1 Višestruka diskriminantna analiza (Multiple Discriminant Analysis)

Problem c-razreda:
$D_1, D_2, \ldots, D_c$ odnosno $\omega_1, \omega_2, \ldots, \omega_c$
Generalizirana FLD uključuje c-1 diskrimnantnih funkcija :
Projekcija d-dimenzionalnog prostora u (c-1)-dimenzionalni prostor ($d \geq c$)
Generalizirana matrica raspršenosti unutar razreda (within-class scatter matrix):
$\mathrm{S}_W = \sum_{i=1}^{c} S_i$, gdje je $S_i = \sum_{\vec{X} \in D_i} (\vec{X} - \vec{m_i})(\vec{X} - \vec{m_i})^T$ 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 $\vec{m}$ $\vec{m} = \frac{1}{n} \sum_{\vec{X}} \vec{X} = \frac{1}{n} \sum_{i=1}^{c} n_i \vec{m_i}$, gdje je n_i broj uzoraka u razredu D_i (ω_i), $\vec{m_i}$ - vektor srednje vrijednosti vektora iz razreda ω_i

- ukupna matrica raspršenosti S_T
  $S_T = \sum_{\vec{X}} (\vec{X} - \vec{m})(\vec{X} - \vec{m})^T$
  $S_T = \sum_{i=1}^{c} \sum_{\vec{X} \in D_i} (\vec{X} - \vec{m_i} + \vec{m_i} - \vec{m})(\vec{X} - \vec{m_i} + \vec{m_i} - \vec{m})^T$
  $S_T = \sum_{i=1}^{c} \sum_{\vec{X} \in D_i} (\vec{X} - \vec{m_i})(\vec{X} - \vec{m_i})^T + \sum_{i=1}^{c} \sum_{\vec{X} \in D_i} (\vec{m_i} - \vec{m})(\vec{m_i} - \vec{m})^T$
  $S_T = S_W + \sum_{i=1}^{c} n_i(\vec{m_i} - \vec{m})(\vec{m_i} - \vec{m})^T$

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

Projekcija iz d-dimenzionalnog prostora u (c-1)dimenzionalni prostor uporabom (c-1)diskriminantnih funkcija $y_i = \vec{W_i}^T \vec{X} \ \ i = 1, 2, \ldots, c-1$
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:
$\vec{Y} = W^T \vec{X}$
Uzorci $\vec{x_1}, \vec{x_2}, \ldots, \vec{x_n}$ projiciraju se u odgovarajući skup uzoraka $\vec{y_1}, \vec{y_2}, \ldots, \vec{y_n}$ koji mogu biti opisani svojim srednjim vektorima i matricama raspršenosti:
$\tilde{m_i} = \frac{1}{n_i} \sum_{\vec{y} \in \mathcal{Y}_i} \vec{y}$

$\tilde{m} = \frac{1}{n} \sum_{n_i \tilde{m_i}}$
$\tilde{S_W} = \sum_{i=1}^{c} \sum_{\vec{y} \in \mathcal{Y}_i} (\vec{y} - \tilde{m_i})(\vec{y} - \tilde{m_i})^T$
$\tilde{S_B} = \sum_{i=1}^{c} n_i(\tilde{m_i} - \tilde{m})(\tilde{m_i} - \tilde{m})^T$
$\tilde{S_W} = W^T S_W W$
$\tilde{S_B} = W^T S_B W$
$\tilde{S_W} = \sum_{i=1}^{c} \sum_{\vec{y} \in \mathcal{Y}_i} (\vec{y} - \tilde{m_i})(\vec{y} - \tilde{m_i})^T$
$\vec{y} = W^T \vec{X} \qquad \tilde{m_i} = W^T \vec{m_i}$
$\tilde{S_W} = \sum_{i=1}^{c} \sum_{\vec{y} \in \mathcal{Y}_i} (W^T \vec{X} - W^T \vec{m_i})(W^T \vec{X} - W^T \vec{m_i})^T$
$\tilde{S_W} = \sum_{i=1}^{c} \sum_{\vec{y} \in \mathcal{Y}_i} W^T (\vec{X} - \vec{m_i})(\vec{X} - \vec{m_i})^T W$

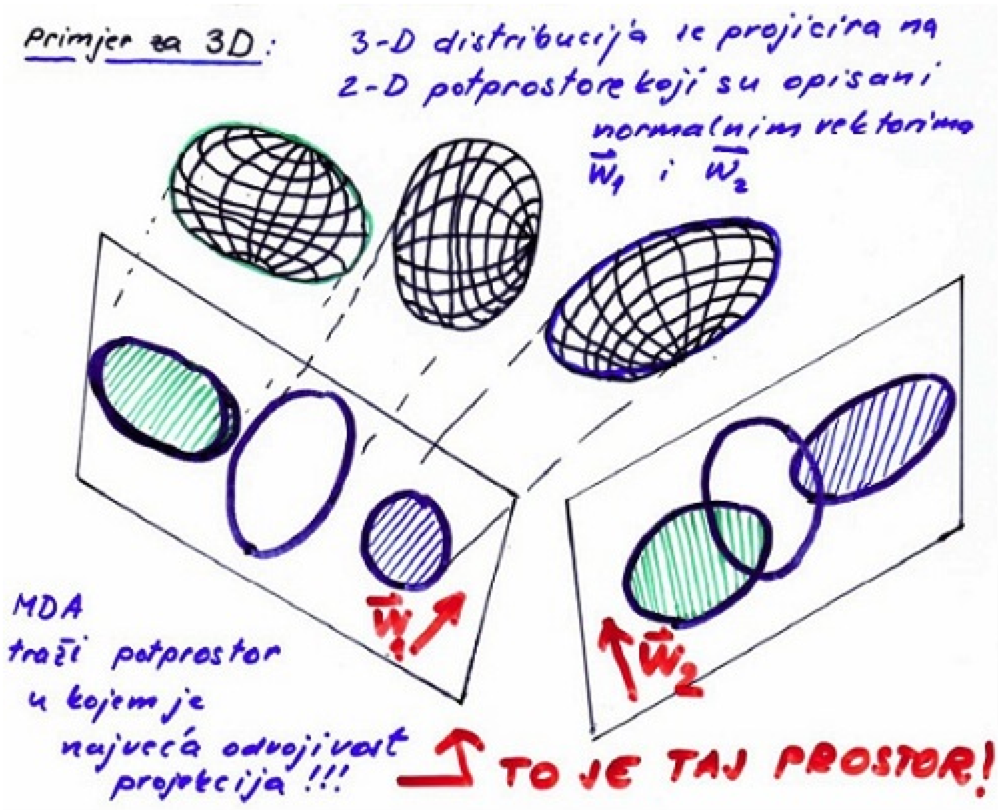
$$\tilde{S_W} = W^T \underbrace{\sum_{i=1}^{c} \sum_{\vec{y} \in \mathcal{Y}_i} (\vec{X} - \vec{m_i})(\vec{X} - \vec{m_i})^T}_{S_W = \sum_{i=1}^{c} S_i} W$$

$S_i = \sum_{\vec{X} \in D_i} (\vec{X} - \vec{m_i})(\vec{X} - \vec{m_i})^T$
$\tilde{S_W} = W^T S_W W$
$\tilde{S_B} = W^T S_B W$
$\tilde{S_W} = W^T S_W W$

Gornje jednadžbe pokazuju kako se tzv. within-class i between-class matrice raspršenja TRANSFORMIRAJU projekcijom u višedimenzionalni prostor.
TRAŽIMO TRANSFORMACIJSKU MATRICU W koja maksimizira omjer between-class raspršenost s within-class raspršenost!! $J(W) = \frac{|\tilde{S_B}|}{|\tilde{S_W}|}$
Jednostavna SKALARNA mjera raspršenosti je determinanta matrice raspršenosti
Determinanta je PRODUKT svojstvenih vrijednosti matrice.
PROBLEM TRAŽENJA (I NALAŽENJA) PRAVOKUTNE(!) MATRICE W koja maksimizira J(.) 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 zahtjeva 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 rješiti: $(S_B - \lambda_i S_W)\vec{W_i} = \vec{0}$



Slika 1: Primjer MDA

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 rješiti $(S_B - \lambda_i S_W)\vec{W}_i = \vec{0}$ izravno po $\vec{W}_i$.
$S_B = \sum_{i=1}^{c} n_i(\vec{m}_i - \vec{m})(\vec{m}_i - \vec{m})^T$
$S_B = n_1(\vec{m}_1-\vec{m})(\vec{m}_1-\vec{m})^T + n_2(\vec{m}_2-\vec{m})(\vec{m}_2-\vec{m})^T + \ldots + n_c(\vec{m}_c-\vec{m})(\vec{m}_c-\vec{m})^T$
Matrica S_B je suma c matrica! Matrice su ranga 1 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 RAZLIČITIH OD 0 I DA (ŽELJENI) SVOJSTVENI VEKTORI ODGOVARAJU TIM SVOJSTVENIM VRIJEDNOSTIMA RAZLIČITIM OD 0.

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

- problem nalaženja (pravokutne) matrice W koja maksimizira J(.).->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 konvencionalni problem svojstvenih vrijednosti
- međutim, umjesto toga možemo naći svojstvene vrijednosti kao korijene karakterističnog polinoma $|S_B - \lambda_i S_W| = 0$ i onda riješiti $(S_B - \lambda_i S_W)\vec{W}_i = \vec{0}$

<u>IZRAVNO</u> za svojstvene vektore $\vec{W}_i$ budući da je S_B suma c matrica ranga jedan ili manje, i budući da su samo c-1 od njih nezavisni, S_B je ranga c-1 ili manje

- najviše c-1 svojstvenih vrijednosti različito od 0.

2 SKUP UZORAKA ZA UČENJE I SKUP UZORAKA ZA ISPITIVANJE - METODE ISPITIVANJA

Skupa uzoraka za učenje - uzorci 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 veliki skup uzoraka s poznatom klasifikacijom

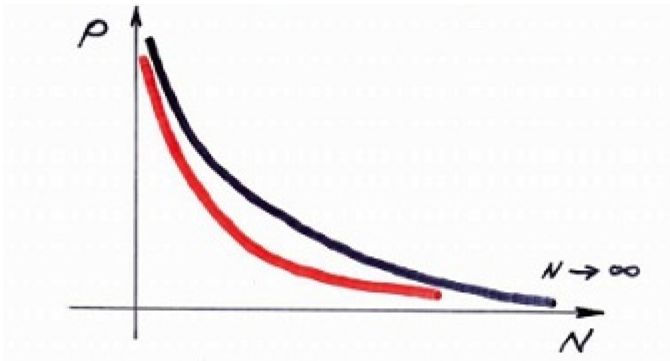
Slika 2: Skup uzoraka

S_u- skup uzoraka za učenje, $\#S_u = N$

S_i- skup uzoraka za ispitivanje

$S = S_u \bigcup S_i; \quad S_u \bigcap S_i = \emptyset$
S - skup uzoraka s poznatom klasifikacijom
Nedostaci Holdout metode:

- smanjuje se veličina skupa za učenje i skupa za isptivanje
- 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 je odgovarajuća asimptotska vjerojatnost pogreške ($N- > \infty$)



Slika 3: Vjerojatnost pogreške

2. Leave-One-Out metoda

    - metoda pokušava "zaobići" problem podjele 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 procjenu vjerojatnosti pogreške klasifikatora.
    Nedostatak metode: velika računska složenost

3. Resubstitution metoda (Metoda ponovne zmajene)
    Isti se skup podataka koristi, prvo za učenje, a zatim za ispitivanje.

    - optimistička procjena vjerojatnosti pogreške klasifiraktora

    Od skupa uzoraka za učenje zahtijeva se (za svaki uzorak):

    - dovoljnost informacija
    - postojanost značajki
    - geometrijska postojanost (mala udaljenost među uzorcima u prostoru značajki znači i malu razliku u svojstvima objekata)

N? Idealno $N- > \infty$
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 (M=580)
110 - komponentni vektor značajki
N=5*110*580=319000!!! (slika lica)
Primjer: Klasifikacija brojčano-slovčanih znakova M=30+10=40
$\omega_1, \omega_2, \ldots, \omega_{40}$
Dimenzionalnost vektora značajki n=18
N=5*18*40=3600 (slika brojčano-slovčanih znakova)