SVEUČILIŠTE U ZAGREBU

FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

SEMINAR

Fisherova linearna diskriminantna analiza

Ivana Sučić

Voditelj: *Siniša Šegvić*

Zagreb, travanj, 2010

Sadržaj

1. Uvod

Jedan od čestih problema koji se odnosi na primjenu statističkih metoda u raspoznavanju uzoraka naziva se „prokletstvo dimenzionalnosti“. Postupci koji su analitički ili računski praktični u prostorima nižih dimenzija mogu biti potpuno nepraktični u prostorima od 50 ili 100 dimenzija. Upravo zbog toga razvijene su različite tehnike za reduciranje dimenzija.

D-dimenzionalni uzorci mogu se reducirati na jednu dimenziju ako projiciramo d -dimenzionalan uzorak na pravac. Ako pravac odaberemo proizvoljno , projicirani uzorci unatoč tome da li su originalni uzorci razdvojivi, mogu biti zbunjujuće pomiješani. Ideja je pronaći smjer pravca na kojeg će projicirani uzorci biti dobro odvojivi. Upravo to predstavlja cilj klasične diskriminantne analize.

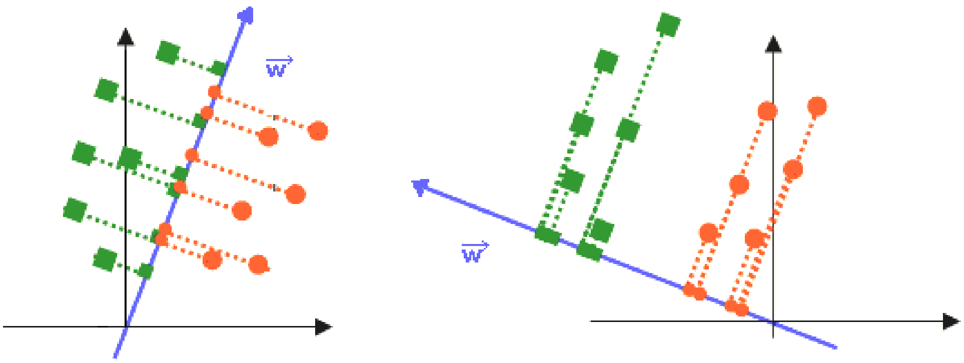
U ovom seminarskom radu obrađena je Fisherova linearna diskriminanta (FLD) koja predstavlja jedan od glavnih pristupa linearnoj klasifikaciji. Njena primjena je raznolika. Jedna od popularnih metoda za raspoznavanje lica naziva se Fisherfaces i temelj joj je upravo spomenuta FLD poznata još kao i linearna diskriminantna analiza (LDA).

Prvo poglavlje rada opisuje FLD za dva razreda, drugo poglavlje bavi se poopćenjem FLD na više razreda odnosno višestruka diskriminantna analiza . Zadnje poglavlje ukratko uspoređuje analizu glavnih komponenti (PCA) s FLD.

2. Fisherova linearna diskriminanta

Fisherova linearna diskrimantna analiza jedan je od glavnih pristupa linearnoj klasifikaciji. Osnovna ideja Fisherove analize je d-dimenzionalan vektor značajki reducirati na jednu dimenziju i tada ga upotrijebiti za klasifikaciju.

Potrebno je pronaći orijentaciju pravca na koji se projiciraju d-dimenzionalni uzorci $\vec{x}_i$ pri čemu je $i = 1,2,...,n$, ali tako da su projicirani uzorci separatibilni. Upravo to predstavlja cilj klasične diskriminantne analize.



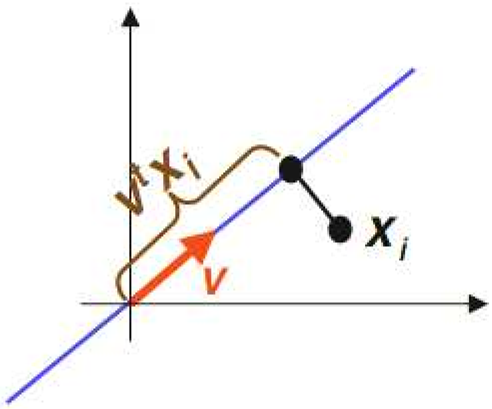
Slika 2.1 Projekcije uzoraka na dva različita pravca. Desna slika prikazuje bolju razdvojenost projekcija crvenih i zelenih točaka nego lijeva slika

Slika 2.1. predstavlja projekcije uzoraka na dva različita pravca. Desna slika prikazuje bolju razdvojenost projekcija crvenih i zelenih točaka

Pretpostavimo da imamo skup od n d-dimenzionalnih uzoraka $x_1,...,x_n$ od čega n_1 uzoraka čini podskup D_1 (označeni s w_1) i n_2 uzoraka podskup D_2 (označeni s w_2). Tvorimo linearnu kombinaciju komponenti $\vec{x}$:

$$y_i = \vec{w}^T \vec{x}_i, i = 1,2,...,n \tag{2.1}$$

Dobivamo odgovarajući skup od n uzoraka $y_1,..., y_n$ koji su podijeljeni u podskupove y_1 i y_2. Ako je $\|w\| = 1$, svaki y_i je projekcija odgovarajućeg vektora $\vec{x}_i$ na pravac (u smjeru $\vec{w}$, slika 2.2). Veličina vektora $\vec{w}$ nema posebno značenje, ona samo skalira y_i. Bitan je smjer vektora $\vec{w}$.



Slika 2.2 Projekcija vektora $\vec{x}_i$ na pravac u smjeru $\vec{w}$

Tražimo $\vec{w}$ za koji se projekcije uzoraka mogu dobro odvojiti. Mjera odvojivosti između projiciranih točaka je razlika srednjih vrijednosti projiciranih uzoraka. Srednja vrijednost zadanih uzoraka definirana je izrazom:

$$m_i = \frac{1}{n_i} \sum_{x \in D_i} x \qquad (2.2)$$

Srednja vrijednost projiciranih uzoraka:

$$\tilde{m}_i = \frac{1}{n_i} \sum_{y \in Y_i} y \,. \qquad (2.3)$$

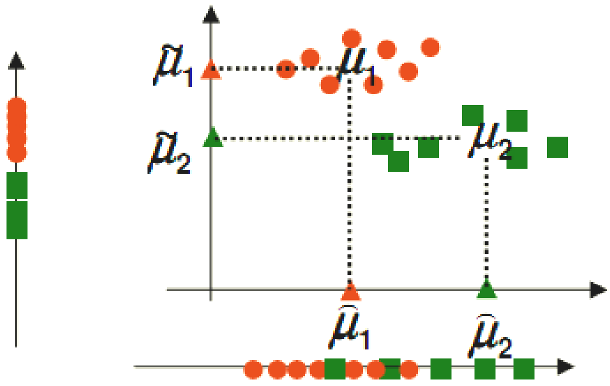
Uvrštavanjem (2.1) i (2.2) u (2.3) dobivamo relaciju:

$$\tilde{m}_i = \frac{1}{n_i} \sum_{y \in Y_i} w^T x = w^T m_i \,. \qquad (2.4)$$

Slijedi da je razlika između srednjih vrijednosti projiciranih uzoraka:

$$\left|\tilde{m}_1 - \tilde{m}_2\right| = \left|w^T (m_1 - m_2)\right| . \qquad (2.5)$$

Da li je udaljenost srednjih vrijednosti projiciranih uzoraka „dobra mjera“ da se odijele projicirani uzorci različitih razreda?



Slika 2.3 Utjecaj udaljenosti srednjih projiciranih uzoraka na klasifikaciju uzoraka

Na slici 2.3. vidimo da vertikalna os bolje razdvaja uzorke unatoč manjoj udaljenosti srednjih projiciranih uzoraka naspram udaljenosti srednjih projiciranih uzoraka na horizontalnu os. Razlog tomu jest činjenica da definirana mjera odvojivosti ne uračunava varijancu unutar razreda zadanih uzoraka.

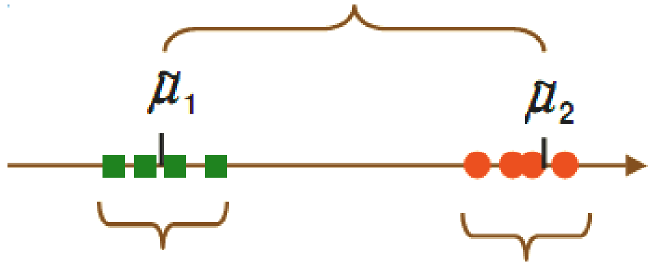
Kako bismo to izbjegli, definirat ćemo mjeru raspršenosti za projicirane uzorke:

$$\tilde{s}_i^2 = \sum_{y \in Y_i} (y - \tilde{m}_i)^2 \ . \tag{2.6}$$

$(1/n)(\tilde{s}_1^2 + \tilde{s}_2^2)$ je procjena varijance podataka, a $\tilde{s}_1^2 + \tilde{s}_2^2$ je ukupna mjera raspršenosti podataka unutar razreda. Fisherova linearna diskriminanta određuje da linearna funkcija $\vec{w}^T \vec{x}_i$ za koju je kriterijska funkcija $J(\vec{w})$ maksimalna vodi najboljem razdvajanju između projiciranih skupova.

$$J(\vec{w}) = \frac{|\tilde{m}_1 - \tilde{m}_2|^2}{\tilde{s}_1^2 + \tilde{s}_2^2} \tag{2.7}$$

Kako bi gore navedeni izraz bio maksimalan, potrebno je imati što veću razliku srednjih vrijednosti i što manju raspršenost unutar projiciranih uzoraka istog razreda(slika 2.4).



Slika 2.4 Da bi razdvajanje projiciranih uzoraka bilo najbolje, potrebno je imati što veću razliku srednjih vrijednosti različitih razreda i što manju raspršenost unutar projiciranih uzoraka istog razreda

Potrebno je izraziti kriterijsku funkciju $J(\vec{w})$ kao eksplicitnu funkciju od $\vec{w}$.

Definirajmo matricu S_i:

$$S_i = \sum_{x \in D_i} (x - m_i)(x - m_i)^T. \tag{2.8}$$

I matricu S_w:

$$S_w = S_1 + S_2. \tag{2.9}$$

Uvrstimo (2.1) i (2.4) u izraz (2.6):

$$\begin{aligned} \tilde{s}_i^2 &= \sum_{x \in D_i} (w^T x - w^T m_i)^2 \\ &= \sum_{x \in D_i} w^T (x - m_i)(x - m_i)^T w \\ &= w^T S_i w \end{aligned} \tag{2.10}$$

Ukupna mjera raspršenosti je:

$$\tilde{s}_1^2 + \tilde{s}_2^2 = w^T S_i w. \tag{2.11}$$

Slično razlika srednjih vrijednosti projiciranih uzoraka je:

$$\begin{aligned} (\tilde{m}_1 - \tilde{m}_2)^2 &= (w^T m_1 - w^T m_2)^2 \\ &= w^T (m_1 - m_2)(m_1 - m_2)^T w \\ &= w^T S_B w \end{aligned} \tag{2.12}$$

Gdje je

$$S_B = (m_1 - m_2)(m_1 - m_2)^T \tag{2.13}$$

S_w je matrica raspršenosti unutar razreda, a S_B matrica raspršenosti između razreda. Rang matrice S_B za slučaj dva razreda je jedan.

Uvrštavanjem (2.9) i (2.13) u (2.7) dolazimo do kriterijske funkcije koja ovisi samo o parametru $\vec{w}$:

$$J\left(\vec{w}\right) = \frac{\vec{w}^T S_B \vec{w}}{\vec{w}^T S_w \vec{w}}. \tag{2.14}$$

Preostalo je pronaći maksimum funkcije $J\left(\vec{w}\right)$. Do njega se dolazi deriviranjem funkcije i izjednačavanjem sa nulom:

$$\begin{aligned}
&\frac{\partial J(\vec{w})}{\partial w} = 0 \\
&\frac{\partial J(\vec{w})}{\partial w} = \frac{\partial}{\partial \vec{w}}\left(\frac{\vec{w}^T S_B \vec{w}}{\vec{w}^T S_w \vec{w}}\right) = 0 \\
&\frac{(\vec{w}^T S_w \vec{w})\frac{\partial}{\partial w}(\vec{w}^T S_B \vec{w}) - (\vec{w}^T S_B \vec{w})\frac{\partial}{\partial w}(\vec{w}^T S_w \vec{w})}{(\vec{w}^T S_w \vec{w})^2} = 0 \\
&(\vec{w}^T S_w \vec{w}) \cdot 2S_B \vec{w} - (\vec{w}^T S_B \vec{w}) \cdot 2S_w \vec{w} = 0 \\
&S_w \vec{w}(\vec{w}^T S_B \vec{w})(\vec{w}^T S_w \vec{w})^{-1} = S_B \vec{w} \\
&\lambda = (\vec{w}^T S_B \vec{w})(\vec{w}^T S_w \vec{w})^{-1}
\end{aligned} \tag{2.15}$$

$$\lambda S_w w = S_B w \tag{2.16}$$

Dobiveni izraz $\lambda S_w w = S_B w$ naziva se generalizirani problem svojstvenih vektora.

Međutim, nije uvijek potrebno pristupati ovom problemu tako da tražimo svojstvene vrijednosti i svojstvene vektore za $S_w^{-1} S_B$ zato što je $S_B \vec{w}$ uvijek usmjeren kao $\left(\vec{m_1} - \vec{m_2}\right)$. Budući da nas faktor skaliranja za vektor $\vec{w}$ ne zanima (zanima nas samo smjer), možemo napisati rješenje za w:

$$w = S_w^{-1}(m_1 - m_2)\,. \tag{2.17}$$

Dobiveni izraz za vektor $\vec{w}$ nam određuje linearnu funkciju koja maksimizira omjer između udaljenosti srednjih vrijednosti projiciranih uzoraka i raspršenja unutar razreda. Drugim riječima, pronađen je smjer vektora $\vec{w}$ za kojeg se može vršiti najbolja klasifikacija.

3. FLD za više razreda

Fisherova linearna diskriminanta se može poopćiti odnosno primijeniti na problem više razreda. Imamo c razreda dimenzionalnosti $D_1,...,D_c$ ililti $w_1,...,w_c$. Generalizirana FLD $c-1$ diskriminantnih funkcija tj. projekcija d-dimenzinalnog prostora u $(c-1)$-dimenzionalni prostor $(d \geq c)$.

Generalizirana matrica rapršenosti unutar razreda je:

$$S_w = \sum_{i=1}^{c} S_i \tag{3.1}$$

gdje je

$$S_i = \sum_{x \in D_i} (\vec{x} - \vec{m}_i)(\vec{x} - \vec{m}_i)^T \tag{3.2}$$

i

$$\vec{m}_i = \frac{1}{n_i} \sum_{x \in D_i} \vec{x} \quad . \tag{3.3}$$

Matrica raspršenosti između razreda S_B ne dobiva se tako očigledno. Ukupan vektor srednje vrijednosti $\vec{m}$ jest:

$$\vec{m} = \frac{1}{n} \sum_{x} \vec{x} = \frac{1}{n} \sum_{i=1}^{c} n_i \vec{m}_i \tag{3.4}$$

gdje je n_i broj uzoraka u razredu $D_i(w_i)$, a $\vec{m}_i$ vektor srednjih vrijednosti iz razreda w_i. Ukupna matrica raspršenosti S_T je:

$$S_T = \sum_{\vec{x}} (\vec{x} - \vec{m})(\vec{x} - \vec{m})^T$$

$$S_T = \sum_{i=1}^{c} \sum_{\vec{x} \in D_i} (\vec{x} - \vec{m}_i + \vec{m}_i - \vec{m})(\vec{x} - \vec{m}_i + \vec{m}_i - \vec{m})^T$$

$$S_T = \sum_{i=1}^{c} \sum_{\vec{x} \in D_i} (\vec{x} - \vec{m}_i)(\vec{x} - \vec{m}_i)^T + \sum_{i=1}^{c} \sum_{\vec{x} \in D_i} (\vec{m}_i - \vec{m})(\vec{m}_i - \vec{m})^T \tag{3.5}$$

$$+ \sum_{i=1}^{c} \sum_{\vec{x} \in D_i} (\vec{m}_i - \vec{m})(\vec{x} - \vec{m}_i)^T + \sum_{i=1}^{c} \sum_{\vec{x} \in D_i} (\vec{m}_i - \vec{m})^T (\vec{x} - \vec{m}_i)$$

$$S_T = S_w + \sum_{i=1}^{c} n_i (\vec{m}_i - \vec{m})(\vec{m}_i - \vec{m})^T$$

Drugi član $\sum_{i=1}^{c} n_i (\vec{m}_i - \vec{m})(\vec{m}_i - \vec{m})^T$ je poopćena matrica raspršenosti između razreda S_B:

$$S_B = \sum_{i=1}^{c} n_i (\vec{m}_i - \vec{m})(\vec{m}_i - \vec{m}) \tag{3.6}$$

$$S_T = S_W + S_B \tag{3.7}$$

Projekcija iz d -dimenzionalnog prostora u $(c-1)$ -dimenzionalni prostor postiže se uporabom $(c-1)$ diskriminantnih funkcija te je dana izrazom $y_i = \vec{w}_i^T \vec{x}$, $i = 1,2,...,c-1$. Ako y_i promatramo kao komponentu vektora $\vec{y}$ i težinske faktore $\vec{w}_i$ kao stupce $d \times (c-1)$ matrice W tada je projekcija :

$$\vec{y} = W^T \vec{x} \tag{3.8}$$

Uzorci $\vec{x}_1,...,\vec{x}_n$ projiciraju se u odgovarajući skup uzoraka $\vec{y}_1,...,\vec{y}_n$ koji mogu biti opisani svojim srednjim uzorcima i matricama raspršenosti:

$$\tilde{m}_i = \frac{1}{n_i} \sum_{\vec{y} \in Y_i} \vec{y} \tag{3.9}$$

$$\tilde{m} = \frac{1}{n} \sum_{i=1}^{c} n_i \tilde{m}_i \tag{3.10}$$

$$\tilde{S}_w = \sum_{i=1}^{c} \sum_{\vec{y} \in Y_i} (\vec{y} - \tilde{m}_i)(\vec{y} - \tilde{m}_i)^T \tag{3.11}$$

$$\tilde{S}_B = \sum_{i=1}^{c} n_i (\tilde{m}_i - \tilde{m})(\tilde{m}_i - \tilde{m})^T \quad (3.12)$$

$$\tilde{S}_w = W^T S_w W \quad (3.13)$$

$$\tilde{S}_B = W^T S_B W \quad (3.14)$$

Jednadžbe 3.13 i 3.14 pokazuju kako se matrice raspršenosti unutar i između razreda transformiraju projekcijom u niževdimenzionalni prostor. Tražimo transformacijsku matricu W koja maskimizira omjer raspršenosti između razreda s raspršenošću unutar razreda. Jednostavna skalarna mjera raspršenosti je determinanta matrice raspršenosti.

$$J(W) = \frac{|\tilde{S}_B|}{|\tilde{S}_w|} = \frac{|W^T S_B W|}{|W^T S_w W|} \quad (3.15)$$

Determinanta je produkt svojstvenih vrijednosti matrice. Problem traženja (i nalaženja) pravokutne matrice W koja maksimizira $J(\bullet)$ je težak problem.

Rješenje je slijedeće. Stupci optimalne matrice W su generalizirani svojstveni vektori koji odgovaraju najvećim svojstvenim vrijednostima u :

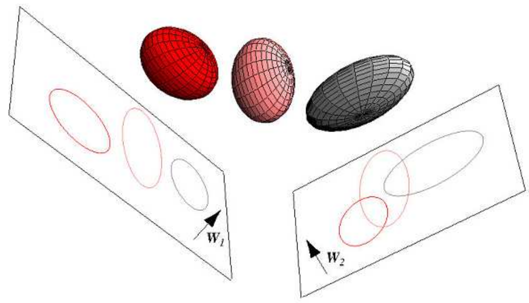
$$S_B \vec{w}_i = \lambda_i S_w \vec{w}_i \ . \quad (3.16)$$

Ako je S_w nesingularna onda se problem može pretvoriti u konvencionalni problem svojstvenih vrijednosti. Međutim, to zahtijeva računanje inverzne matrice S_w ($S_w^{-1} S_B \vec{w} = \lambda \vec{w}$). Umjesto toga možemo naći svojstvene vrijednosti kao korijene karakterističnog polinoma:

$$|S_B - \lambda_i S_w| = 0 \quad (3.17)$$

i zatim riješiti:

$$(S_B - \lambda_i S_w)\vec{w}_i = 0 \ . \quad (3.18)$$



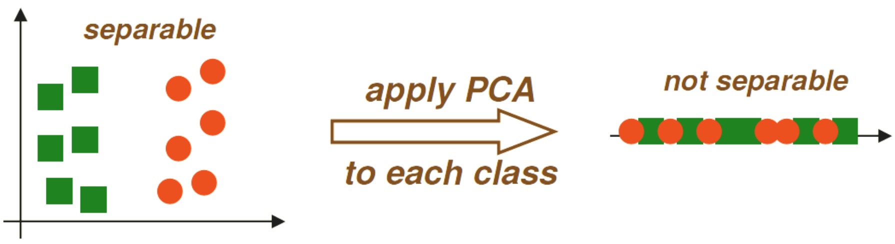
Slika 3.1

Slika 3.1 prikazuje kako se 3-D distribucija projicira na 2-D potprostore koji su opisani normaliziranim vektorima $\vec{w}_1$ i $\vec{w}_2$.

4. Usporedba PCA i FLD

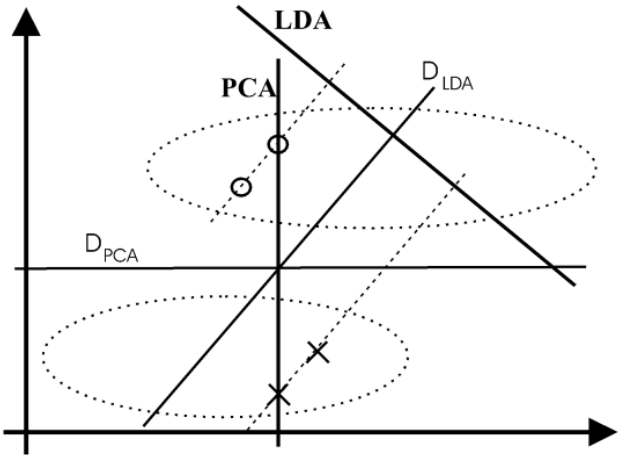
PCA (Principal Component Analysis) ili analiza glavnih komponenti je statistički postupak za reduciranje dimenzije podataka. Ideja je da se pronađe takav podprostor zadane niže dimenzionalnosti koji najbolje prikazuje pojedini razred uzoraka u smislu da pogreška između originalnog uzorka i projekcije uzorka u taj podprostor bude minimalna.

PCA pronalazi podprostor u kojem podatci najviše variraju što je za samu redukciju dobro, ali za klasifikaciju može biti beznačajno. Fisherova diskriminatna analiza također reducira dimenzionalnost prostora uzoraka, no razlika u odnosu na PCA je što nastoji smanjiti dimenzionalnost uzimajući još u obzir i razrede projiciranih uzoraka, tj. nastoji zadržati što bolju razdvojivost razreda u reduciranom prostoru.



Slika 4.1

Iako se intuitivno može zaključiti da je FLD nadmoćnija od PCA, situacija ipak nije uvijek takva. U području raspoznavanja lica, pokazalo se da ako je skup za učenje mali PCA može dati bolje rezultate od FLD. PCA je također manje osjetljiva na različite skupove za učenje [5].



Slika 4.2 Iscrtani krugovi prikazuju ukupan opseg uzoraka dva različita razreda. Samo su po dva uzorka dana u skup za učenje (PCA ili FLD). Klasifikacija dobivena PCA bolja je od one dobivene FLD.

Slika 4.2 prikazuje dva različita razreda. Samo su po dva uzorka dana u skup za učenje (PCA ili FLD). Klasifikacija dobivena PCA bolja je od one dobivene FLD.

Kao najbolja opcija jest korištenje PCA i FLD u kombinaciji. Dakle pomoću PCA reduciramo podatke kako bi proračun svojstvenih vektora bio lakši , a potom primijenimo FLD, daljnja redukcija informacija na $(ukupan\_broj\_razreda - 1)$.

5. Zaključak

FLD ili Fisherova linearna diskriminatna analiza jedan je od glavnih pristupa linearnoj klasifikaciji. Ako zamislimo da uzorci iz $\vec{w}_1$ čine nakupinu, a uzorci iz $\vec{w}_2$ čine drugu nakupinu želimo da projekcije na pravac određen s $\vec{w}$ budu takve da su dobro odvojive. Ideja je dakle pronaći najbolji smjer $\vec{w}$.

FLD moguće je poopćiti na problem više razreda što nazivamo generalizirani FLD. Generalizirana FLD traži transformacijsku matricu W koja maksimizira omjer raspršenosti između razreda s raspršenošću unutar razreda.

Usporedba analize glavnih komponenti (PCA) s FLD pokazala je da optimum čini kombinacija obiju metoda. Dok PCA traži glavne komponente u smjeru najveće varijacije, FLD projicira uzorke u takav prostor u kojem će najbolje razdvojiti klase uzoraka.

6. Literatura

1. Szeliski, Richard. Computer Vision: Algorithms and Applications.Recogniton. Str.642-654.
2. Belhumeur, P.; Hespanha, J. & Kriegman, D. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Transactions on Pattern Analysis and Machine Intelligence. (1997).
3. Chrisopher M. Bishop, Pattern Recognition and Machine Learning.Diskriminant Funtions. Str.181-192.
4. S. Theodorids, K. Koutroumbas. Pattern Recognition.Class separability mesures. 2003,str.174-181.
5. Aleix M. Martinet, Avinash C. Kak. PCA versus LDA. IEEE Transactions on Pattern Analysis and Machine Intelligence.(2001).
6. http://www.csd.uwo.ca/~olga/Courses//CS434a_541a//Lecture8.pdf
7. S. Ribarić, Raspoznavanje uzoraka, predavanje 3, str.34-50.

7. Sažetak

FLD ili Fisherova linearna diskriminatna analiza jedan je od glavnih pristupa linearnoj klasifikaciji. Glavna ideja je pronaći orijentaciju pravca na koji se projiciraju d-dimenzionalni uzorci $\vec{x}_i$ pri čemu je $i = 1,2,...,n$, ali tako da su projicirani uzorci separatibilni. FLD moguće je poopćiti na problem više razreda što nazivamo generalizirani FLD. PCA (Principal Component Analysis) ili analiza glavnih komponenti je statistički postupak za reduciranje dimenzije podataka. FLD kao i PCA nastoji smanjiti dimenzionalnost no u obzir uzima i razrede projiciranih uzoraka. Da bi razdvajanje projiciranih uzoraka bilo najbolje, potrebno je imati što veću razliku srednjih vrijednosti različitih razreda i što manju raspršenost unutar projiciranih uzoraka istog razreda.