Višestruka diskriminantna analiza

(Multiple Discriminant Analysis)

Problem c-razreda:

$\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_c$ odnosno $\omega_1, \omega_2, \ldots \omega_c$

Generalizirana FLD uključuje $c-1$ diskriminantnih funkcija:

Projekcija d-dimenzionalnog prostora u (c-1)-dimenzionalni prostor ($d \geq c$).

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

$$S_W = \sum_{i=1}^{c} S_i \text{ , gdje je}$$

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

i

$$\vec{m}_i = \frac{1}{n_i} \sum_{\vec{x} \in \mathcal{D}_i} \vec{x}$$

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

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

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

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

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

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

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

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:

$$\tilde{m}_i = \frac{1}{n_i} \sum_{\vec{y} \in \mathcal{Y}_i} \vec{y}$$

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

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

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

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

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

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

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

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

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

$$\tilde{S}_W = W^T \underbrace{\sum_{i=1}^{c} \sum_{\vec{y} \in \mathcal{Y}} (\vec{x} - \vec{m}_i)(\vec{x} - \vec{m}_i)}_{S_W = \sum_{i=1}^{c} S_i} \cdot W$$

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

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

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

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

Gornje jednadžbe pokazuju kako se tzv. within-class i between-class matrice raspršenja TRANSFORMIRAJU projekcijom u niže-dimenzionalni prostor.

TRAŽIMO TRANSFORMACIJSKU MATRICU W koja maksimizira omjer

between-class raspršenosti s within-class raspršenosti

(!!)

- Jednostavna SKALARNA mjera raspršenosti je determinanta matrice raspršenosti

$$\boxed{J(W) = \frac{|\tilde{S}_B|}{|\tilde{S}_W|}}$$

Determinanta je PRODUKT svojstvenih vrijednosti matrice.

PROBLEM TRAŽENJA (I NALAŽENJA) PRAVOKUTNE(!) MATRICE W koja maksimizira $J(\cdot)$ je težak problem.

RJEŠENJE:

lfd 15a2

Stupci optimalne matrice W su generalizirani svojstveni vektori koji odgovaraju najvećim svojstvenim vrijednostima u

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

Ako je S_W nesingularna onda se problem može pretvoriti u konvencionalni problem svojstvenih vrijednosti. Međutim, to zahtijeva računanje inverzne matrice S_W.

$$\boxed{S_W^{-1} S_B \vec{w} = \lambda \vec{w}}$$

Umjesto toga možemo naći svojstvene vrijednosti kao korijene karakterističnog polinoma:

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

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

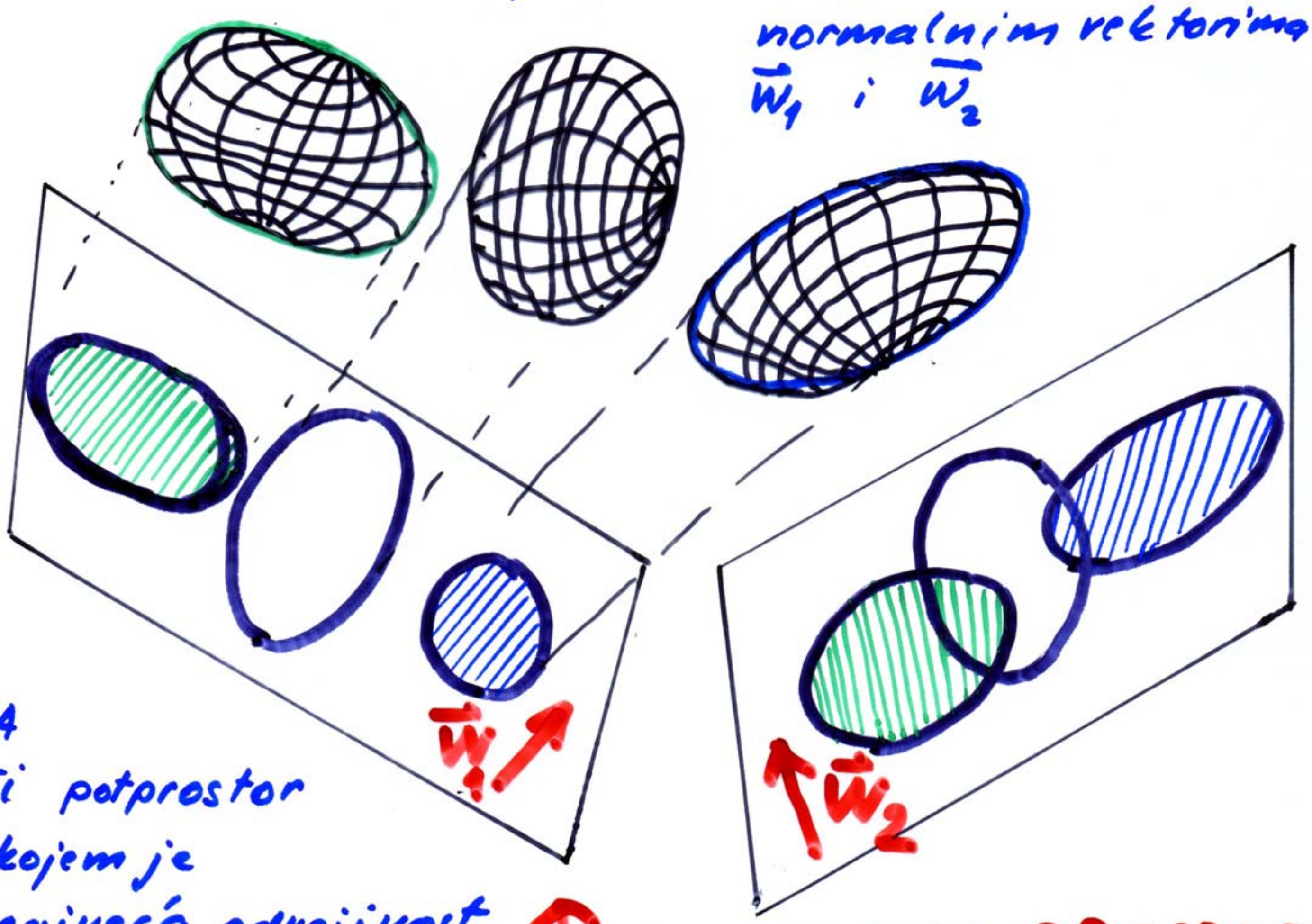
Naći
W koja maksimizira $J(W)$!

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

Naći svojstvene vrijednosti kao korijene karakt. 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: 3-D distribucija se projicira na 2-D potprostore koji su opisani normalnim vektorima $\vec{w}_1$ i $\vec{w}_2$



MDA traži potprostor u kojem je najveća odvojivost projekcija!!!

TO JE TAJ PROSTOR!

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 $\vec{w}_i$.

edf 16c

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

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

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 RAZLIČITO OD 0
I DA (RJEŠENI) 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 $J(\cdot)$. → složen!

- stupci optimalne matrice W su generalizirani svojstveni vektori koji odgovaraju najvećim svojstvenim vrijednostima u

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

- 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

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

<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.

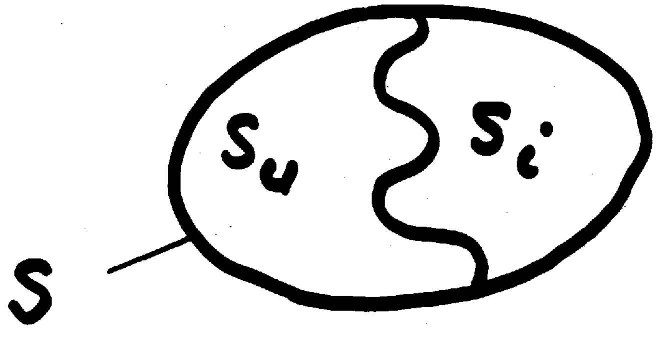
- najviše je $c-1$ svojstvenih vrijednosti različito od 0.

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

Skup uzoraka za učenje - uzorci s poznatom klasifikacijom (označeni uzorci)

<u>Važna pretpostavka:</u> u uzorcima za učenje sadržana je većina informacije o svojstvima razreda kojima uzorci pripadaju

1. Ako imamo dovoljno veliki skup uzoraka s poznatom klasifikacijom



Holdout metoda

$\boxed{\# S_u = N}$

S_u - skup uzoraka za učenje

S_i - skup uzoraka za ispitivanje

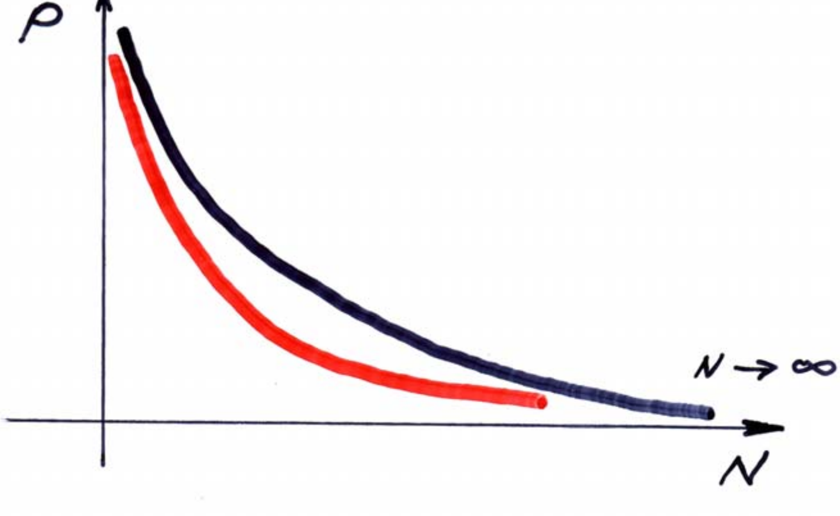
$$S = S_u \cup S_i \quad ; \quad S_u \cap S_i = \emptyset$$

S - 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 je odgovarajuća asimptotska vjerojatnost pogreške ($N \to \infty$)



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 procjenjenu vjerojatnost pogreške klasifikatora

Nedostatak metode: velika računska složenost

3. Resubstitution metoda
(Metoda ponovne zamjene)

Isti se skup podataka 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 malu razliku u svojstvima objekta)

N ?

Idealno $N \to \infty$

<u>Preporuka za N</u>

- 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

$$N \doteq 5 * 110 * 580 = 319\,000 \; !!!$$

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