# 4. Višestruka diskriminantna analiza (Multiple Discriminant Analysis)

Problem c-razreda:

$D_1, D_2, \cdots, D_c$ odnosno $\omega_1, \omega_2, \cdots, \omega_c$

**Generalizirana FLD** uključuje (c-1) diskriminantnih funkcija:
Projekcija d-dimenzionalnog prostora u (c-1)-dimenzionalni prostor (d ≥ c).

Generalizirana matrica raspršenosti unutar razreda (engl. *within-class scatter matrix*):

$$S_W = \sum_{i=1}^{c} S_i$$

gdje je

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

i

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

Matrica $S_B$ kao generalizirana ne dobiva se tako očigledno.

Ukupan vektor srednje vrijednosti $\vec{m}$

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

gdje je $n_i$ broj uzoraka u razredu $D_i$ odnosno $\omega_i$, a $\vec{m}_i$ vektor srednje vrijednosti vektora iz razreda $\omega_i$.

Ukupna matrica raspršenosti $S_T$

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

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

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

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

Drugi član

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

je poopćena matrica raspršenosti između razreda $S_B$

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

$$S_T=S_W+S_B.$$

Projekcija iz d-dimenzionalnog prostora u (c-1)-dimenzionalni prostor uporabom (c-1) diskriminantnih funkcija

$$y_i=\vec{w}_i^T\vec{x}\,,\, i=1,2,\cdots,c-1.$$

Ako $y_i$ promatramo kao komponentu vektora $\vec{Y}$ i težinske vektore $\vec{w}_i$ kao stupce $d\times(c-1)$ matrice $W$ tada je projekcija:

$$\vec{Y}=W^T\vec{x}\,\cdot$$

Uzorci $\vec{x}_1,\vec{x}_2,\cdots,\vec{x}_n$ projiciraju se u odgovarajući skup uzoraka $\vec{Y}_1,\vec{Y}_2,\cdots,\vec{Y}_n$ koji mogu biti opisani svojim srednjim vektorima i matricama raspršenosti:

$$\widetilde{m}_i=\frac{1}{n_i}\sum_{\vec{y}\in Y_i}\vec{y}$$

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

$$\widetilde{S}_W=\sum_{i=1}^{c}\sum_{\vec{y}\in Y_i}(\vec{y}-\widetilde{m}_i)(\vec{y}-\widetilde{m}_i)^T$$

$$\widetilde{S}_B=\sum_{i=1}^{c}n_i(\widetilde{m}_i-\vec{m})(\widetilde{m}_i-\vec{m})^T$$

$$\widetilde{S}_W=W^TS_WW$$

$$\widetilde{S}_B = W^T S_B W \cdot$$

---

$y_i = \vec{w}_i^T \vec{x}\,, i = 1,2,\cdots,c-1.$**Dodatak 1.**

$$\widetilde{S}_W = \sum_{i=1}^{c} \sum_{\vec{y}\in Y_i} (\vec{y} - \widetilde{m}_i)(\vec{y} - \widetilde{m}_i)^T$$

$$\vec{y} = W^T \vec{x} \qquad \widetilde{m}_i = W^T \vec{m}_i$$

$$\widetilde{S}_W = \sum_{i=1}^{c} \sum_{\vec{y}\in Y_i} (W^T \vec{x} - W^T \vec{m}_i)(W^T \vec{x} - W^T \vec{m}_i)^T$$

$$\widetilde{S}_W = \sum_{i=1}^{c} \sum_{\vec{y}\in Y_i} W^T (\vec{x} - \vec{m}_i)(\vec{x} - \vec{m}_i)^T W$$

$$\widetilde{S}_W = W^T \sum_{i=1}^{c} \sum_{\vec{y}\in Y_i} (\vec{x} - \vec{m}_i)(\vec{x} - \vec{m}_i)^T W$$

$$S_W = \sum_{i=1}^{c} S_i$$

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

$$\widetilde{S}_W = W^T S_W W$$

$$\widetilde{S}_B = W^T S_B W$$

$$\widetilde{S}_W = W^T S_W W$$

Gornje jednadžbe pokazuju kako se tzv. *within-class* i *between-class* matrice raspršenja transformiraju projekcijom u niždimenzionalni prostor.

Tražimo transformacijsku matricu $W$ koja maksimizira omjer *between-class* i *within-class* raspršenosti. Jednostavna skalarna mjera raspršenosti je determinanta matrice raspršenosti:

$$J(W) = \frac{\left|\widetilde{S}_B\right|}{\left|\widetilde{S}_W\right|}.$$

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

**Rješenje:**
Stupci optimalne matrice $W$ su generalizirani svojstveni vektori koji odgovaraju najvećim svojstvenim vrijednostima u

$$S_B \vec{w}_i = \lambda_i S_W \vec{w}_i.$$

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:

$$|S_B - \lambda_i S_W| = 0.$$

---

$$J(W) = \frac{|\tilde{S}_B|}{|\tilde{S}_W|} = \frac{|W^T S_B W|}{|W^T S_W W|}$$

Naći $W$ koja maksimizira $J(W)$.

$$S_B \vec{w}_i = \lambda_i S_W \vec{w}_i$$

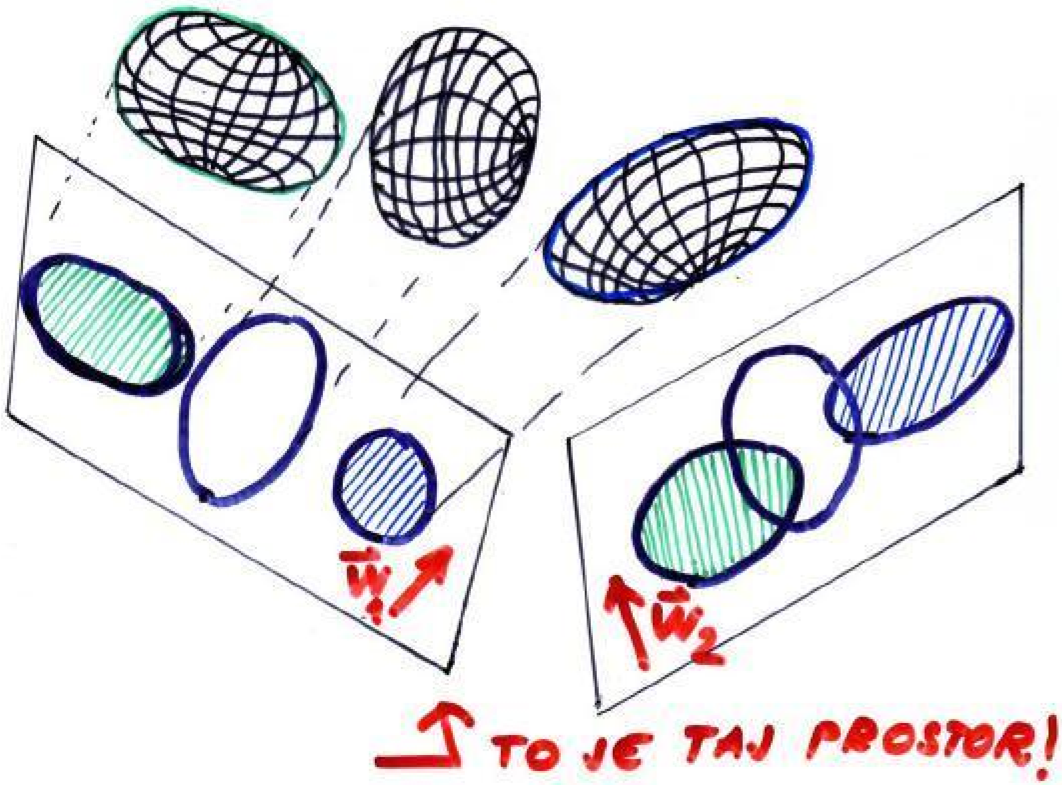
Naći svojstvene vrijednosti kao korijene karakterističnog polinoma:

$$|S_B - \lambda_i S_W| = 0$$

i zatim riješiti:

$$(S_B - \lambda_i S_W)\vec{w}_i = \vec{0}.$$

**Primjer za 3D:**
3D distribucija se projicira na 2D podprostore koji su opisani normalnim vektorima $\vec{w}_1$ i $\vec{w}_2$. MDA traži podprostor u kojem je odvojivost projekcija najveća.




---

## Dodatak 2.

Stupci optimalne (pravokutne) matrice $W$ (koja maksimizira $J(W)$) su generalizirani svojstveni vektori koji odgovaraju najvećim svojstvenim vrijednostima u

$$S_B \vec{w}_i = \lambda_i S_W \vec{w}_i.$$

Ako je $S_W$ nesingularna onda se problem pretvara u konvencionalan problem svojstvenih vrijednosti. Međutim, umjesto računanja inverzne matrice $S_W^{-1}$ mogu se naći svojstvene vrijednosti karakterističnog polinoma

$$|S_B - \lambda_i S_W| = 0$$

i riješiti

$$(S_B - \lambda_i S_W)\vec{w}_i = \vec{0}$$

izravno po $\vec{w}_i$.

$$J(W) = \frac{\left|\widetilde{S}_B\right|}{\left|\widetilde{S}_W\right|} = \frac{|W^T S_B W|}{|W^T S_W W|}$$

Problem nalaženja (pravokutne) matrice $W$ koja maksimizira $J(\cdot)$ je složen. Stupci optimalne matrice $W$ su generalizirani svojstveni vektori koji odgovaraju najvećim svojstvenim vrijednostima u

$$S_B \vec{w}_i = \lambda_i S_W \vec{w}_i.$$

Ako je $S_W$ nesingularna onda se problem može transformirati u konvencionalan problem svojstvenih vrijednosti. Međutim, umjesto toga možemo naći svojstvene vrijednosti kao korijene karakterističnog polinoma

$$\left|S_B - \lambda_i S_W\right| = 0$$

i onda riješiti

$$(S_B - \lambda_i S_W)\vec{w}_i = \vec{0}$$

izravno za svojstvene vektoren $\vec{w}_i$.

Budući da je $S_B$ suma c matrica ranga jedan ili manje i budući da su samo (c-1) od njih nezavisne $S_B$ je ranga (c-1) ili manje. Najviše je (c-1) svojstvenih vrijednosti različito od 0.

$$\widetilde{S}_B = \sum_{i=1}^{c} n_i (\widetilde{m}_i - \vec{m})(\widetilde{m}_i - \vec{m})^T$$

$$S_B = n_1(\vec{m}_1 - \vec{m})(\vec{m}_1 - \vec{m})^T + n_2(\vec{m}_2 - \vec{m})(\vec{m}_2 - \vec{m})^T + \cdots + n_c(\vec{m}_c - \vec{m})(\vec{m}_c - \vec{m})^T$$

Matrica $S_B$ je suma c matrica. Matrice su ranga jedan ili manje i samo (c-1) od njih su nezavisne. $S_B$ je ranga (c-1) ili manje. To znači da je najviše (c-1) svojstvenih vrijednosti različito od 0 i da (željeni) svojstveni vektori odgovaraju tim svojstvenim vrijednostima različitima od 0.

---

# 5. Skup uzoraka za učenje i skup uzoraka za ispitivanje - metode ispitivanja

Skup uzoraka za učenje je skup uzoraka s poznatom klasifikacijom (označeni uzorci).
**Važna pretpostavka:** U uzorcima za učenje sadržana je većina informacija o svojstvima razreda kojima uzorci pripadaju.

**1. Holdout metoda**

Ako imamo dovoljno velik skup uzoraka s poznatom klasifikacijom.

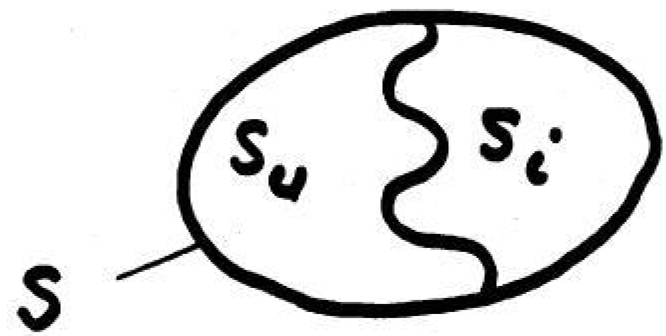
$$S_u - \text{skup uzoraka za ucenje} \; (|S_u| = N)$$

$$S_i - \text{skup uzoraka za ispitivanje}$$

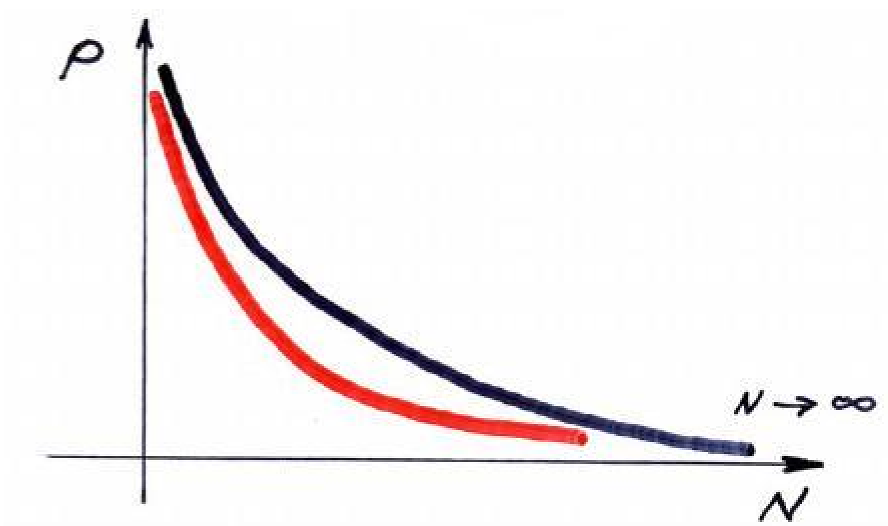
$$S = S_u \cup S_i$$

$$S_u \cap S_i = \emptyset$$

$$S = \text{skup uzoraka s poznatom klasifikacijom}$$




Nedostaci Holdout metode:
- smanjuje se veličina skupa za učenje i skupa za ispitivanje;
- kako podijeliti skup $S$ na $S_u$ i $S_i$?
- vjerojatnost greške klasifikatora koji se oblikuje uporabom konačnog skupa za učenje $N$ je
uvijek veća negoli odgovarajuća asimptotska vjerojatnost pogreške ($N \to \infty$).



## 2. Leave-One-Out metoda

Metoda pokušava *zaobići* problem podijele skupa označenih uzoraka. Učenje se obavlja uporabom $(N-1)$ uzoraka, a ispitivanje se izvodi uporabom onog jednog preostalog uzorka. Ako je taj uzorak pogrešno razvrstan inkrementira se brojilo pogreške. postupak se ponavlja $N$ puta, ali tako da je svaki put isključen drugi uzorak. Ukupan broj pogrešaka nas upućuje na procjenjenu vjerojatnost pogreške klasifikatora. Nedostatak metode je velika računska složenost.

## 3. Resubstitution metoda (Metoda ponovne zamjene)

Isti se skup uzoraka koristi prvo za učenje, a zatim za ispitivanje. Optimistička procjena vjerojatnosti pogreške klasifikatora.

Od skupa uzoraka za učenje zahtijeva se (za svaki uzorak):
- dovoljnost informacije;
- postojanost značajki;
- geometrijska postojanost (mala udaljenost među uzorcima u prostoru značajki znači i mali
razliku u svojstvima objekta).

$N$?
Idealno $N \to \infty$.

**Preporuka za N:**
Barem 3 do 5 puta više uzoraka za učenje po razredu od broja značajki (dimenzionalnost vektora značajki).

**Primjer:**
Sustav za autorizaciju osoba na temelju lica 580 korisnika - 110-komponentni vektor značajki.

$$M = 580$$

$$N = 5 \times 110 \times 580 = 319000 \text{ slika lica.}$$

**Primjer:**
Klasifikacija brojčano-slovčanih znakova.

$$M = 30 + 10 = 40$$

$$\omega_1, \omega_2, \cdots, \omega_{40}$$

Dimenzionalnost vektora značajki $n = 18$.

$$N = 5 \times 18 \times 40 = 3600 \text{ slika brojcano} - \text{slovcanih znakova.}$$