P(B \cap A_i) \\ &\quad \sum_{i=1}^{N} P(B|A_i)P(A_i). \end{aligned} \tag{A.4}$$

Odavde dobijamo praktičniji oblik Bajesove teoreme

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{N} P(B|A_i)P(A_i)} \tag{A.5}$$

U problemima klasifikacije, B je često događaj koji opserviramo, a A_j je klasa modela. Izraz **a priori verovatnoća** se često koristi za $P(A_i)$, a **a posteriori verovatnoća** za $P(A_i|B)$.

Zajednička raspodela

Slučajan vektor dodeljuje svakoj tački iz prostora događaja Ω tačku iz $\mathbb{R}^p$. Podsetimo se da slučajna promenljiva svakoj tački iz Ω dodeljuje tačku iz $\mathbb{R}$.

Zajednička raspodela (eng. *joint distribution*) je p-dimenzionalna generalizacija kumulativne raspodele (videti A.1, $F_X \equiv P_X$)

$$P_X(x) = P_X(x_1, \ldots, x_p) = P(X_1 \leq x_1, \ldots, X_p \leq x_p).$$

Zajednička funkcija gustine

Zajednička funkcija gustine data je sa (videti A.2, $f_x \equiv p$)

$$p(x) = \frac{\partial^p P(x)}{\partial x_1 \ldots \partial x_p}.$$

Marginalna raspodela

Neka je data zajednička funkcija gustine slučajnih promenljivih $X_1, \ldots, X_p$. Tada je zajednička funkcija gustine za manji skup

$X_1, \ldots, X_m$ $(m < p)$ data sa

$$p(x_1, \ldots, x_m) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} p(x_1, \ldots, x_p) dx_{m+1} \ldots dx_p.$$

Ovo se ponekad zove i marginalna raspodela $X_1, \ldots, X_m$, iako je mnogo češća upotreba za jednu gustinu

$$p(x_i) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} p(x_1, \ldots, x_p) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_p.$$

Očekivani vektor

Očekivani vektor slučajne promenljive x je dat sa (videti [40])

$$m = E[X] = \int xp(x)dx,$$

gde dx označava $dx_1 \ldots dx_p$, a integral je nad celim prostorom, dok $E[\cdot]$ označava očekivanje. Ako posmatramo samo i-tu komponentu, važi

$$\begin{aligned} m_i = E[X_i] &= \int \ldots \int x_i p(x_1, \ldots, x_p) dx_1 \ldots dx_p \\ &= \int x_i p(x) dx \\ &= \int_{-\infty}^{\infty} x_i p(x_i) dx_i. \end{aligned}$$

Kovarijansa

Kovarijansa C_{ij} dve slučajne promenljive X_i i X_j je data sa

$$C_{ij} = E[(X_j - E[X_j])(X_i - E[X_i])]$$

odnosno

$$C_{ij} = E[X_i X_j] - E[X_i]E[X_j]$$

gde je $E[X_i X_j]$ autokorelacija. Matrica čija je $ij-$ta komponenta C_{ij} zove se **kovarijansna matrica**

$$\mathbf{C} = E[(\mathbf{X} - \mathbf{m})(\mathbf{X} - \mathbf{m})^T].$$

Slučajne promenljive X_i i X_j nisu u korelaciji ako je $C_{ij} = 0$, što implicira

$$E[X_i X_j] = E[X_i]E[X_j].$$

Dakle, dva vektora nisu u korelaciji ako važi

$$E[\mathbf{XY}] = E[\mathbf{X}]E[\mathbf{Y}].$$

Za dve slučajne promenljive X_i i X_j kažemo da su nezavisne ako važi

$$p(X_i, X_j) = p(X_i)p(X_j).$$

Dakle, ako su $X_1, \ldots, X_p$ nezavisne tada je

$$p(X_1, \ldots, X_p) = p(X_1) \ldots p(X_p).$$

Uslovna rapodela

Uslovna raspodela, $P_X(x|A)$, slučajne promenljive X događaja A je definisana kao uslovna verovatnoća događaja $\{X \leq x\}$

$$P(x|A) = \frac{P(\{X \leq x\}, A)}{P(A)}$$

pri čemu je $P(\infty|A) = 1$, $P(-\infty|A) = 0$. **Uslovna raspodela** $p(x|A)$ je izvod

$$p(x|A) = \frac{dP}{dx} = \lim_{\Delta x \to 0} \frac{P(x \leq X \leq x + \Delta x|A)}{\Delta x}$$

Kada A.4 prilagodimo za neprekidan slučaj dobijamo

$$p(x) = \sum_{i=1}^{N} p(x|A_i)$$

gde je $p(x)$ *mixture* gustina. Ako u Bajesovoj teoremi stavimo $B = X \leq x$, dobijamo neprekidan slučaj za Bajesovu teoremu

$$p(x|A) = \frac{p(A|x)p(x)}{p(A)} = \frac{p(A|X = x)p(x)}{\int_{-\infty}^{\infty} p(A|X = x)p(x)dx}$$

Uslovna gustina x ako je opserviran vektor Y koji uzima određenu vrednost y je data sa

$$p(x|y) = \frac{p(x,y)}{p(y)} \tag{A.6}$$

gde je $p(x,y)$ zajednička gustina od X i Y, a $p(y)$ je marginalna gustina

$$p(y) = \int p(x,y)dx. \tag{A.7}$$

Koristeći prethodna dva izraza dobijamo još jednu formu Bajesove teoreme

$$\begin{aligned} p(x|y) &= \frac{p(y|x)p(x)}{p(y)} \\ &= \frac{p(y|x)p(x)}{\int p(y|x)p(x)dx} \end{aligned}$$

Generalizacijom A.6 dobijamo uslovnu raspodelu za slučajne promenljive $X_{k+1}, \ldots, X_p$ ako su date $X_1, \ldots, X_k$

$$p(x_{k+1}, \ldots, x_p|x_1, \ldots, x_k) = \frac{p(x_1, \ldots, x_p)}{p(x_1, \ldots, x_k)}. \tag{A.8}$$

Ovo vodi **pravilu lanca** (eng. *chain rule*)

$$p(x_1, \ldots, x_p) = p(x_p|x_1, \ldots, x_{p-1})p(x_{p-1}|x_1, \ldots, x_{p-2}) \ldots p(x_2|x_1), p(x_1).$$

A.6 i A.8 nam omogućavaju da "neželjene" promenljive u uslovnoj gustini zamenimo. Primeri tih zamena su

$$\begin{aligned} p(a|l,m,n) &= \int p(a,b,c|l,m,n)dbdc \\ p(a,b,c|m) &= \int p(a,b,c|l,m,n)p(l,n|m)dldn. \end{aligned} \tag{A.9}$$

Posmatrajmo prvi primer i desnu stranu jednakosti. S' leve strane znaka | smo dodali promenljive b, c i zato se integrali po b, c.
Posmatrajmo drugi primer i desnu stranu jednakosti. Sad smo sa desne strane znaka | dodali l, n zato se $p(a,b,c|l,m,n)$ množi sa uslovnom gustinom $p(l,n|m)$ i integrali po l, n.

Standardna normalna raspodela

Standardna normalna gustina slučajne promenljive X koja ima za centorid nulu, a za varijansu jedan, je sledećeg oblika

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp -\frac{x^2}{2}, -\infty < x < \infty.$$

Raspodela je data sa

$$P(X) = \int_{-\infty}^{X} p(x)dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{X} \exp\{-\frac{1}{2}x^2\}dx.$$

Neka je $Y = \mu + \sigma X$, gde je X slučajna promenljiva. Funkcija gustine Y data je sa

$$p(y) = \frac{1}{\sqrt{2\sigma^2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right]$$

gde je μ centorida, a σ^2 varijansa. Dakle, $Y \sim N(\mu, \sigma^2)$

Multivarijantna normalna raspodela

Ako su $X_1, X_2, \ldots, X_p$ nezavisne slučajne promenljive koje su identično raspodeljene sa standardnom normalnom raspodelom, tada je zajednička gustina data sa

$$p(x_1, x_2, \ldots, x_p) = \prod_{i=1}^{p} p(x_i) = \frac{1}{(2\pi)^{p/2}} \exp\{-\frac{1}{2}\sum_{i=1}^{p} x^2\}.$$

Koristeći transformaciju $Y = AX + \mu$ dobijamo gustinu za Y

$$p(y) = \frac{1}{(2\pi)^{p/2}|A|} \exp\left\{-\frac{1}{2}(y-\mu)^T(A^{-1})^T A^{-1}(y-\mu)\right\}. \quad \text{(A.10)}$$

Kako je Σ kovarijansna matrica od Y, vazi

$$\Sigma = E[(Y-\mu)(Y-\mu)^T] = AA^T.$$

Izraz A.10 se obično zapisuje na sledeći način

$$p(y) = N(y|\mu, \Sigma) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(y-\mu)^T\Sigma^{-1}(y-\mu)\right\}. \quad \text{(A.11)}$$

Ovo je **multivarijantna normalna raspodela**.

Multinominalna raspodela

Multinominalna raspodela je generalizacija **binomne raspodel**e. Neka je X diskretna slučajna promenljiva i neka ona ima binomnu raspodelu sa parametrima n i p; to pišemo $X \sim B(x|n,p)$. Binomna raspodela se primenjuje samo u slučaju binomnog eksperimenta (eksperiment se izvodi n puta, imamo samo mogućnost uspeh/neuspeh, verovatnoća uspeha je p, a neuspeha $1-p$, realizacije su nezavisne). Binomna raspodela je zadata izrazom

$$p(x) = \left(\begin{array}{c} n \\ x \end{array} \right) p^x (1-p)^{n-x}, \tag{A.12}$$

gde je $p \in (0,1)$, $n = 1, 2, \ldots$, $x = 0, 1, \ldots, n$.

Ako imamo pak više mogućih ishoda, a ne samo uspeh i neuspeh koristi se multinominalna raspodela. Pretpostavimo da imamo eksperiment sa kuglicama različite boje. Neka ima ukupno n kuglica i neka ima k različitih boja kuglica. Označimo sa X_i promenljivu koja označava da je izvučena kuglica boje i, a sa p_i verovatnoću izvlačenja te kuglice. Izraz za **multinominalnu raspodelu** $Mu_k(x|p,n)$ je

$$p(x) = \frac{n!}{\prod_{i=1}^{k} x_i!} \prod_{i=1}^{k} p_i^{x_i}$$

gde je $p_i \in (0,1)$, $\sum_{i=1}^{k} p_i = 1$,$\sum_{i=1}^{k} x_i = n$, $x_i = 0, 1, \ldots, n$.

Prilog B

Linearna algebra

U ovom prilogu ćemo prezentovati osnovne pojmove iz linearne algebre koji su neophodni za razumevanje ove teze (videti npr. [40]).

Neka je data matrica A, koja je dimenzija $n \times m$. Označimo sa a_{ij} element i-te vrste i j-te kolone. Transponovana matrica matrice A se označava sa A^T. Važi

$$(AB)^T = B^T A^T.$$

Kvadratna matrica je **simetrična** ako važi $a_{ij} = a_{ji}$ za svako i, j.

Trag kvadratne matrice A označavamo sa TrA. Važi

$$TrA = \sum_{i=1}^{n} a_{ii}$$

i $TrAB = TrBA$, pri čemu ako je AB kvadratna nije neophodno da A i B budu kvadratne.

Determinantu matrice A označavamo sa $|A|$. Važi

$$|A| = \sum_{j=1}^{n} a_{ij} A_{ij}$$

gde $i = 1, \ldots, n$ i A_{ij} je **kofaktor**. Kofaktor je determinanta matrice koja se dobije kada se uklone i-ta vrsta i j-ta kolona matrice A, pomnožena sa $(-1)^{i+j}$. Matrica kofaktora C ($c_{ij} = A_{ij}$) zove se **adjungovana matrica** matrice A. Ako su A i B kvadratne matrice

istog reda važi $|AB| = |A||B|$.

Inverzna matrica A^{-1} matrice A je jedinstvena matrica za koju važi

$$A^{-1}A = AA^{-1} = I$$

gde je I jedinična matrica. Ako matrica ima inverznu matricu kažemo da je **nesingularna**. Ako za datu matricu ne postoji inverzna matrica kažemo da je **singularna** i dalje je det $A = 0$. Važi: $(A^T)^{-1} = (A^{-1})^T$, $(AB)^{-1} = B^{-1}A^{-1}$, tj. ako je A simetrična onda je i A^{-1} simetrična.

Skup k vektora je **linearno zavisan** ako postoji skup skalara $c_1, \ldots, c_k$ koji nisu svi jednaki nuli, tako da važi

$$c_1x_1 + \ldots + c_kx_k = 0.$$

Ako je nemoguće naći takvih k skalara $c_1, \ldots, c_k$, onda kažemo da su vektori **linearno nezavisni**. **Rang matrice** je, po teoremi, maksimalan broj linearno nezavisnih vrsta (ili ekvivalentno maksimalan broj linearno nezavisnih kolona). Matrica dimenzije $n \times n$ je **potpunog ranga** ako joj je rang jednak n. U ovom slučaju determinanta matrice je različita od nule, tj. matrica je regularna. Ako je A dimenzije $m \times n$ važi $rang(A) \leq \min(m, n)$ i

$$rang(A) = rang(A^T) = rang(A^TA) = rang(AA^T).$$

Kvadratna matrica je **ortogonalna** ako važi

$$A^TA = AA^T = I$$

tada važi da su kolone i vrste matrice A **ortonormirane** ($x^Ty = 0$, $x^Tx = 1$, $y^Ty = 1$ za dve različite kolone x i y). Ortogonalna matrica je nesingularna. Inverzna matrica ortogonalne matrice je jednaka njenoj transponovanoj: $A^{-1} = A^T$.

Kvadratna matrica je **pozitivno definitna** ako je kvadratna forma $x^TAx > 0$ za svako $x \neq 0$. Matrica je **pozitivno semidefinitna** ako je $x^TAx \geq 0$ za svako $x \neq 0$. Pozitivno definitne matrice su potpunog ranga.

Karakteristične vrednosti kvadratne matrice A koja je dimenzije $p \times p$ su rešenja **karakteristične jednačine**

$$\det A - \lambda I = 0.$$

Ova karakteristična jednačina je polinom stepena p po λ, što znači da ima p rešenja. Označimo ih sa $\lambda_1, \ldots, \lambda_p$. Ona ne moraju biti različita i mogu biti i realna i kompleksna. Svakoj karakterističnoj vrednosti λ_i dodeljujemo **karakteristični vektor** u_i za koji važi

$$Au_i = \lambda_i u_i.$$

Ovi karakteristični vektori nisu jedinstveni, jer kada u_i pomnožimo bilo kojim skalarom novodobijeni vektor opet gornju jednakost pretvara u identitet.

Zato karakteristične vektore obično normalizujemo $u_i^T u_i = 1$.

Navedimo neke osobine karakterističnih vrednosti i karakterističnih vektora:

- Proizvod karakterističnih vrednosti jednak je determinanti matrice, tj. $\prod_{i=1}^{p} \lambda_i = |A|$.
- Suma karakterističnih vrednosti jednaka je tragu matrice, tj. $\sum_{i=1}^{p} \lambda_i = TrA$.
- Ako je A realna simetrična matrica, karakteristični vektori i karaktristične vrednosti su sve realne.
- Ako je A pozitivno definitna, sve karateristične vrednosti su pozitivne.
- Ako je A pozitivno semidefinitna matrica ranga m, onda ona ima m ne-nula karakterističnih vrednosti i $p-m$ karakterističnih vrednosti jednakih nuli.
- Svaka realna simetrična matrica ima skup ortonormiranih karakterističnih vektora. Matrica U, čije su kolone karakteristični vektori realne simetrične matrice ($U = [u_1, \ldots, u_p]$), je ortogonalna.

Važi $U^TU = UU^T = I$, $U^TAU = \Lambda$ gde je $\Lambda = diag(\lambda_1, \ldots, \lambda_p)$. Možemo pisati i

$$A = U\Lambda U^T = \sum_{i=1}^{p} \lambda_i u_i u_i^T.$$

Ako je A pozitivno definitna važi $A^{-1} = U\Lambda^{-1}U^T$, gde je $\Lambda^{-1} = diag(1/\lambda_1, \ldots, 1/\lambda_p)$.

Uopštena simetrična karakteristična jednačina je oblika

$$Au = \lambda Bu$$

gde su A i B realne simetrične matrice. Ako je B pozitivno definitna, tada gornja jednačina ima p karakterističnih vektora, $(u_i, \ldots, u_p)$, koji su ortonormirani u odnosu na B. Dakle

$$u_i^T B u_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

i

$$u_i^T A u_j = \begin{cases} 0 & i \neq j \\ \lambda_j & i = j \end{cases} .$$

Prethodno se može zapisati i u obliku

$$U^TBU = I$$

i

$$U^TAU = \Lambda$$

gde je $U = [u_1, \ldots, u_p]$, $\Lambda = diag(\lambda_1, \ldots, \lambda_p)$.

U mnogim problemima prepoznavanja oblika potrebno je izvršiti minimizaciju kvadratne greške. Uopšteni problem ovog tipa se rešava dekompozicijom matrice. Neka je A matrica dimenzije $m \times n$. Ona se može zapisati na sledeći način

$$A = U\Sigma V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T,$$

gde je r rang matrice A, U je $m \times r$ matrica čije su kolone $u_1, \ldots, u_r$ levi singularni vektori, $U^TU = I_r$ ($r \times r$ jedinična matrica), V je $n \times r$

matrica sa kolonama $v_1, \ldots, v_r$ desnih singularnih vektora, $V^T V = I_r$, $\Sigma = diag(\sigma_1, \ldots, \sigma_r)$ matrica singularnih vrednosti σ_i.

Singularne vrednosti od A su kvadratni koreni ne-nula karakterističnih vrednosti os AA^T ili $A^T A$. Pseudo-inverzija ili generalizovana inverzija je $n \times m$ matrica $A^\dagger$

$$A^\dagger = V\Sigma^{-1}U^T = \sum_{i=1}^{r} \frac{1}{\sigma_i} v_i u_i^T$$

i rešenje x koje minimizuje kvadratnu grešku $\|Ax - b\|^2$ je dato sa

$$x = A^\dagger b.$$

Ako je rang od A manji od n, tada ne postoji jedinstveno x i dekompozicija na singularne vrednosti daje rešenje sa minimalnom normom.

Pseudo-inverzija ima sledeće osobine:

$$AA^\dagger A = A, \qquad (AA^\dagger)^T = AA^\dagger, \qquad A^\dagger AA^\dagger = A^\dagger, \qquad (A^\dagger A)^T = A^\dagger A.$$

Sada ćemo uvesti i neke osobine izvoda. Važi

$$\frac{\partial}{\partial x} = \left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_p} \right)^T.$$

Prema tome, izvod skalarne funkcije f po x je

$$\frac{\partial f}{\partial x} = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_p} \right)^T.$$

Slično, izvod skalarne funkcije f po matrici A se označava sa $\frac{\partial f}{\partial A}$ i važi

$$\left[\frac{\partial f}{\partial A} \right]_{ij} = \frac{\partial f}{\partial a_{ij}}.$$

Važi

$$\frac{\partial |A|}{\partial A} = |A|(A^{-1})^T$$

ako A^{-1} postoji. Za simetričnu matricu važi

$$\frac{\partial}{\partial x} x^T Ax = 2Ax.$$

Literatura

[1] A. Antić, B. Popović, L. Krstanović, R. Obradović, M. Milošević (2017). Novel Texture-Based Descriptors for Tool Wear Condition Monitoring, in: Mechanical Systems and Signal Processing, ISSN: 0888-3270, Elsevier, Vol. 98:1-15, doi.org/10.1016/j.ymssp.2017.04.030

[2] P.Belhumeur, J.Hepanha, D.Kriegman (1997). Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Analysis and Machine Intelligence, 19 (7): 711-720.

[3] M. Belkin, P. Niyogi (2001). Laplacian eigenmaps and spectral techniques for embedding and clustering. 14th International Conference on Neural Information Processing Systems: Natural and Synthetic (NIPS), 585-591.

[4] C. M. Bishop (2006). Pattern recognition and machine learning, Springer, Verlag-New York.

[5] Vandenberghe, S. Boyd, Shao-Po Wu (1998). Determinant maximaization with linear matrix inequality constraints, SIAM Journal on Matrix Analysis and Applications, 19(2):499-533.

[6] R. L. Burden, J. D. Faires (2010). Numerical Analysis.(9th ed.). London, UK.

[7] T. Cover, J.Thomas (1991). Elements of information theory.(2nd ed.). New York, USA: Wiley Series in Telecommunications, John Wiley and Sons.

[8] K. J. Dana, B. van Ginneken, S. K. Nayar, and J. J. Koenderink (1999). Reflectance and texture of real-world surfaces. ACM Transactions on Graphics (TOG), 18(1):1-34.

[9] J.L. Durrieu, J.P. Thiran, F. Kelly (2012). Lower and upper bounds for approximation of the Kullback-Leibler divergence between Gaussian mixture models. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 4833 - 4836.

[10] A. Goh, R. Vidal (2008). Clustering and Dimensionality Reduction on Riemannian Manifolds. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1-7.

[11] J. Goldberger S. Gordon H. Greenspan (2003). An Efficient Image Similarity Measure based on Approximations of KL-Divergence Between Two Gaussian Mixtures. Ninth IEEE International Conference on Computer Vision, 1: 487 - 493.

[12] J. Goldberger, H. Aronowitz (2005). A Distance measure Between GMMs Based on the Unscented Transform and its Application to Speaker Recognition. INTERSPEECH, 1985-1988.

[13] Z. Guo, L. Zhang, and D. Zhang (2010). A completed modeling of local binary pattern operator for texture classification. TIP, 19(6):1657-1663.

[14] T. Hastie, R. Tibshirani, J. Friedman (2009). The Elements of Statistial Learning. Springer, Verlag-New York.

[15] E. Hayman, B. Caputo, M. Fritz, and J.-O. Eklundh (2004). On the significance of real-world conditions for material classification. In European Conference on Computer Vision (ECCV), 253-256.

[16] X.He, P.Niyogi, Locality preserving projections (2003). Proceedings of Conference on Advances in Neural Information Processing Systems 16 (NIPS), 153-160.

[17] X. He, D. Cai, S. Yan and H.-J. Zhangr (2005). Neighborhood Preserving Embedding, ICCV 2005. Computer Vision, Tenth IEEE International Conference, Vol. 2: 1208-1213.

[18] J. R. Hershey, P. A. Olsen (2007). Aproximating the Kullback Leibler divergence between Gaussian mixture models. IEEE

International Conference on Acoustics, Speech and Signal Processing, ICASSP, 4: IV-317 - IV-320.

[19] S. Jayasumana, R. Hartley, M. Salzmann, H. Li, and M. Harandi (2013). Kernel Methods on the Riemannian Manifold of Symmetric Positive Definite Matrices. IEEE Computer Vision and Pattern Recognition (CVPR), 73-80.

[20] H. Ji, X. Yang, H. Ling, and Y. Xu (2013). Wavelet domain multifractal analysis for static and dynamic texture classification. IEEE Transactions on Image Processing (TIP), 22(1):286-299.

[21] S. Julier, J. K. Uhlmann (1996). A general method for approximating nonlinear transformations of probability distributions, Technical report, RRG, Dept. of Engineering Science, University of Oxford.

[22] G. J. Klir, B. Yuan (1995). Fuzzy sets and fuzzy logic: Theory and Applicaions. London, UK.

[23] L. Krstanović, N. Ralević, V. Zlokolica, R. Obradović, D. Mišković, M. Janev, B. Popović (2016). GMMs similarity measure based on LPP-like projection of the parameter space, Expert Systems with Applications ISSN: 0957-4174, Elsevier, Vol 66: 136-148, DOI:10.1016/j.eswa.2016.09.014

[24] S. Kullback (1968). Information Theory and Statistics. Dover Publications Inc.(new ed.), Mineola, New York.

[25] S. Lazebnik, C. Schmid, J. Ponce (2005). A Sparse Texture Representation Using Local Affine Regions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27 (8): 1265-1278.

[26] H.Li, T.Jiang, K.Zhang (2006). Efficient and robust feature extraction by maximum margin criterion, IEEE Trans. Neural Networks, 17(1): 157-165.

[27] P. Li, Q. Wang, L. Zhang (2013). A Novel Earth Mover's Distance Methodology for Image Matching with Gaussian Mixture Models, IEEE International Conference on Computer Vision, ICCV, pp. 1689-1696.

[28] H. Ling, K. Okada (2007). An efficient Earth Movers Distance algorithm for robust histogram comparison. IEEE Trans on Pattern Anal. and Mach. Intell. (PAMI), 29(5):840-853.

[29] J.Liu, S.Chen, X.Tan, D.Zhang (2007). Comments on efficient and robust feature extraction by maximum margin criterion, IEEE Trans. Neural Networks 18(6): 1862-1864.

[30] L. Liu, P. Fieguth, G. Kuang, and H. Zha(2012). Sorted random projections for robust texture classification. Pattern Recognition, Vol. 45(6):2405–2418.

[31] M. Lovric, M. Min-Oo, E. A. Ruh (2000). Multivariate normal distributions parametrized as a Riemannian symmetric space. Journal of Multivariate Analysis (JMVA), 74(1):36-48.

[32] A. Ng (2004). CS229 Lecture Notes, Simon Fraser University

[33] L. Qiao, S. Chen, X. Tan (2010). Sparsity preserving projections with applications to face recognition. Pattern Recognition 43: 331-341.

[34] S.Roweis, L.Saul (2000). Nonlinear dimensionality reduction by locally linear embedding. Science, 290(5500): 2323-2326.

[35] Y. Rubner, C. Tomasi, L. Guibas (2000). The Earth Mover's Distance as a Metric for Image Retrieval. International Journal of Computer Vision, 40(2): 99-121.

[36] R. Sivalingam, D. Boley, V. Morellas, N. Papanikolopoulos (2010). Tensor Sparse Coding for Region Covariances. Computer Vision (ECCV), 722-735.

[37] J.Tenenbaum (1998). Mapping a manifold of perceptual observations. Advances in Neural Information Processing Systems (NIPS), 682-688.

[38] M.Turk, A.Pentland (1991). Eigenfaces for recognition. J. Cognitive Neurosci, 3(1): 71-86.

[39] M. Varma and A. Zisserman(2003). Texture classification: Are filter banks necessary?, IEEE Conference on Computer Vision and Pattern Recognition, 2: 691-698

[40] A. R. Webb (2002). Statistical Pattern Recognition. (2nd ed.). New York, USA: John Wiley and Sons.

[41] Y. Wu, K. Chan, and H. Wang (2003). Texture classification based on finite gaussian mixture model, IEEE, Workshop on Texture Analysis and Synthesis.

[42] Y. Xu, H. Ji, and C. Fermuller (2006). A projective invariant for texture, IEEE Conference on Computer Vision and Pattern Recognition, pp. 1932-1939.

[43] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid (2007). Local features and kernels for classification of texture and object categories: A comprehensive study. IJCV, 73(2):213-238.

[44] Y. Zhang, Z. Jiang, L. S. Davis (2013). Discriminative Tensor Sparse Coding for Image Classification, British Machine Vision Conference.