Pravilo razvrstavanja - Ako je $d_i(\vec{x}) > d_j(\vec{x})$ te $j \neq i$ tada uzorak $\vec{x}$ pripada razredu ω_i

$$d(\vec{x}) = \underbrace{w_1}_{} x_1 + \underbrace{w_2}_{\text{težinski koeficijenti}} x_2 + \ldots + \underbrace{w_n}_{} x_n + \underbrace{w_{n+1}}_{\text{pomaknuće}} = \vec{W}^T \vec{X} + w_{n+1}$$

DVA RAZREDA

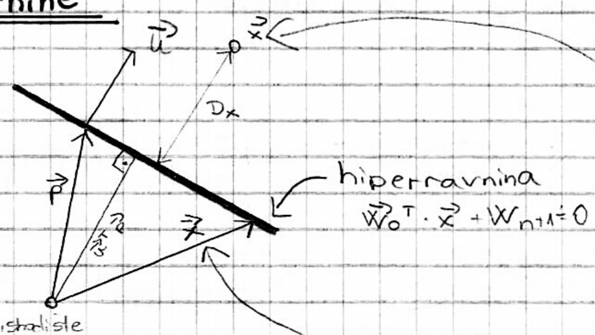
ako $d(\vec{x}) > 0$ onda $\vec{x} \in \omega_1$

$d(\vec{x}) < 0$ $\quad \vec{x} \in \omega_2$

$d(\vec{x}) = 0$ $\quad$ nedefinirano

$$d(\vec{x}) = \vec{w_0}^T \vec{x} + w_{n+1} = 0$$

Jednadžba hiperravnine



$$\| \vec{w_0} \| = \sqrt{w_1^2 + w_2^2 + \ldots + w_n^2}$$

$$\vec{w_0}^T \vec{x} + w_{n+1} = 0$$

$$\vec{w_0}^T \vec{x} = - w_{n+1} \quad / \cdot \frac{1}{\|\vec{w_0}\|}$$

$$\frac{\vec{w_0}^T \vec{x}}{\|\vec{w_0}\|} = - \frac{w_{n+1}}{\|\vec{w_0}\|}$$

$$\vec{u}^T(\vec{x} - \vec{p}) = 0$$

$$\vec{u}^T \vec{x} - \vec{u}^T \vec{p} = 0$$

$$\vec{u}^T \vec{x} = \vec{u}^T \vec{p}$$

$$\vec{u}^T \vec{x} = \frac{\vec{w_0}^T \vec{x}}{\|\vec{w_0}\|}$$

$$\boxed{\vec{u}^T \vec{p} = - \frac{w_{n+1}}{\|\vec{w_0}\|}} = \frac{|w_{n+1}|}{\|\vec{w_0}\|} = D_0$$

udaljenost hiperravnine od ishodišta

$$\vec{u}^T = \frac{\vec{w_0}^T}{\|\vec{w_0}\|}$$

$$\boxed{\vec{u} = \frac{\vec{w_0}}{\|\vec{w_0}\|}}$$

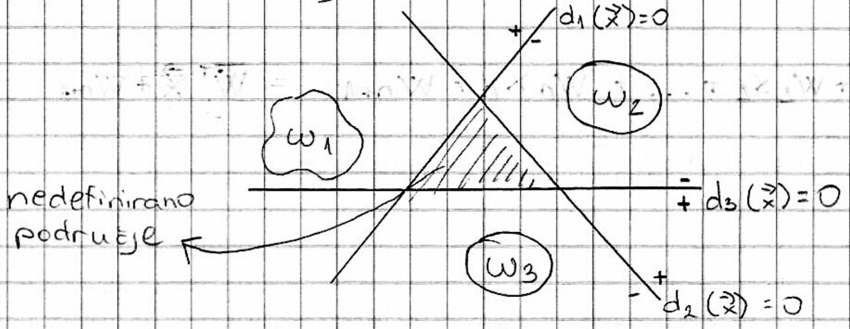
orjentacija hiperravnine

$$\boxed{D_x = |\vec{u}^T \vec{p} - \vec{u}^T \vec{x}|}$$

udaljenost točke $\vec{x}$ od hiperravnine

Separabilnost više razreda

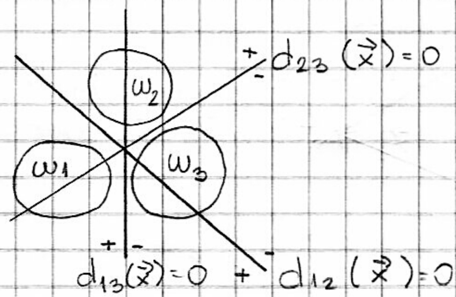
(1) Svaki je razred uzoraka separabilan od ostalih razreda jednom decizijskom ravninom



Broj linearnih f-ja: c

$$d_i(\vec{x}) = \vec{w}_i^T \vec{x} = \begin{cases} >0 & \text{ako } \vec{x} \in \omega_i \\ <0 & \text{ako } \vec{x} \notin \omega_i \end{cases}$$

(2) Svaki razred uzoraka je separabilan sa svakim pojedinim drugim razredom i to jednom decizijskom ravninom



Broj linearnih f-ja: $\frac{c(c-1)}{2}$

$d_{ij}(\vec{x}) \geq 0$ za svaki $j \neq i$

↳ $\vec{x}$ pripada ω_i

Granica između ω_i i ω_j zadana je s

$$d_{ij}(\vec{x}) = w_{ij1}x_1 + w_{ij2}x_2 + \dots + w_{ijn}x_n + w_{ijn+1} = 0$$

(3.) Postoji c decizijskih funkcija sa svojstvom da $\vec{x}$ pripada razredu ω_i ako

$$d_i(\vec{x}) > d_j(\vec{x}) \quad \text{za sve } j \neq i$$

Možemo definirati

$$d_{ij}(\vec{x}) = d_i(\vec{x}) - d_j(\vec{x}) = (\vec{w}_i - \vec{w}_j)^T \vec{x} = \vec{w}_{ij}^T \vec{x}$$

ako je $d_i(\vec{x}) > d_j(\vec{x})$ za sve $j \neq i$ onda je $d_{ij}(\vec{x}) > 0$ za sve $j \neq i$ tj.

ako su razredi separabilni za (3) onda su automatski separabilni i za (2.)

$$d_{ij}(\vec{x}) = d_i(\vec{x}) - d_j(\vec{x}) = (w_{i1} - w_{j1})x_1 + (w_{i2} - w_{j2})x_2 + \dots + (w_{in} - w_{jn})x_n + (w_{in+1} - w_{jn+1}) = 0$$

$$\Rightarrow \boxed{(\vec{w}_i - \vec{w}_j)^T \vec{x} + (w_{in+1} - w_{jn+1}) = 0}$$

Razdvojni vektor

→ Vektor $\vec{w}$ koji zadovoljava sustav linearnih nejednadžbi

$$[x] \cdot \vec{w} > \vec{0}$$

gdje je

$$\vec{x}_{11}, \vec{x}_{12}, \dots \in \omega_1$$
$$\vec{x}_{21}, \vec{x}_{22}, \dots \in \omega_2$$

→ dimenzionalnost vektora povećava se za jedan (dodaje se 1 na kraj)

$$i \quad [x] = \begin{bmatrix} \vec{x}_{11} \\ \vec{x}_{12} \\ \vdots \\ \vec{x}_{21} \\ \vec{x}_{22} \\ \vdots \end{bmatrix}$$

} uzorci iz ω_2 pomnoženi su sa -1

Uzorak je pravilno razvrstan ako za sve $\vec{x}$ iz ω_1 i ω_2 vrijedi

$$\vec{w}^T \vec{x} > 0$$

gdje su uzorci iz ω_2 pomnoženi s -1

Tražimo funkciju koja dostiže minimum kada je ispunjen uvjet

$$\vec{w}^T \vec{x} > 0$$

Gradijentni postupak

Korak po korak povećavamo argument $\vec{w}$ u smjeru negativnog gradijenta funkcije $J(\vec{w}, \vec{x})$ dok ne postignemo njen minimum

$$\vec{w}(k+1) = \vec{w}(k) - c \left\{ \frac{\partial J(\vec{w}, \vec{x})}{\partial \vec{w}} \right\}_{\vec{w} = \vec{w}(k)}$$

a)

Uzimamo funkciju: $\boxed{J(\vec{w}, \vec{x}) = \frac{1}{2}\left(|\vec{w}^T \vec{x}| - \vec{w}^T \vec{x}\right)}$

$$(|x|)' = \operatorname{sgn}(x)$$

$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{ako } x > 0 \\ -1 & \text{ako } x \leq 0 \end{cases}$$

$$\frac{\partial J}{\partial \vec{w}} = \frac{1}{2}\left[\vec{x} \operatorname{sgn}(\vec{w}^T \vec{x}) - \vec{x}\right]$$

$$\vec{w}(k+1) = \vec{w}(k) - c \cdot \frac{\partial J}{\partial \vec{w}} = \vec{w}(k) + \frac{c}{2}\left[\vec{x}(k) - \vec{x}(k) \operatorname{sgn}(\vec{w}^T(k) \vec{x}(k))\right]$$

$$\vec{w}(k+1) = \vec{w}(k) + \frac{c}{2} \cdot \begin{cases} 0 & \text{ako } \vec{w}^T(k)\, \vec{x}(k) > 0 \\ 2\vec{x}(k) & \text{ako } \vec{w}^T(k) \vec{x}(k) \leq 0 \end{cases}$$

$$\vec{w}(k+1) = \vec{w}(k) + c \cdot \begin{cases} 0 & \text{ako } \vec{w}^T(k) \cdot \vec{x}(k) > 0 \\ \vec{x}(k) & \text{ako } \vec{w}^T(k) \cdot \vec{x}(k) \leq 0 \end{cases}$$

b)

Uzimamo funkciju: $\boxed{J(\vec{w},\vec{x}) = \frac{1}{4\vec{x}^T\vec{x}}\left(|\vec{w}^T\vec{x}|^2 - |\vec{w}^T\vec{x}|\cdot \vec{w}^T\vec{x}\right)}$

$$\frac{\partial J}{\partial \vec{w}} = \frac{1}{4\vec{x}^T\vec{x}}\left[2\cdot|\vec{w}^T\vec{x}|\cdot \mathrm{sgn}(\vec{w}^T\vec{x})\cdot\vec{x} - \mathrm{sgn}(\vec{w}^T\vec{x})\cdot\vec{x}\cdot\vec{w}^T\vec{x} - |\vec{w}^T\vec{x}|\cdot\vec{x}\right]$$

$$= \frac{1}{4\vec{x}^T\vec{x}}\left[2|\vec{w}^T\vec{x}|\cdot\vec{x}\,\mathrm{sgn}(\vec{w}^T\vec{x}) - |\vec{w}^T\vec{x}|\cdot\vec{x} - |\vec{w}^T\vec{x}|\cdot\vec{x}\right]$$

$$= \frac{1}{4\vec{x}^T\vec{x}}\left[2|\vec{w}^T\vec{x}|\vec{x}\,\mathrm{sgn}(\vec{w}^T\vec{x}) - 2|\vec{w}^T\vec{x}|\vec{x}\right]$$

$$= \frac{1}{2\vec{x}^T\vec{x}}\left[|\vec{w}^T\vec{x}|\vec{x}\left(\mathrm{sgn}(\vec{w}^T\vec{x})-1\right)\right]$$

$$= \frac{|\vec{w}^T\vec{x}|}{2\vec{x}^T\vec{x}}\left[\vec{x}\left(\mathrm{sgn}(\vec{w}^T\vec{x})-1\right)\right] = \frac{|\vec{w}^T\vec{x}|}{2\vec{x}^T\vec{x}}\cdot\begin{cases}0 & \text{ako } \vec{w}^T\vec{x}>0\\ -2\vec{x} & \text{ako } \vec{w}^T\vec{x}\le 0\end{cases}$$

$$\vec{w}(k+1) = \vec{w}(k) - \eta\frac{\partial J}{\partial \vec{w}} = \vec{w}(k) - \eta\cdot\frac{|\vec{w}^T\vec{x}|}{2\vec{x}^T\vec{x}}\cdot\begin{cases}0 & \text{ako } \vec{w}^T\vec{x}>0\\ -2\vec{x} & \text{ako } \vec{w}^T\vec{x}\le 0\end{cases} =$$

$$\vec{w}(k+1) = \vec{w}(k) + \eta\frac{|\vec{w}^T(k)\cdot\vec{x}(k)|}{\vec{x}^T(k)\cdot\vec{x}(k)}\cdot\begin{cases}\vec{0} & \text{ako } \vec{w}^T(k)\vec{x}(k)>0\\ \vec{x}(k) & \text{ako } \vec{w}^T(k)\vec{x}(k)\le 0\end{cases}$$

- za $0<\eta<2 \Rightarrow$ postupak učenja konvergira
- za $\eta > 1 \Rightarrow$ uzorak se pravilno razvrstava nakon svake korekcije $\vec{w}$

Varijacije algoritma perceptrona

a) C = konst. >0 → ALGORITAM SA STALNIM PRIRASTOM

b) C izabran tako da je dovoljno velik da jamči da će uzorak biti ispravno klasificiran nakon ugađanja težinskog faktora $\vec{w}$

– ako je $\vec{w}^T(k)\cdot x(k)\le 0$ izabire se C tako da vrijedi

ALGORITAM S APSOLUTNOM KOREKCIJOM

$$\vec{w}^T(k+1)\,\vec{x}(k) = \left[\vec{w}(k) + c\,\vec{x}(k)\right]^T\vec{x}(k) > 0$$

$$\vec{w}^T(k)\vec{x}(k) + c\,\vec{x}^T(k)\,\vec{x}(k) > 0$$

$$c > \frac{|\vec{w}^T(k)\vec{x}(k)|}{\vec{x}^T(k)\vec{x}(k)}$$

C je najmanji cijeli broj veći od

c) C je izabran tako da vrijedi

$$\left|\vec{w}^T(k)\vec{x}(k) - \vec{w}^T(k+1)\vec{x}(k)\right| = \lambda \left|\vec{w}^T(k)\vec{x}(k)\right| \qquad (1)$$

izraz $\vec{w}(k+1) = \vec{w}(k) + c\vec{x}(k)$ tj. $\vec{w}^T(k+1) = \vec{w}^T(k) + c\vec{x}^T(k)$ uvrstimo u (1)

ALGORITAM S DJELOMIČNOM KOREKCIJOM

$$\left|\cancel{\vec{w}^T(k)\vec{x}(k)} - \cancel{\vec{w}^T(k)\vec{x}(k)} - c\vec{x}^T(k)\vec{x}(k)\right| = \lambda \left|\vec{w}^T(k)\vec{x}(k)\right|$$

$$\left|c\vec{x}^T(k)\vec{x}(k)\right| = \lambda \left|\vec{w}^T(k)\vec{x}(k)\right|$$

$$c \cdot \underbrace{\left|\vec{x}^T(k)\vec{x}(k)\right|}_{\text{kvadriranje je pozitivno}} = \lambda \left|\vec{w}^T(k)\vec{x}(k)\right|$$

$$c = \lambda \frac{\left|\vec{w}^T(k)\vec{x}(k)\right|}{\vec{x}^T(k)\vec{x}(k)}$$

PERCEPTRON - klasifikator uzoraka u dva razreda - ako su razredi linearno separabilni

Perceptron za $M > 2$

M = broj razreda

PRAVILO - uzorak se svrstava u razred ω_i akko

$$R_i > R_j \quad \text{za sve } i \neq j$$

gdje je $R = \vec{w}^T \cdot \vec{x}$

Algoritam perceptrona

1) Ako $\vec{x}(k) \in \omega_1$ i $\vec{w}^T(k)\vec{x}(k) \leq 0$

zamijeni $\vec{w}(k)$ sa $\vec{w}(k+1) = \vec{w}(k) + c\vec{x}(k)$

2) Ako $\vec{x}(k) \in \omega_2$ i $\vec{w}^T(k)\vec{x}(k) > 0$

zamijeni $\vec{w}(k)$ sa $\vec{w}(k+1) = \vec{w}(k) - c\vec{x}(k)$

3) Inače $\vec{w}(k+1) = \vec{w}(k)$

c → korekcijski faktor, mora biti pozitivan

$\vec{w}(1)$ → početna vrijednost vektora težinskih koeficijenata - proizvoljno izabrana

RJEŠENJE - dobiva se kada se pojavi iteracija bez pogrešaka tj. iteracija u kojoj za sve uzorke vrijedi 3)

Ho-Kashyap algoritam (LMSE - Last Mean Square Error)

- Traži se vektor $\vec{w}$ tako da vrijedi

$$[x]\vec{w} = \vec{b}$$

$\vec{b}$ - vektor kojemu su sve komponente pozitivne

Kriterijska funkcija:

$$J(\vec{w},\vec{x},\vec{b}) = \frac{1}{2}\sum_{j=1}^{N}(\vec{w}^T\vec{x}_j - b_j)^2 = \frac{1}{2}\underbrace{\left\| [x]\vec{w} - \vec{b} \right\|}_{\text{veličina vektora } [x]\vec{w}-b}^2$$

Gradijenti

$$\frac{\partial J}{\partial \vec{w}} = \frac{1}{\cancel{2}}\cdot[x]^T\cdot\cancel{2}([x]\vec{w}-\vec{b}) = [x]^T([x]\vec{w}-\vec{b})$$

$$\frac{\partial J}{\partial \vec{b}} = \frac{1}{2}\cdot(-1)\cdot 2([x]\vec{w}-\vec{b}) = -([x]\vec{w}-\vec{b})$$

- Pošto se umjesto uvjeta $[x]\vec{w} > \vec{0}$ koristi $[x]\vec{w} = \vec{b}$ to znači da $\vec{w}$ nije ograničen i možemo uzeti $\frac{\partial J}{\partial \vec{w}} = \vec{0}$

$$\frac{\partial J}{\partial \vec{w}} = \vec{0}$$

$$[x]^T([x]\vec{w}-\vec{b}) = \vec{0}$$

$$[x]^T[x]\vec{w} - [x]^T\vec{b} = \vec{0}$$

$$[x]^T[x]\vec{w} = [x]^T\vec{b} \quad / \cdot ([x]^T[x])^{-1}$$

$$\vec{w} = ([x]^T[x])^{-1}[x]^T\vec{b}$$

$$[x]^{\#} = ([x]^T[x])^{-1}[x]^T$$ → generalizirani inverz od $[x]$

$$\vec{w} = [x]^{\#}\vec{b} \qquad (1)$$

- Vektor $\vec{b}$ se mijenja tako da uvijek vrijedi $b_i > 0$ za $i = 0,1,2,\ldots N$

$$\vec{b}(k+1) = \vec{b}(k) + \delta\vec{b}(k) \qquad (2)$$

$$\delta b_i(k) = \begin{cases} 2c([x]\vec{w}(k) - \vec{b}(k))_i & \text{ako } ([x]\vec{w}(k)-\vec{b}(k))_i > 0 \\ 0 & \text{ako } ([x]\vec{w}(k)-\vec{b}(k))_i \le 0 \end{cases}$$

$$\vec{e}(k) = [x]\vec{w}(k) - \vec{b}(k) \Rightarrow \delta\vec{b}(k) = c(\vec{e}(k) + |\vec{e}(k)|)$$

(2) se uvrštava u (1)

$$\vec{w}(k+1) = [x]^{\#}\,\vec{b}(k+1)$$
$$= [x]^{\#}\left(\vec{b}(k) + \delta\vec{b}(k)\right)$$
$$= \underbrace{[x]^{\#}\,\vec{b}(k)} + [x]^{\#}\,\delta\vec{b}(k)$$
$$= \vec{w}(k) + [x]^{\#}\cdot c\left(\vec{e}(k) + |\vec{e}(k)|\right)$$

$$\vec{w}(k+1) = \vec{w}(k) + c\,[x]^{\#}\left(\vec{e}(k) + |\vec{e}(k)|\right)$$

$$\vec{b}(k+1) = \vec{b}(k) + c\left(\vec{e}(k) + |\vec{e}(k)|\right)$$

Pravilo – ako su u bilo kojoj iteraciji sve komponente vektora $\vec{e}(k) < 0$ razredi NISU linearno separatibilni

– kada je $\vec{e}(k) = \vec{0}$ znači da vrijedi $[x]\vec{w}(k) = b(k)$ pa je $\vec{w}(k)$ RJEŠENJE.

Poopćeni algoritam perceptrona za M razreda

- istodobno se određuje M težinskih koeficijenata: $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_M$
- računa se M decizijskih funkcija

$$d_j\left(\vec{x}(k)\right) = \vec{w}_j^{\,T}(k)\,\vec{x}(k) \qquad \text{za } j = 1, 2, \ldots, M$$

Pravilo: Ako

$$d_i\left(\vec{x}(k)\right) > d_j\left(\vec{x}(k)\right) \qquad \text{za } j = 1, 2, \ldots, M$$

ONDA $\vec{w}_j(k+1) = \vec{w}_j(k)$ za $j = 1, 2, \ldots, M$

Ako

$$d_i\left(\vec{x}(k)\right) \le d_l\left(\vec{x}(k)\right)$$

ONDA

$$\vec{w}_i(k+1) = \vec{w}_i(k) + c\,\vec{x}(k)$$
$$\vec{w}_l(k+1) = \vec{w}_l(k) - c\,\vec{x}(k)$$
$$\vec{w}_j(k+1) = \vec{w}_j(k) \qquad \text{za } j = 1, 2, \ldots, M,\ j \ne i,\ j \ne l$$

– Postupak konvergira u konačnom broju ponavljanja ako su razredi linearno separatibilni

FISHEROVA LINEARNA DISKRIMINANTA

- d-dimenzionalan vektor značajki $\vec{x}$ reducira se na jednu dimenziju

$$R^d \rightarrow R^1$$

- ZADATAK - naći orjentaciju pravca na koji se projiciraju d-dimenzionalni uzorci, ali tako da su projicirani uzorci separatibilni

n - broj uzoraka $= n_1 + n_2$ → broj uzoraka razreda ω_1 / broj uzoraka razreda ω_2

$\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$ - d-dimenzionalni uzorci

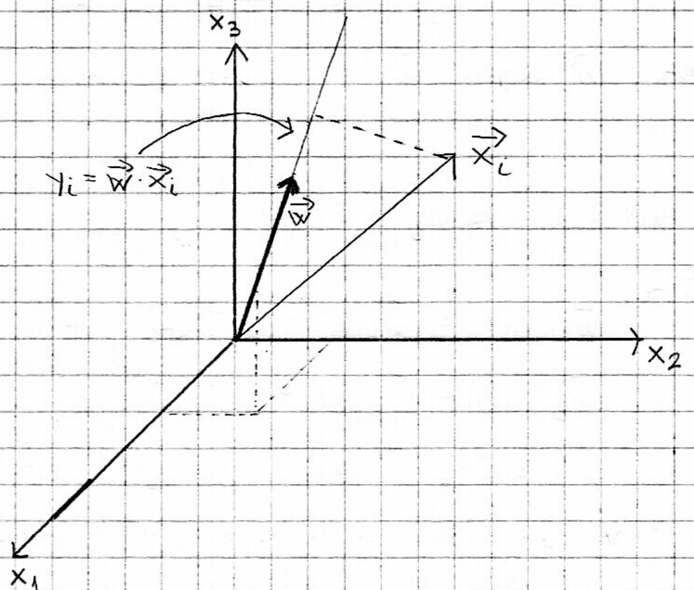
$$y_i = \vec{w}^T \vec{x}_i \qquad i = 1, 2, \ldots, n$$

→ linearna kombinacija komponenti $\vec{x}$

$$D_1, D_2 \longrightarrow Y_1, Y_2$$

D_1 → d-dimenzionalno; Y_1 → jednodimenzionalno

ako $\|\vec{w}\| = 1$ → svaki y_i je projekcija odgovarajućeg $\vec{x}_i$ na pravac u smjeru $\vec{w}$

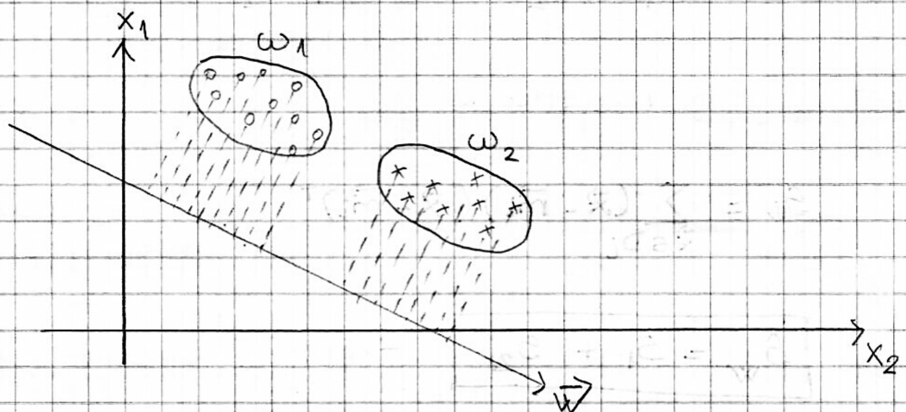


Cilj – naći najbolji smjer od $\vec{w}$

MJERA ODVOJIVOSTI između projiciranih uzoraka je razlika srednjih vrijednosti projiciranih uzoraka

$$\vec{m}_i = \frac{1}{n_i} \sum_{\vec{x} \in D_i} \vec{x} \quad ; \; i = 1,2$$
→ srednja vrijednost uzoraka svakog razreda u d-dimenzionalnom prostoru

$$\tilde{m}_i = \frac{1}{n_i} \sum_{y \in Y_i} y \quad ; \; i = 1,2$$
→ srednja vrijednost projiciranih uzoraka



$$\tilde{m}_i = \frac{1}{n_i} \sum_{y \in Y_i} y = \frac{1}{n_i} \sum_{\vec{x} \in D_i} \vec{w}^T \vec{x} = \vec{w}^T \cdot \left[\frac{1}{n_i} \sum_{\vec{x} \in D_i} \vec{x}\right] = \vec{w}^T \vec{m}_i$$

→ $\tilde{m}_i$ je projekcija vektora $\vec{m}_i$

Razlika srednjih vrijednosti projiciranih uzoraka

$$|\tilde{m}_1 - \tilde{m}_2| = |\vec{w}^T \vec{m}_1 - \vec{w}^T \vec{m}_2| = |\vec{w}^T (\vec{m}_1 - \vec{m}_2)|$$

→ skaliranjem $\vec{w}$ razliku činimo proizvoljno velikom

Kriterijska funkcija:

$$J(\vec{w}) = \frac{(\tilde{m}_1 - \tilde{m}_2)^2}{\sigma_1^2 - \sigma_2^2}$$
→ razlika sr. vrijednosti
→ varijanca uzoraka unutar razreda

- Za dobro odvajanje projiciranih uzoraka udaljenost između srednjih vrijednosti projiciranih točaka mora biti relativno velika u odnosu na standardnu devijaciju svakog razreda

Linearna funkcija $\vec{w}^T \vec{x}$ za koju je kriterijska funkcija $J(\vec{w})$ MAKSIMALNA vodi najboljem razdvajanju između projiciranih skupova

- Umjesto varijanci σ_i^2 raspršenost unutar razreda definiramo kao

$$\tilde{s}_i^2 = \sum_{y \in \mathcal{Y}_i} (y - \tilde{m}_i)^2$$

$$\frac{1}{n} \underbrace{\left(\tilde{s}_1^2 + \tilde{s}_2^2 \right)}_{\text{mjera raspršenosti projiciranih podataka unutar razreda}} \rightarrow \text{procjena varijance podataka}$$

$$J(\vec{w}) = \frac{(\tilde{m}_1 - \tilde{m}_2)^2}{\sigma_1^2 + \sigma_2^2} = \frac{(\tilde{m}_1 - \tilde{m}_2)^2}{\tilde{s}_1^2 + \tilde{s}_2^2}$$

Raspršenost unutar razreda ω_i:

$$S_i = \sum_{\vec{x} \in D_i} (\vec{x} - \vec{m}_i)(\vec{x} - \vec{m}_i)^T \rightarrow \text{matrica raspršenosti unutar razreda } \omega_i$$

matrica

$$\boxed{S_w = S_1 + S_2} \rightarrow \text{matrica raspršenosti unutar razreda}$$

$$\tilde{s}_i^2 = \sum_{y \in \mathcal{Y}_i} (y - \tilde{m}_i)^2 = \left\{ \begin{array}{l} y = \vec{w}^T \vec{x} \\ \tilde{m}_i = \vec{w}^T \vec{m}_i \end{array} \right\}$$

$$= \sum_{\vec{x} \in D_i} (\vec{w}^T \vec{x} - \vec{w}^T \vec{m}_i)^2$$

$$(\vec{x}^T \vec{w} - \vec{m}_i^T \vec{w}) = (\vec{x} - \vec{m}_i)^T \vec{w}$$

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

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

$$= \vec{w}^T \cdot \left[\sum_{\vec{x} \in D_i} (\vec{x} - \vec{m}_i)(\vec{x} - \vec{m}_i)^T \right] \cdot \vec{w}$$

$$= \vec{w}^T S_i \vec{w}$$

$$\tilde{s}_i^2 = \vec{w}^T S_i \vec{w}$$

$$\tilde{s}_1^2 + \tilde{s}_2^2 = \vec{w}^T S_1 \vec{w} + \vec{w}^T S_2 \vec{w} = \vec{w}^T (S_1 + S_2) \vec{w} = \vec{w}^T S_w \vec{w}$$

$$\tilde{s}_1^2 + \tilde{s}_2^2 = \vec{w}^T S_w \vec{w}$$

$$(\tilde{m}_1 - \tilde{m}_2)^2 = (\vec{w}^T \vec{m}_1 - \vec{w}^T \vec{m}_2)^2$$
$$= (\vec{w}^T \vec{m}_1 - \vec{w}^T \vec{m}_2)(\vec{w}^T \vec{m}_1 - \vec{w}^T \vec{m}_2)^T$$
$$= (\vec{w}^T \vec{m}_1 - \vec{w}^T \vec{m}_2)(\vec{m}_1^T \vec{w} - \vec{m}_2^T \vec{w})$$
$$= \vec{w}^T \underbrace{(\vec{m}_1 - \vec{m}_2)(\vec{m}_1 - \vec{m}_2)^T}_{S_B} \vec{w}$$
$$= \vec{w}^T S_B \vec{w}$$

$$\boxed{S_B = (\vec{m}_1 - \vec{m}_2)(\vec{m}_1 - \vec{m}_2)^T}$$
$$(\tilde{m}_1 - \tilde{m}_2)^2 = \vec{w}^T S_B \vec{w}$$

$$\boxed{J(\vec{w}) = \frac{\vec{w}^T S_B \vec{w}}{\vec{w}^T S_W \vec{w}}}$$

$$\left(\frac{a}{b}\right)' = \frac{b a' - a b'}{b^2}$$

Traži se maksimum funkcije $J(\vec{w})$

$$\frac{\partial J}{\partial \vec{w}} = 0$$

$$\frac{\partial J(\vec{w})}{\partial \vec{w}} = \frac{(\vec{w}^T S_W \vec{w})(\vec{w}^T S_B \vec{w})' - (\vec{w}^T S_B \vec{w})(\vec{w}^T S_W \vec{w})'}{(\vec{w}^T S_W \vec{w})^2}$$

$$= \frac{(\vec{w}^T S_W \vec{w}) \cdot 2 S_B \vec{w} - (\vec{w}^T S_B \vec{w}) \cdot 2 S_W \vec{w}}{(\vec{w}^T S_W \vec{w})(\vec{w}^T S_W \vec{w})^T} = 0$$

$$(\vec{w}^T S_W \vec{w}) \cdot 2 S_B \vec{w} - (\vec{w}^T S_B \vec{w}) \cdot 2 S_W \vec{w} = 0 \quad / : 2$$

$$(\vec{w}^T S_W \vec{w}) \cdot S_B \vec{w} = (\vec{w}^T S_B \vec{w}) \cdot S_W \vec{w} \quad / \cdot (\vec{w}^T S_W \vec{w})^{-1}$$

$$S_B \vec{w} = \underbrace{(\vec{w}^T S_B \vec{w})(\vec{w}^T S_W \vec{w})^{-1}}_{\text{skalar}} S_W \vec{w}$$

$$\lambda = (\hat{\vec{w}}^T S_B \hat{\vec{w}})(\hat{\vec{w}}^T S_W \hat{\vec{w}})^{-1}$$

$$\boxed{S_B \vec{w} = \lambda S_W \vec{w}}$$ → generalizirani problem svojstvenog vektora

Ako S_W^{-1} postoji onda je smjer $\vec{w}$-a

$$\hat{\vec{w}} = (S_W^{-1} S_B) \vec{w}$$

Pošto je $S_B \cdot \vec{w}$ uvijek usmjeren kao $(\vec{m}_1 - \vec{m}_2)$, a faktor skaliranja za $\vec{w}$ nas ne zanima nego samo smjer, onda pišemo

$$\boxed{\hat{\vec{w}} = S_W^{-1} (\vec{m}_1 - \vec{m}_2)}$$ → $\vec{w}$ maksimizira omjer raspršenja između razreda i raspršenja unutar razreda

Višestruka FLD → generalizirana FLD

C – broj razreda

Postoji $(c-1)$ diskriminantnih funkcija

Cilj: Projekcija d-dimenzionalnog prostora u $(c-1)$-dimenzionalni

$d \geq c$

$$\boxed{S_W = \sum_{i=1}^{c} S_i} \rightarrow \text{generalizirana matrica raspršenosti unutar razreda}$$

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

$$\vec{m}_i = \frac{1}{n_i} \sum_{\vec{x} \in D_i} \vec{x} \qquad / \cdot n_i$$

$$n_i \vec{m}_i = \sum_{\vec{x} \in D_i} \vec{x}$$

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

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

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

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

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

$$= \sum_{i=1}^{c} \sum_{\vec{x} \in D_i} \left[(\vec{x} - \vec{m}_i)(\vec{x} - \vec{m}_i)^T + \underbrace{(\vec{x} - \vec{m}_i)(\vec{m}_i - \vec{m})^T + (\vec{m}_i - \vec{m})(\vec{x} - \vec{m}_i)^T}_{} + (\vec{m}_i - \vec{m})(\vec{m}_i - \vec{m})^T\right]$$

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

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

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

Drugi član je

$$\boxed{S_B = \sum_{i=1}^{c} n_i (\vec{m}_i - \vec{m})(\vec{m}_i - \vec{m})^T}$$ → generalizirana matrica raspršenosti između razreda

$$S_T = S_W + S_B$$

Projekcija iz d-dimenz. prostora u (c-1)-dimenz. uz (c-1) diskriminantnih funkcija:

$$y_i = \vec{w}_i^T \vec{x} \quad ; \quad i = 1, 2, \dots, c-1$$

$\vec{y}$ → vektor čije su komponente y_i

W → matrica dimenzija $d \times (c-1)$
→ stupci su joj vektori $\vec{w}_i$

$$\boxed{\vec{y} = W^T \vec{x}}$$

Uzorci $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n$ projiciraju se u $\vec{y}_1, \vec{y}_2, \dots, \vec{y}_n$ čiji vektori srednjih vrijednosti i matrica raspršenosti iznose:

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

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

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

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

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

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

$$= W^T S_W W$$

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

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

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

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

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

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

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

$$= W^T S_B W$$

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

Cilj: Naći matricu W koja maksimizira omjer $\tilde{S_B} / \tilde{S_W}$

- jednostavna skalarna mjera raspršenosti je DETERMINANTA

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

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

$$S_B \vec{w_i} = \lambda_i S_W \vec{w_i}$$

Traže se svojstvene vrijednosti kao korijeni karakterističnog polinoma

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

Zatim se traže svojstveni vektori rješavanjem

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

METODE ISPITIVANJA

1) Holdout metoda

Skup uzoraka s poznatom klasifikacijom dijeli se na:

- Skup uzoraka za učenje = S_u
- Skup uzoraka za ispitivanje = S_i

NEDOSTACI - smanjene skupova za učenje i ispitivanje
- kako pravilno razdijeliti skupove?

2) Leave - One - Out metoda

- Učenje se obavlja na N-1 uzoraka a ispitivanje na onom preostalom
- ovisno je li preostali uzorak točno ili netočno klasificiran povećava se brojilo točnosti ili pogreške
- postupak se ponavlja N puta tako da je svaki put za ispitivanje korišten drugi uzorak

NEDOSTATAK - velika računska složenost

3) Resubstitution metoda (metoda ponovne zamjene)

- Isti skup se koristi prvo za učenje pa za ispitivanje
- OPTIMISTIČKA procjena vjerojatnosti klasifikatora

SVM - SUPPORT VECTOR MACHINES

(strojevi s potpornim vektorima)

Cilj: Konstrukcija decizijske hiperravnine ali tako da margina odvajanja između pozitivne i negativne skupine uzoraka bude MAKSIMALNA

Skup uzoraka za učenje:

$$\{(\vec{x}_i, d_i)\}_{i=1}^{N}$$

N → broj primjera

$\vec{x}_i$ → ulazni vektor uzoraka za i-ti primjer

d_i → željeni odgovor klasifikatora

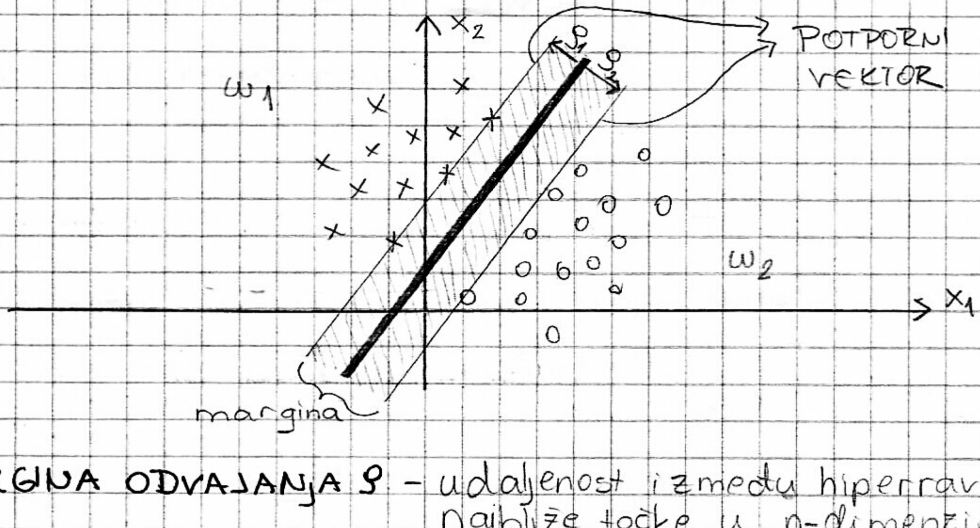
za $\omega_1 \rightarrow d_i = +1$
$\omega_2 \rightarrow d_i = -1$

Jednadžba decizijske ravnine:

$$\vec{w}^T \vec{x} + b = 0$$

$\vec{w}$ → vektor težinskih koeficijenata
$\vec{x}$ → ulazni vektor
b → pomaknuće (w_0)

za: $d_i = +1$ vrijedi $\vec{w}^T \vec{x} + b \geq 0$
$d_i = -1$ vrijedi $\vec{w}^T \vec{x} + b < 0$



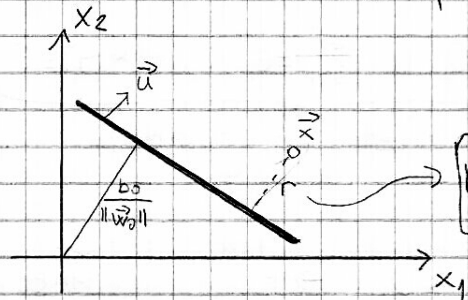
MARGINA ODVAJANJA ρ - udaljenost između hiperravnine i najbliže točke u n-dimenzionalnom prostoru

OPTIMALNA RAVNINA - hiperravnina za koju je margina odvajanja maksimalna

$\vec{w}_o, b_o$ - optimalne vrijednosti $\vec{w}$ i b

$\vec{w}_o^T \vec{x} + b_o = 0$ → optimalna hiperravnina

$\boxed{g(\vec{x}) = \vec{w}_o^T \vec{x} + b_o}$ → decizijska funkcija
- daje mjeru udaljenosti $\vec{x}$-a od optimalne hiperravnine



$$\boxed{r = \frac{g(\vec{x})}{\|\vec{w}_o\|}}$$

Par $(\vec{w}_o, b_o)$ mora zadovoljiti (ako su uzorci lin. odvojivi)

$$\vec{w}_o^T \vec{x}_i + b_o \geq 1 \quad \text{za } d_i = +1$$

$$\vec{w}_o^T \vec{x}_i + b_o \leq -1 \quad \text{za } d_i = -1$$

POTPORNI VEKTORI - uzorci iz skupa za učenje za koje vrijedi

$\vec{x}^{(s)}$

$$\vec{w}_o^T \vec{x} + b_o = 1 \quad \text{za } d_i = +1$$

i

$$\vec{w}_o^T \vec{x} + b_o = -1 \quad \text{za } d_i = -1$$

$$g(\vec{x}^{(s)}) = \vec{w}_o^T \vec{x}^{(s)} + b_o = \pm 1 \quad \text{za } d_i = \pm 1$$

$$r = \frac{g(\vec{x}^{(s)})}{\|\vec{w}_o\|} = \frac{\pm 1}{\|\vec{w}_o\|} = \begin{cases} 1/\|\vec{w}_o\| & \text{za } d^{(s)} = +1 \\ -1/\|\vec{w}_o\| & \text{za } d^{(s)} = -1 \end{cases}$$

Optimalna vrijednost margine odvajanja

$$\boxed{\rho = 2r = \frac{2}{\|\vec{w}_o\|}}$$

→ minimiziranjem norme od $\vec{w}_o$ maksimizira se margina odvajanja

CILJ: Pronaći optimalne vrijednosti koeficijenata $\vec{w}$ i pomaknuće b tako da je zadovoljeno ograničenje

$$d_i \cdot (\vec{w}^T \vec{x}_i + b) \geq 1 \quad , \; i = 1, 2, \dots, N$$

a da pri tomu vektor $\vec{w}$ minimizira **KRITERIJSKU FUNKCIJU**

$$J(\vec{w}) = \frac{1}{2} \underbrace{\vec{w}^T \vec{w}}_{\vec{w}^T \vec{w} = \|w\|^2}$$

Problem optimizacije rješava se Lagrangeovim multiplikatorima:

Određivanje ekstrema funkcije $f(x,y)$ uz uvjet $\varphi(x,y)=0$

$$F = f(x,y) + \lambda \cdot \varphi(x,y)$$

λ → Lagrangeov multiplikator

$$\underbrace{\frac{\partial F}{\partial x} = 0 \qquad \frac{\partial F}{\partial y} = 0 \qquad \varphi(x,y) = 0}_{\text{ovim sustavom jednadžbi određuju se } x, y \text{ i } \lambda}$$

AKO

$d^2F < 0$ → $f(x,y)$ ima MAKSIMUM

$d^2F > 0$ → $f(x,y)$ ima MINIMUM

$$d^2F = \frac{\partial^2 F}{\partial x^2} dx^2 + 2\frac{\partial^2 F}{\partial x \partial y} dx\,dy + \frac{\partial^2 y}{\partial y^2} dy^2$$

ODREĐIVANJE OPTIMALNE HIPERRAVNINE

$$J(\vec{w}) = \frac{1}{2}\vec{w}^T\vec{w}$$

$$d_i \cdot (\vec{w}^T\vec{x}_i + b) \geq 1 \quad ; \; i = 1, 2, \dots, N$$

↳ $\varphi(\vec{w}, b) = d_i(\vec{w}^T\vec{x}_i + b) - 1$

$$J(\vec{w}, b, \lambda) = \frac{1}{2}\vec{w}^T\vec{w} - \sum_{i=1}^{N} \lambda_i \underbrace{\left[d_i(\vec{w}^T\vec{x}_i + b) - 1\right]}_{} \quad (*)$$

$\lambda_i d_i(\vec{w}^T\vec{x}_i + b) - \lambda_i$
$= \lambda_i d_i \vec{w}^T\vec{x}_i + \lambda_i d_i b - \lambda_i$

$$\frac{\partial J}{\partial \vec{w}}, \quad \frac{\partial J}{\partial b}, \quad \varphi(\vec{w}, b) = 0$$

1) $$\frac{\partial J(\vec{w}, b, \lambda)}{\partial \vec{w}} = \frac{1}{2} \cdot 2\vec{w} - \sum_{i=1}^{N} \lambda_i d_i \vec{x}_i = \vec{w} - \sum_{i=1}^{N} \lambda_i d_i \vec{x}_i = 0$$

$$\boxed{\vec{w} = \sum_{i=1}^{N} \lambda_i d_i \vec{x}_i}$$

$$\frac{\partial J(\vec{w}, b, \lambda)}{\partial b} = -\sum_{i=1}^{N} \lambda_i d_i = 0$$

$$\boxed{\sum_{i=1}^{N} \lambda_i d_i = 0}$$

(**) vrijednosti za koje $J(\vec{w})$ dostiže minimum

$$\lambda_i\left[d_i(\vec{w}^T\vec{x}_i+b)-1\right]=0$$

$$\boxed{\lambda_i\left[d_i(\vec{w}^T\vec{x}_i+b)-1\right]=0 \qquad ; \quad i=1,2,\dots,N}$$

$$\boxed{\lambda_i \geq 0}$$

$$J(\vec{w},b,\lambda)=\frac{1}{2}\vec{w}^T\vec{w}-\sum_{i=1}^{N}\lambda_i d_i \vec{w}^T\vec{x}_i-\sum_{i=1}^{N}\lambda_i d_i b+\sum_{i=1}^{N}\lambda_i$$

$$=\frac{1}{2}\vec{w}^T\vec{w}-\sum_{i=1}^{N}\lambda_i d_i \underbrace{\vec{w}^T}_{\vec{w}=\sum_{j=1}^{N}\lambda_j d_j \vec{x}_j}\vec{x}_i-b\underbrace{\sum_{i=1}^{N}\lambda_i d_i}_{=0}+\sum_{i=1}^{N}\lambda_i$$

$**$ se uvrštava u $*$ da bi se dobilo $\min_{\vec{w}} J(\vec{w},\vec{\lambda})$

$$=\frac{1}{2}\vec{w}^T\vec{w}-\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i d_i \lambda_j d_j \vec{x}^T_j\vec{x}_i+\sum_{i=1}^{N}\lambda_i$$

$$=\frac{1}{2}\vec{w}^T\vec{w}-\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i \lambda_j d_i d_j \vec{x}^T_i\vec{x}_j+\sum_{i=1}^{N}\lambda_i$$

$$=\left\{\vec{w}^T\cdot\vec{w}=\left[\sum_{i=1}^{N}\lambda_i d_i \vec{x}^T_i\right]\cdot\left[\sum_{j=1}^{N}\lambda_j d_j \vec{x}_j\right]=\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j d_i d_j \vec{x}_i^T\vec{x}_j\right\}$$

$$=\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j d_i d_j \vec{x}_i^T\vec{x}_j-\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j d_i d_j \vec{x}_i^T\vec{x}_j+\sum_{i=1}^{N}\lambda_i$$

$$=\sum_{i=1}^{N}\lambda_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j d_i d_j \vec{x}_i^T\vec{x}_j$$

$$\boxed{Q(\lambda)=J(\vec{w},b,\lambda)=\sum_{i=1}^{N}\lambda_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j d_i d_j \vec{x}_i^T\vec{x}_j}=\min_{\vec{w}} J(\vec{w},\vec{\lambda})$$

Lagrangeov dualni problem

Optimizacijski zadatak – $\overbrace{\text{minimizacija } J(\vec{w})}^{\min_{\vec{w}} J(\vec{w})}$ uz uvjet $f_i(\vec{w})\geq 0$

Lagrangeova funkcija

$$J(\vec{w},\lambda)=J(\vec{w})-\underbrace{\sum_{i=1}^{N}\lambda_i f_i(\vec{w})}_{=0}$$

ima maksimum kada je

$$\lambda_i=0 \quad ; \quad i=1,2,\dots,N$$

$$\text{ili} \quad f_i(\vec{w})=0 \quad ; \quad i=1,2,\dots,N$$

tj. $\max_{\lambda}\left(J(\vec{w},\vec{\lambda})\right)=J(\vec{w})$

Pa se originalni problem može prikazati kao

$$\min_{\vec{w}} J(\vec{w})=\min_{\vec{w}}\max_{\lambda} J(\vec{w},\vec{\lambda})=\max_{\lambda}\min_{\vec{w}} J(\vec{w},\vec{\lambda})$$

Zadatak se svodi na $\max_{\lambda} Q(\lambda)$

SVM za $M > 2$ razreda

- Za svaki od razreda traži se optimalna decizijska funkcija $g_i(\vec{x})$ –

  tako da vrijedi $g_i(\vec{x}) > g_j(\vec{x}) \quad \forall\, j \neq i$

  ako je $\vec{x} \in \omega_i$

ZADATAK – Naći decizijsku funkciju $g_i(\vec{x}) = 0$ koja je optimalna hiperravnina koja odvaja razred ω_i od svih ostalih:

$$g_i(\vec{x}) > 0 \quad \text{za } \vec{x} \in \omega_i$$

$$g_i(\vec{x}) < 0 \quad \text{inače}$$

KLASIFIKACIJSKO PRAVILO

Dodijeli $\vec{x}$ u ω_i ako

$$i = \arg \underset{k}{\text{maks}} \{ g_k(\vec{x}) \}$$

SVM za linearno neodvojive razrede

Cilj: Pronalazak optimalne hiperravnine koja minimizira vjerojatnost klasifikacijske pogreške

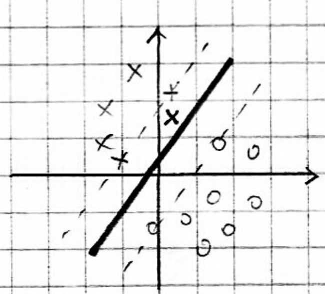
Meka margina razdvajanja – margina kod koje za uzorak $(\vec{x}_i, d_i)$ NE VRIJEDI

$$d_i \cdot (\vec{w}^T \vec{x}_i + b) \geq +1$$

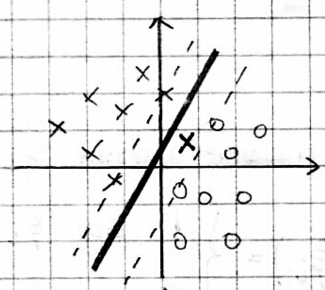
Povreda uvjeta $d_i (\vec{w}^T \vec{x}_i + b) \geq +1$ u dva slučaja:

1) uzorak $(\vec{x}_i, d_i)$ pada unutar područja odvajanja ali s prave strane decizijske hiperravnine

2) uzorak $(\vec{x}_i, d_i)$ pada unutar područja odvajanja na krivu stranu decizijske hiperravnine



1)



2)

Kategorije uzoraka za učenje

1) Uzorci koji padaju izvan margine i ispravno su klasificirani. Za njih vrijedi:

$$d_i (\vec{w}^T \vec{x}_i + b) \geq 1 \qquad \xi_i = 0$$

2) Uzorci koji padaju unutar margine ali su pravilno razvrstani. Za njih vrijedi

$$0 \leq d_i (\vec{w}^T \vec{x}_i + b) < 1 \qquad 0 < \xi_i \leq 1$$

3) Uzorci koji su pogrešno razvrstani (mogu biti i unutar i izvan margine. Za njih vrijedi:

$$d_i (\vec{w}^T \vec{x}_i + b) < 0 \qquad \xi_i > 1$$

Jedinstveni oblik: $d_i (\vec{w}^T \vec{x}_i + b) \geq 1 - \xi_i$

$\xi_i \rightarrow$ Labava varijabla

- mjera odstupanja vektora od idealnog uvjeta separabilnosti uzoraka

Potporni vektori zadovoljavaju:

$$d_i(\vec{w}^T \vec{x} + b) \geq 1 - \xi_i$$

čak i za slučaj $\xi_i > 0$

Cilj: - Naći decizijsku funkciju za koju je pogreška klasifikacije minimalna tj.

- učiniti marginu što je moguće većom ali uz uvjet da je broj uzoraka sa

$$\xi_i > 0$$

što je moguće manji

$\xi \rightarrow$ vektor s komponentama ξ_i

$$I(\xi_i) = \begin{cases} 1 & \text{za } \xi_i > 0 \quad \rightarrow \text{loš slučaj} \\ 0 & \text{za } \xi_i = 0 \quad \rightarrow \text{dobar slučaj} \end{cases}$$

Funkcija koštanja

$$J(\vec{w}, b, \xi) = \frac{1}{2}\|\vec{w}\|^2 + C \underbrace{\sum_{i=1}^{N} I(\xi_i)}_{\text{broj loših slučajeva}}$$

C - pozitivna konstanta kojom se upravlja relativan međuutjecaj $\frac{1}{2}\|\vec{w}\|^2$ i $\sum_{i=1}^{N} I(\xi_i)$

Optimizacija je teška jer je $I(\xi_i)$ diskontinuirana funkcija pa se koristi nova funkcija "bliska" funkciji koštanja:

$$J(\vec{w}, b, \xi) = \frac{1}{2}\vec{w}^T \vec{w} + C \sum_{i=1}^{N} \xi_i$$

Primarni problem za neseparatibilne razrede

Za zadani skup uzoraka $\{\vec{x}_i, d_i\}_{i=1}^{N}$ pronaći optimalne vrijednosti vektora $\vec{w}$ tako da su zadovoljena ograničenja

$$d_i(\vec{w}^T \vec{x}_i + b) \geq 1 - \xi_i \qquad ; \; i = 1, 2, \ldots, N$$

$$uz \qquad \xi_i \geq 0$$

te da labave varijable ξ_i minimiziraju funkciju koštanja

$$J(\vec{w}, b, \vec{\xi}) = \frac{1}{2} \vec{w}^T \vec{w} + C \sum_{i=1}^{N} \xi_i$$

Uporaba Lagrangeovih multiplikatora

$$\mathcal{L}(\cdot) = J(\vec{w}, b, \vec{\xi}) - \lambda\, f_i(\cdot)$$

$$\mathcal{L}(\vec{w}, b, \vec{\xi}, \vec{\lambda}, \vec{\mu}) = \frac{1}{2}\vec{w}^T\vec{w} + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\underbrace{\lambda_i\left[d_i(\vec{w}^T\vec{x}_i + b) - 1 + \xi_i\right]} - \sum_{i=1}^{N}\mu_i\,\xi_i$$

$$= \lambda_i\left[d_i\vec{w}^T\vec{x}_i + d_i b - 1 + \xi_i\right]$$
$$= \lambda_i d_i \vec{w}^T\vec{x}_i + \lambda_i d_i b - \lambda_i + \lambda_i\xi_i$$

Tražimo $\frac{\partial \mathcal{L}}{\partial \vec{w}}, \frac{\partial \mathcal{L}}{\partial b}, \frac{\partial \mathcal{L}}{\partial \vec{\xi}}$

$$\frac{\partial \mathcal{L}}{\partial \vec{w}} = \frac{1}{2}\cdot 2\vec{w} - \sum_{i=1}^{N}\lambda_i d_i \vec{x}_i = \vec{w} - \sum_{i=1}^{N}\lambda_i d_i \vec{x}_i = 0$$

$$\boxed{\vec{w} = \sum_{i=1}^{N}\lambda_i d_i \vec{x}_i} \qquad i = 1, 2, \ldots, N$$

$$\frac{\partial \mathcal{L}}{\partial b} = -\sum_{i=1}^{N}\lambda_i d_i = 0$$

$$\boxed{\sum_{i=1}^{N}\lambda_i d_i = 0} \qquad i = 1, 2, \ldots, N$$

$$\frac{\partial \mathcal{L}}{\partial \xi_i} = C\cdot 1 - \lambda_i - \mu_i = C - \lambda_i - \mu_i = 0$$

$$\boxed{C - \lambda_i - \mu_i = 0} \qquad i = 1, 2, \ldots, N$$

$$\boxed{\lambda_i\left[d_i(\vec{w}^T\vec{x}_i + b) - 1 + \xi_i\right] = 0} \qquad i = 1, 2, \ldots, N$$

$$\boxed{\mu_i\,\xi_i = 0} \qquad i = 1, 2, \ldots, N$$

$$\boxed{\mu_i \geq 0} \qquad \boxed{\lambda_i \geq 0}$$

Novi zadatak:

Maksimizacija

$$J(\vec{w}, b, \vec{\xi}, \pi_i, \mu_i) = \frac{1}{2}\vec{w}^T\vec{w} + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\underbrace{\pi_i\left[d_i(\vec{w}^T\vec{x}_i + b) - 1 + \xi_i\right]}_{\substack{=\pi_i\left[d_i\vec{w}^T\vec{x}_i + d_i b - 1 + \xi_i\right] \\ =\pi_i d_i\vec{w}^T\vec{x}_i + \pi_i d_i b - \pi_i + \pi_i\xi_i}} - \underbrace{\sum_{i=1}^{N}\mu_i\xi_i}_{=0}$$

uz uvjete

1) $\vec{w} = \sum_{i=1}^{N}\pi_i d_i \vec{x}_i$

2) $C - \mu_i - \pi_i = 0$

3) $\pi_i \geq 0, \quad \mu_i \geq 0$

4) $\sum_{i=1}^{N}\pi_i d_i = 0$

5) $\mu_i \xi_i = 0$

$$\vec{w}^T\vec{w} = \sum_{i=1}^{N}\pi_i d_i \vec{x}_i^T \cdot \sum_{j=1}^{N}\pi_j d_j \vec{x}_j = \sum_{i=1}^{N}\sum_{j=1}^{N}\pi_i\pi_j d_i d_j \vec{x}_i^T\vec{x}_j$$

$$J(\vec{w}, b, \vec{\xi}, \pi_i, \mu_i) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\pi_i\pi_j d_i d_j \vec{x}_i^T\vec{x}_j + \sum_{i=1}^{N}(\mu_i + \pi_i)\xi_i - \sum_{i=1}^{N}\pi_i d_i\vec{w}^T\vec{x}_i - \sum_{i=1}^{N}\pi_i d_i b + \sum_{i=1}^{N}\pi_i - \sum_{i=1}^{N}\pi_i\xi_i$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\pi_i\pi_j d_i d_j \vec{x}_i^T\vec{x}_j + \underbrace{\sum_{i=1}^{N}\mu_i\xi_i}_{=0} + \cancel{\sum_{i=1}^{N}\pi_i\xi_i} - \sum_{i=1}^{N}\pi_i d_i\vec{w}^T\vec{x}_i - b\underbrace{\sum_{i=1}^{N}\pi_i d_i}_{=0} + \sum_{i=1}^{N}\pi_i - \cancel{\sum_{i=1}^{N}\pi_i\xi_i}$$

$$= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\pi_i\pi_j d_i d_j \vec{x}_i^T\vec{x}_j - \sum_{i=1}^{N}\pi_i d_i\vec{w}^T\vec{x}_i + \sum_{i=1}^{N}\pi_i$$

$$= \left\{\sum_{i=1}^{N}\pi_i d_i\vec{w}^T\vec{x}_i = \sum_{i=1}^{N}\sum_{j=1}^{N}\pi_i\pi_j d_i d_j \vec{x}_i^T\vec{x}_j\right\}$$

$$= \sum_{i=1}^{N}\pi_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\pi_i\pi_j d_i d_j \vec{x}_i^T\vec{x}_j = Q(\pi)$$

Novi zadatak (Lagrangeov dualni problem)

Traži se maksimum funkcije

$$Q(\vec{\pi}) = \sum_{i=1}^{N}\pi_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\pi_i\pi_j d_i d_j \vec{x}_i^T\vec{x}_j$$

tj. traži se

$$\max_{\vec{\pi}} Q(\vec{\pi}) = \max_{\vec{\pi}}\left(\sum_{i=1}^{N}\pi_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\pi_i d_i \pi_j d_j \vec{x}_i^T\vec{x}_j\right)$$

Labave varijable ξ_i i njihovi multiplikatori μ_i ne ulaze izravno u problem nego su prisutni kroz uvjet

$$0 \leq \pi_i \leq C$$

kod lin. sep. je $\pi_i \geq 0$

Skalarna produktna jezgra za SVM

Cilj: Pronalazak optimalne hiperravnine za linearno neseparatibilne razrede uporabom SVM-a

KORACI:

1) Nelinearno preslikavanje ulaznog vektora u prostor značajki većih dimenzija

2) Konstrukcija optimalne hiperravnine za odvajanje vektora značajki dobivenih u 1)

m_0 – dimenzija ulaznog prostora

m_1 – dimenzija prostora značajki većih dimenzija

$\vec{x}$ – vektor iz ulaznog prostora

$\{\varphi_j(\vec{x})\}_{j=1}^{m_1}$ – skup nelinearnih transformacija iz ulaznog prostora u prostor značajki

$$\sum_{j=1}^{m_1} w_j \varphi_j(\vec{x}) + b = 0 \rightarrow \text{DECIZIJSKA RAVNINA}$$

$$\Downarrow$$

$$\sum_{j=0}^{m_1} w_j \varphi_j(\vec{x}) = 0 \rightarrow \begin{array}{l} \varphi_0(\vec{x}) = 1 \\ w_0 = b \end{array}$$

$$\vec{\varphi}(\vec{x}) = [\varphi_0(\vec{x}), \varphi_1(\vec{x}), \dots, \varphi_{m_1}(\vec{x})]^T$$ → slika inducirana u prostoru značajki zbog ulaznog vektora $\vec{x}$

$$\boxed{\vec{w}^T \vec{\varphi}(\vec{x}) = 0} \quad *$$

U prostoru značajki postoji linearna separatibilnost za koju je SVM dao:

$$\vec{w} = \sum_{i=1}^{N} \alpha_i d_i \vec{x}_i \Rightarrow \boxed{\vec{w} = \sum_{i=1}^{N} \alpha_i d_i \vec{\varphi}(\vec{x}_i)}$$

$\vec{w}$ se uvrštava u $*$:

$$\vec{w}^T \vec{\varphi}(\vec{x}) = \sum_{i=1}^{N} \alpha_i d_i \underbrace{\vec{\varphi}^T(\vec{x}_i) \cdot \vec{\varphi}(\vec{x})}_{} = 0$$

$$\boxed{K(\vec{x}, \vec{x}_i) = \vec{\varphi}^T(\vec{x}) \cdot \vec{\varphi}(\vec{x}_i)}$$ → Skalarna produktna jezgra

skalarni produkt dvaju vektora iz prostora značajki koji su inducirani ulaznim vektorima $\vec{x}$ i $\vec{x}_i$

$$K(\vec{x}, \vec{x}_i) = \sum_{j=0}^{m_1} \varphi_j(\vec{x}) \cdot \varphi_j(\vec{x}_i) \quad \text{za } i = 1, 2, \dots, N$$

Optimalna hiperravnina u prostoru značajki je stoga zadana s:

$$\boxed{\sum_{i=1}^{N} \alpha_i d_i K(\vec{x}, \vec{x}_i) = 0}$$

u izvornom prostoru je nelinearna a u prostoru značajki linearna

Prema SVM-u imamo

$$J(\vec{\alpha}) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j \vec{x}_i^T \vec{x}_j$$

pa ista formula u prostoru značajki iznosi

$$J(\vec{\alpha}) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j \cdot K(\vec{x}_i, \vec{x}_j)$$

uz ograničenja

$$\sum_{i=1}^{N} \alpha_i d_i = 0 \qquad 0 \leq \alpha_i \leq C \qquad \text{za } i = 1, 2, \ldots N$$

pozitivni parametar

optimalni vektor

$$\vec{w}_0 = \sum_{i=1}^{N} \alpha_{0,i} d_i \varphi(\vec{x}_i)$$

prva komponenta vektora $\vec{w}_0$ predstavlja optimalno pomaknuće b_0

POOPĆENE LINEARNE DECIZIJSKE FUNKCIJE

Poopćeni oblik linearne decizijske funkcije

$$\boxed{d(\vec{x}) = \sum_{i=1}^{k+1} w_i f_i(\vec{x})}$$

$$d(\vec{x}) = w_1 f_1(\vec{x}) + w_2 f_2(\vec{x}) + \dots + w_k f_k(\vec{x}) + w_{k+1}$$

$$f_{k+1} = 1$$

$$\vec{x}^* = \begin{bmatrix} f_1(\vec{x}) \\ f_2(\vec{x}) \\ \vdots \\ f_k(\vec{x}) \\ 1 \end{bmatrix}$$

$$\boxed{d(\vec{x}) = \vec{w}^T \vec{x}^*}$$

- linearna funkcija u k-dimenzijskom prostoru
- "virtualno linearna

$$\vec{w} = (w_1, w_2, \dots, w_k, w_{k+1})^T$$

Obavlja se transformacija iz n-dimenzijskog prostora u k-dimenzijski.

Zadatak: Odabrati $\{f_i(\vec{x})\}_{i=1}^{k}$ - u obliku polinoma

2 stupnja

$$d(\vec{x}) = \sum_{j=1}^{n} w_{jj} x_j^2 + \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} w_{jk} x_j x_k + \sum_{j=1}^{n} w_j x_j + w_{n+1}$$

$$k = \frac{(n+1)(n+2)}{2}$$

$1, 2$

$d(x) =$ $\quad \vec{x} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \begin{matrix} \to x_1 \\ \to x_2 \end{matrix}$

$n = 2 \quad r = 3$

r-tog stupnja

$$k = \frac{(n+r)!}{r!\,n!}$$

1 1 1	→	x_1^3
1 1 2	→	$x_1^2 x_2$
1 2 2	-	$x_1 x_2^2$
2 2 2	-	x_2^3
1 1	-	x_1^2
1 2	-	$x_1 x_2$
2 2	-	x_2^2
1	-	x_1
2	-	x_2

$\vec{x}^* =$

BAYESOVA DECIZIJSKA TEORIJA

$P(\omega_i)$ - a' priori vjerojatnost

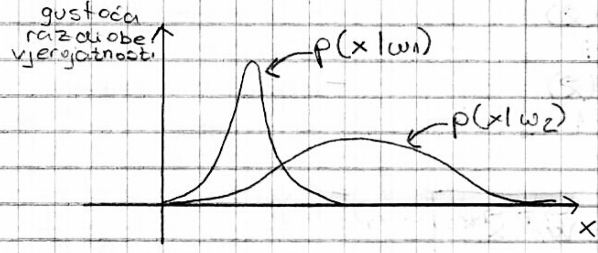
$p(x|\omega_j)$ - gustoća razdiobe vjerojatnosti

Funkcija gustoće razdiobe vjerojatnosti - funkcija $f(x)$ sa svojstvima:

1) $f(x) \geq 0 \quad , \forall x$

2) $\int_{-\infty}^{+\infty} f(x)\,dx = 1$

3) $\int_{x_1}^{x_2} f(x)\,dx = P\{x_1 < x < x_2\}$



$$P(\omega_j|x) = \frac{p(x|\omega_j)\cdot P(\omega_j)}{p(x)} \rightarrow \text{BAYESOVO PRAVILO}$$

a' posteriori vjerojatnost

$$p(x) = \sum_{j=1}^{N} p(x|\omega_j) P(\omega_j)$$

$$\boxed{\text{Ako } P(\omega_1|x) > P(\omega_2|x) \text{ odabiremo } \omega_1 \text{ inače } \omega_2}$$

→ Bayesovo decizijsko pravilo

Poopćenje

1) Upotreba više od jedne značajke
2) Više od dva stanja prirode
3) Uvođenje akcije umjesto samo odluke o stanju prirode
4) Funkcija gubitka $\lambda(\alpha_i|\omega_j)$

$\lambda(\alpha_i|\omega_j)$ - gubitak nastao poduzimanjem akcije α_i kada je stanje prirode ω_j

$$\boxed{R(\alpha_i|\vec{x}) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j) P(\omega_j|\vec{x})}$$

OČEKIVANI GUBITAK - nastaje kada poduzmemo akciju α_i a stanje prirode je ω_j

Bayesovo decizijsko pravilo

Da bi minimizirali ukupan rizik izračunavamo uvjetni rizik

$$R(\alpha_i|\vec{x}) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j) P(\omega_j|\vec{x})$$

za $i = 1, 2, \ldots, a$ i tada izabiremo akciju α_i za koju je $R(\alpha_i|\vec{x})$ MINIMUM

M = 2

$$R(\alpha_1|\vec{x}) = \lambda(\alpha_1|\omega_1) P(\omega_1|\vec{x}) + \lambda(\alpha_1|\omega_2) P(\omega_2|\vec{x})$$

$$= \{\lambda(\alpha_1|\omega_1) = \lambda_{11} \ ; \ \lambda(\alpha_1|\omega_2) = \lambda_{12}\}$$

$$= \lambda_{11} P(\omega_1|\vec{x}) + \lambda_{12} P(\omega_2|\vec{x})$$

$$R(\alpha_2|\vec{x}) = \lambda(\alpha_2|\omega_1) P(\omega_1|\vec{x}) + \lambda(\alpha_2|\omega_2) P(\omega_2|\vec{x})$$

$$= \lambda_{21} P(\omega_1|\vec{x}) + \lambda_{22} P(\omega_2|\vec{x})$$

PRAVILO: Odluči se za ω_1 AKO

$$R(\alpha_1|\vec{x}) < R(\alpha_2|\vec{x})$$

$$\lambda_{11} P(\omega_1|\vec{x}) + \lambda_{12} P(\omega_2|\vec{x}) < \lambda_{21} P(\omega_1|\vec{x}) + \lambda_{22} P(\omega_2|\vec{x})$$

$$\lambda_{11} P(\omega_1|\vec{x}) - \lambda_{21} P(\omega_1|\vec{x}) < \lambda_{22} P(\omega_2|\vec{x}) - \lambda_{12} P(\omega_2|\vec{x})$$

$$(\lambda_{11} - \lambda_{21}) P(\omega_1|\vec{x}) < (\lambda_{22} - \lambda_{12}) P(\omega_2|\vec{x})$$

$\lambda_{11} < \lambda_{21}$ $\qquad$ $\lambda_{22} < \lambda_{12}$

$\lambda_{11} - \lambda_{21} < 0 \ /\cdot(-1)$ $\qquad$ $\lambda_{22} - \lambda_{12} < 0 \ /\cdot(-1)$

$\lambda_{21} - \lambda_{11} > 0$ $\qquad$ $\lambda_{12} - \lambda_{22} > 0$

$$(\lambda_{22} - \lambda_{12}) P(\omega_2|\vec{x}) > (\lambda_{11} - \lambda_{21}) P(\omega_1|\vec{x})$$

$$\frac{(\lambda_{22} - \lambda_{12}) p(\vec{x}|\omega_2) P(\omega_2)}{\cancel{p(\vec{x})}} > \frac{(\lambda_{11} - \lambda_{21}) p(\vec{x}|\omega_1) P(\omega_1)}{\cancel{p(\vec{x})}}$$

$$(\lambda_{22} - \lambda_{12}) p(\vec{x}|\omega_2) P(\omega_2) > (\lambda_{11} - \lambda_{21}) p(\vec{x}|\omega_1) P(\omega_1) \quad /\cdot(-1)$$

$$(\lambda_{12} - \lambda_{22}) p(\vec{x}|\omega_2) P(\omega_2) < (\lambda_{21} - \lambda_{11}) p(\vec{x}|\omega_1) P(\omega_2)$$

$$\boxed{\frac{p(\vec{x}|\omega_1)}{p(\vec{x}|\omega_2)} > \underbrace{\frac{(\lambda_{12} - \lambda_{22}) P(\omega_2)}{(\lambda_{21} - \lambda_{11}) P(\omega_1)}}_{\text{prag nezavisan od } \vec{x}}} \rightarrow \text{Omjer vjerodostojnosti}$$

ako vrijedi odlučujemo se za ω_1

FUNKCIJA GUBITKA – simetrična ili nula-jedan funkcija

$$\lambda(\alpha_i | \omega_j) = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases}$$

$$R(\alpha_i | \vec{x}) = \sum_{j=1}^{c} \lambda(\alpha_i | \omega_j) P(\omega_j | \vec{x})$$

$$= \sum_{j \neq i} P(\omega_j | \vec{x})$$

$$= 1 - P(\omega_j | \vec{x})$$

M = C

$g_i(\vec{x})$, $i = 1, 2, \dots, c$ → skup diskriminantnih funkcija

Klasifikator razvrstava $\vec{x}$ u ω_i ako

$$g_i(\vec{x}) > g_j(\vec{x}) \quad \text{za sve } j \neq i$$

$g_i(\vec{x}) = -R(\alpha_i | \vec{x})$ → za opći slučaj s rizikom

$g_i(\vec{x}) = P(\omega_i | \vec{x})$ → za slučaj minimalne pogreške

$$g_i(\vec{x}) = P(\omega_i | \vec{x}) = \frac{p(\vec{x} | \omega_i) P(\omega_i)}{p(\vec{x})}$$

$$= p(\vec{x} | \omega_i) P(\omega_i)$$

$$= \ln p(\vec{x} | \omega_i) + \ln P(\omega_i)$$

$$P(\omega_i | \vec{x}) - P(\omega_j | \vec{x}) = 0$$

Decizijska ploha koja u višedimenzionalnom prostoru značajki dijeli područja R_i i R_j

– s jedne strane plohe razlika je pozitivna a s druge negativna

Klasificiraj $\vec{x}$ u ω_i ako $g_i(\vec{x}) > g_j(\vec{x}) \; \forall \; j \neq i$

Bayesova klasifikacija za normalne distribucije

$$P(\vec{x}|\omega_i) = \frac{1}{(2\pi)^{\ell/2} \, |\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu}_i)^T \Sigma_i^{-1} (\vec{x}-\vec{\mu}_i)\right)$$

$$\vec{\mu}_i = E[\vec{x}]$$ → srednja vrijednost razreda ω_i (očekivanje)

$$\Sigma_i = E\left[(\vec{x}-\vec{\mu}_i)(\vec{x}-\vec{\mu}_i)^T\right]$$ → kovarijacijska matrica

$$\mathcal{N}(\vec{\mu}, \Sigma)$$ → Gaussova distribucija

$$g_i(\vec{x}) \equiv \overset{\to \ln(\cdot)}{f}\left(P(\omega_i|\vec{x})\right)$$

$$= \ln\left(P(\omega_i|\vec{x})\right)$$

$$= \ln\left(p(\vec{x}|\omega_i)P(\omega_i)\right)$$

$$= \ln p(\vec{x}|\omega_i) + \ln P(\omega_i)$$

$$= \ln\left[\frac{1}{(2\pi)^{\ell/2}|\Sigma_i|^{1/2}}\right] + \ln\left[e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_i)^T\Sigma_i^{-1}(\vec{x}-\vec{\mu}_i)}\right] + \ln P(\omega_i)$$

$$= \ln 1 - \ln\left[(2\pi)^{\ell/2}|\Sigma_i|^{1/2}\right] - \frac{1}{2}(\vec{x}-\vec{\mu}_i)^T\Sigma_i^{-1}(\vec{x}-\vec{\mu}_i) + \ln P(\omega_i)$$

$$= -\ln(2\pi)^{\ell/2} - \ln|\Sigma_i|^{1/2} - \frac{1}{2}\left(\vec{x}^T\Sigma_i^{-1} - \vec{\mu}_i^T\Sigma_i^{-1}\right)(\vec{x}-\vec{\mu}_i) + \ln P(\omega_i)$$

$$= \underbrace{-\frac{\ell}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_i|}_{C_i} - \frac{1}{2}\left(\vec{x}^T\Sigma_i^{-1}\vec{x} - \vec{\mu}_i^T\Sigma_i^{-1}\vec{x} - \vec{x}^T\Sigma_i^{-1}\vec{\mu}_i + \vec{\mu}_i^T\Sigma_i^{-1}\vec{\mu}_i\right) + \ln P(\omega_i)$$

$$= C_i - \frac{1}{2}\vec{x}^T\Sigma_i^{-1}\vec{x} + \frac{1}{2}\vec{\mu}_i^T\Sigma_i^{-1}\vec{x} + \frac{1}{2}\vec{x}^T\Sigma_i^{-1}\vec{\mu}_i - \frac{1}{2}\vec{\mu}_i^T\Sigma_i^{-1}\vec{\mu}_i + \ln P(\omega_i)$$

$$\boxed{g_i(\vec{x}) = C_i - \frac{1}{2}\vec{x}^T\Sigma_i^{-1}\vec{x} + \frac{1}{2}\vec{\mu}_i^T\Sigma_i^{-1}\vec{x} + \frac{1}{2}\vec{x}^T\Sigma_i^{-1}\vec{\mu}_i - \frac{1}{2}\vec{\mu}_i^T\Sigma_i^{-1}\vec{\mu}_i + \ln P(\omega_i)}$$

Pojednostavljenja

1) Kovarijacijsa matrica ista za sve razrede
2) Kovarijacijska matrica je dijagonalna s jednakim elementima

Kovarijacijska matrica ista za sve razrede

$$\Sigma_i = \Sigma$$

$\vec{x}^T \Sigma \vec{x}$ → jednak za sve diskriminantne funkcije

c_i

→ Mogu se ispustiti

$$g_i(\vec{x}) = \frac{1}{2} \mu_i^T \overset{\Sigma}{\Sigma_i^{-1}} \vec{x} + \frac{1}{2} \vec{x}^T \overset{\Sigma}{\Sigma_i^{-1}} \vec{\mu}_i - \frac{1}{2} \vec{\mu}_i^T \overset{\Sigma}{\Sigma_i^{-1}} \vec{\mu}_i + \ln P(\omega_i)$$

$$g_i(\vec{x}) = (\Sigma^{-1} \vec{\mu}_i)^T \vec{x} + \left(\ln P(\omega_i) - \frac{1}{2} \vec{\mu}_i^T \Sigma^{-1} \vec{\mu}_i\right)$$

$$\vec{w}_i = \Sigma^{-1} \vec{\mu}_i \qquad w_{i0} = \ln P(\omega_i) - \frac{1}{2} \vec{\mu}_i^T \Sigma^{-1} \vec{\mu}_i$$

$$\boxed{g_i(\vec{x}) = \vec{w}_i \vec{x} + w_{i0}}$$

Kovarijacijska matrica dijagonalna s jednakim elementima

$$\Sigma = \sigma^2 I$$

- značajke koje določuju vektor značajki su međusobno nekorelirane i imaju jednaku varijancu

$$E[(x_i - \mu_i)(x_j - \mu_j)] = \sigma^2 \delta_{ij}$$

$$\begin{aligned} g_i(\vec{x}) &= \vec{w}_i \vec{x} + w_{i0} \\ &= \Sigma^{-1} \vec{\mu}_i \vec{x} + w_{i0} \\ &= \frac{1}{\sigma^2} \vec{\mu}_i \vec{x} + w_{i0} \end{aligned}$$

jer je $\Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma^2} & & & \\ & \frac{1}{\sigma^2} & & \\ & & \frac{1}{\sigma^2} & \\ & & & \ddots \end{bmatrix}$

$$g_i(\vec{x}) = \frac{1}{\sigma^2} \vec{\mu}_i \vec{x} + w_{i0}$$

Decizijska hiperravnina

$$g_{ij}(\vec{x}) = g_i(\vec{x}) - g_j(\vec{x}) = \vec{w}^T(\vec{x}-\vec{x}_0) = 0$$

$$\vec{w} = \vec{\mu}_i - \vec{\mu}_j$$

$$\vec{x}_0 = \frac{1}{2}(\vec{\mu}_i + \vec{\mu}_j) - \sigma^2 \ln\left(\frac{P(\omega_i)}{P(\omega_j)}\right)\frac{\vec{\mu}_i - \vec{\mu}_j}{\|\vec{\mu}_i - \vec{\mu}_j\|^2}$$

AKO $P(\omega_i) = P(\omega_j)$ ONDA hiperravnina prolazi kroz točku $\vec{x}_0$ tj. prolazi srednjom vrijednosti od $\vec{\mu}_i$ i $\vec{\mu}_j$

$$g_{ij}(\vec{x}) = 0 \Rightarrow \vec{w}^T(\vec{x}-\vec{x}_0) = \underbrace{(\vec{\mu}_i - \vec{\mu}_j)^T}_{\text{ortogonalan na decizijsku hiperravninu}}(\vec{x}-\vec{x}_0) = 0$$

AKO $P(\omega_i) < P(\omega_j)$ ONDA je hiperravnina bliža $\vec{\mu}_i$

AKO je σ^2 mala vrijednost u odnosu na $\|\vec{\mu}_i - \vec{\mu}_j\|$ položaj hiperravnine je neosjetljiv na vrijednosti $P(\omega_i)$ i $P(\omega_j)$

Kovarijacijska matrica nije dijagonalna

$$\vec{x}_0 = \frac{1}{2}(\vec{\mu}_i + \vec{\mu}_j) - \ln\left(\frac{P(\omega_i)}{P(\omega_j)}\right)\frac{\vec{\mu}_j - \vec{\mu}_i}{\|\vec{\mu}_i - \vec{\mu}_j\|^2_{\Sigma^{-1}}}$$

$$\boxed{\|\vec{x}\|_{\Sigma^{-1}} = (\vec{x}^T \Sigma^{-1} \vec{x})^{\frac{1}{2}}} \; \Sigma^{-1} \rightarrow \text{norma od } \vec{x}$$

Hiperravnina je ortogonalna na $\Sigma^{-1}(\vec{\mu}_i - \vec{\mu}_j)$

Klasifikator na temelju minimalne udaljenosti

Imamo izraz

$$g_i(\vec{x}) = -\frac{1}{2}(\vec{x}-\vec{\mu}_i)^T \Sigma_i^{-1}(\vec{x}-\vec{\mu}_i) + \ln P(\omega_i) + c_i$$ → zanemaruje se

Pretpostavka → svi su razredi jednako vjerojatni i imaju istu kovarijantnu matricu

Slijedi:
$$g_i(\vec{x}) = -\frac{1}{2}(\vec{x}-\vec{\mu}_i)^T \Sigma^{-1}(\vec{x}-\vec{\mu}_i)$$

SLUČAJ $\Sigma = \sigma^2 I$

Maksimum $g_i(\vec{x})$ podrazumijeva MINIMUM od

$$d_E = \|\vec{x}-\vec{\mu}_i\|$$

→ euklidska udaljenost

- vektor značajki klasificira se u razred u skladu s minimalnom vrijednosti Euklidske udaljenosti

SLUČAJ Σ - nije dijagonalna matrica

Maksimum $g_i(\vec{x})$ postiže se kada je Σ^{-1} norma minimalna

$$d_m = \left((\vec{x}-\vec{\mu}_i)^T \Sigma^{-1}(\vec{x}-\vec{\mu}_i)\right)^{\frac{1}{2}}$$

→ Mahalanobisova udaljenost

KARHUNEN-LOÈVA TRANSFORMACIJA (PCA)

I. PRISTUP

KL transformacija kao postupak minimizacije srednje kvadratne pogreške približnog zapisa uzorka

$\{\vec{e}_j\}$ → skup ortonormiranih vektora

Uzorak $\vec{x}$ može se zapisati kao

$$\vec{x} = \sum_{j=1}^{r} c_j \vec{e}_j$$

c_j i $j=1,2,\dots,r$ → koeficijenti reda

Ako uzorak $\vec{x}$ aproksimiramo pomoću $n<r$ članova reda dobivamo srednju kvadratnu pogrešku

$$\bar{\varepsilon}^2 = E\left\{\left|\vec{x} - \sum_{j=1}^{n} c_j \vec{e}_j\right|^2\right\}$$

E - operator matematičkog očekivanja

$$\vec{\varepsilon}^2 = E\left\{\left|\vec{x} - \sum_{j=1}^{n} c_j \vec{e}_j\right|^2\right\}$$

$$= E\left\{\left|\sum_{j=1} c_j \vec{e}_j - \sum_{j=1}^{n} c_j \vec{e}_j\right|^2\right\}$$

-uzimamo u obzir ortonormalnost jediničnih vektora

$$\vec{e}_j^T \vec{e}_i = \begin{cases} 1, & \text{ako } j=i \\ 0, & \text{ako } j\neq i \end{cases} \qquad *$$

Slijedi $\bar{\varepsilon}^2 = E\left\{\left|\sum_{j=n+1}^{r} c_j\right|^2\right\}$

$$= E\left\{\sum_{j=n+1}^{r} c_j^2\right\} \Rightarrow \bar{\varepsilon}^2 = E\left\{\sum_{j=n+1}^{r} c_j^2\right\}$$

$\vec{x} = \sum_{j=1}^{r} c_j \vec{e}_j \quad / \cdot \vec{e}_j^T$

$\vec{e}_j^T \vec{x} = \vec{e}_j^T \vec{e}_j \vec{c}_j$ → uvrštavamo $*$

$\vec{e}_j^T \vec{x} = 1 \cdot \vec{c}_j$

$$\boxed{\vec{c}_j = \vec{e}_j^T \vec{x}}$$

Izraz $c_j = \vec{e}_j^{\,T} \vec{x}$ uvrštavamo u $\overline{\varepsilon}^2$

$$\overline{\varepsilon}^2 = E\left\{\sum_{j=n+1}^{r} c_j^2\right\}$$

$$= E\left\{\sum_{j=n+1}^{r} \vec{e}_j^{\,T} \vec{x}\, \vec{e}_j \vec{x}^T\right\}$$

→ $\vec{e}_j$ je deterministički pa možemo zamijeniti redoslijed

$$= E\left\{\sum_{j=n+1}^{r} \vec{e}_j^{\,T} \vec{x}\vec{x}^T \vec{e}_j\right\}$$

$$= \sum_{j=n+1}^{r} \vec{e}_j^{\,T}\, E\{\vec{x}\vec{x}^T\}\, \vec{e}_j$$

$R = E\{\vec{x}\vec{x}^T\}$ → korelacijska matrica uzoraka

$$\boxed{\overline{\varepsilon}^2 = \sum_{j=n+1}^{r} \vec{e}_j^{\,T} R \vec{e}_j}$$

Tražimo da srednja kvadratna pogreška bude minimalna tj. tražimo ortonormalne vektore $\vec{e}_j$ (koordinatni sustav) koji minimiziraju pogrešku

Lagrangeova funkcija:

$$\mathcal{L}(\vec{e}_j) = \sum_{j=n+1}^{r} \vec{e}_j^{\,T} R \vec{e}_j - \sum_{j=n+1}^{r} \lambda_j (\vec{e}_j^{\,T}\vec{e}_j - 1)$$

$$\frac{\partial \mathcal{L}}{\partial \vec{e}_j} = \sum_{j=n+1}^{r} 2\vec{e}_j R - \sum_{j=n+1}^{r} \lambda_j \cdot 2\vec{e}_j$$

$$= \sum_{j=n+1}^{r} 2\vec{e}_j (R - \lambda_j I) = 0 \quad /:2$$

$$\Rightarrow (R - \lambda_j I)\vec{e}_j = 0$$

$$R\vec{e}_j = \lambda_j I \vec{e}_j$$

$$\boxed{R\vec{e}_j = \lambda_j \vec{e}_j}$$

KL koordinatne osi → svojstveni vektori korelacijske matrice $R = E\{\vec{x}\vec{x}^T\}$

→ ortonormirani su → koeficijenti KL reda su međusobno nekorelirani

$R = E\{\vec{c}\,\vec{c}^T\}$ - dijagonalna matrica
- elementi jednaki svojstvenim vrijednostima korelacijske matrice $R = E\{\vec{x}\,\vec{x}^T\}$

$$\bar{\varepsilon}^2 = \sum_{j=n+1}^{r} \vec{e_j}^T R \vec{e_j}$$

$$= \{R\vec{e_j} = \lambda_j \vec{e_j}\}$$

$\vec{e_j}^T \vec{e_j} = 1$

$$= \sum_{j=n+1}^{r} \vec{e_j}^T \lambda_j \vec{e_j} = \sum_{j=n+1}^{r} \lambda_j$$

$$\boxed{\bar{\varepsilon}^2 = \sum_{j=n+1}^{r} \lambda_j}$$

$$A = \begin{bmatrix} \vec{e_1}^T \\ \vec{e_2}^T \\ \vdots \\ \vec{e_n}^T \end{bmatrix}$$

→ Transformacijska matrica
- sastoji se od prvih n vektora koordinatnih osi $\vec{e_j}$ uređenih po padajućem redosljedu svojstvenih vrijednosti matrice

$$\boxed{\vec{y} = A \cdot \vec{x}}$$

→ Vektor uzoraka

Raspoznavanje uzoraka - umjesto korelacijske matrice R koristi se kovarijacijska matrica K

$$K\vec{e_j} = \lambda_j \vec{e_j}$$

$$K = E\{(\vec{x}-\vec{m})(\vec{x}-\vec{m})^T\}$$

$$\boxed{\bar{\varepsilon}^2 = \sum_{j=n+1}^{r} E\{c_j\}\vec{e_j}}$$

II. PRISTUP

KL transformacija kojom se maksimizira varijanca

- aproksimacija vektora $\vec{x}$ gradi se iz $n<r$ komponenti koje imaju najveću varijancu
    - srednja kvadratna pogreška između $\vec{x}$ i aproksimiranog $\vec{x}$ jednaka je sumi varijanci komponenti koje su eliminirane iz $\vec{x}$

PRETPOSTAVKA – slučajni vektor $\vec{x}$ ima srednju vrijednost ϕ:

$$E\{\vec{x}\} = \vec{0}$$

E → operator matematičkog očekivanja

- $\vec{x}$ nema srednju vrijednost $\vec{0}$? → srednju vrijednost oduzimamo od $\vec{x}$ prije daljnjeg postupka

$\vec{x}$ – vektor značajki dimenzije r

$\vec{e}$ – jedinični vektor dimenzije r

- Projiciramo $\vec{x}$ na $\vec{e}$

$$A = \vec{x}^T\vec{e} = \vec{e}^T\vec{x}$$

ograničenje: $\|\vec{e}\| = (\vec{e}^T\vec{e})^{\frac{1}{2}} = 1$

$$E[A] = E[\vec{e}^T\vec{x}] = \vec{e}^T E[\vec{x}] = 0$$

A → projekcija – slučajna varijabla sa srednjom vrijednosti i varijancom u kojima se zrcali statistika slučajnog vektora $\vec{x}$

Varijanca od A:

$$\begin{aligned}
\sigma^2 = E[A^2] &= E[\vec{e}^T\vec{x} \cdot \vec{e}\vec{x}^T] \\
&= E[(\vec{e}^T\vec{x})(\vec{x}^T\vec{e})] \\
&= \vec{e}^T E[\vec{x}\vec{x}^T]\vec{e} \\
&= \vec{e}^T R \vec{e}
\end{aligned}$$

$R^T = R \Rightarrow$ simetrična je

$$\boxed{\sigma^2 = \vec{e}^T R \vec{e}} \qquad \boxed{R = E[\vec{x}\vec{x}^T]}$$

→ korelacijska matrica slučajnog vektora $\vec{x}$

$$\boxed{\Psi(\vec{e}) = \sigma^2 = \vec{e}^T R \vec{e}}$$

CILJ – pronaći jedinične vektore $\vec{e}$ uzduž kojih $\Psi(\vec{e})$ ima lokalne minimume ili maksimume (ekstreme)

OPTIMIZACIJA – Tražimo max vrijednost uz ograničenje euklidske norme vektora:

$$\boxed{\vec{e}^T \cdot \vec{e} = 1}$$

$$\mathcal{L}(\vec{e}) = \vec{e}^T R \vec{e} - \lambda(\vec{e}^T \vec{e} - 1)$$

$$\frac{\partial \mathcal{L}}{\partial \vec{e}} = 2\vec{e}R - \lambda \cdot 2\vec{e} = 2R \cdot \vec{e} - 2\lambda \cdot \vec{e} = 0 \quad /:2$$

$$R\vec{e} - \lambda\vec{e} = 0$$

$$(R - \lambda I)\vec{e} = 0$$

$$\downarrow$$

$$\boxed{R\vec{e} = \lambda\vec{e}}$$

KLASIFIKACIJA UZORAKA POMOĆU FUNKCIJA UDALJENOSTI

- Djelotvorna metoda kada razredi pokazuju SVOJSTVO GRUPIRANJA

Klasifikacija pomoću najmanje udaljenosti

MJERA UDALJENOSTI - funkcija koja zadovoljava uvjete:

1) $D(\vec{x}_k, \vec{x}_e) = 0$ ako $\vec{x}_k = \vec{x}_e$

$$\left.\begin{array}{l} D(\vec{x}_k, \vec{x}_e) = \emptyset \\ D(\vec{x}_k, \vec{x}_e) > 0 \end{array}\right\} \text{ako } \vec{x}_k = \vec{x}_e$$

2) $D(\vec{x}_k, \vec{x}_e) = D(\vec{x}_e, \vec{x}_k)$

3) $D(\vec{x}_k, \vec{x}_e) \leq D(\vec{x}_k, \vec{x}_j) + D(\vec{x}_j, \vec{x}_e)$

Euklidska udaljenost

$$D = \|\vec{x}_k - \vec{x}_e\| = \sqrt{\sum_{i=1}^{n} (x_{ki} - x_{ei})^2}$$

- invarijantna na translaciju i rotaciju

Minkovski udaljenost

$$D(\vec{x}_k, \vec{x}_e) = \left(\sum_{j=1}^{n} |x_{kj} - x_{ej}|^s \right)^{\frac{1}{s}}$$

$s = 1$

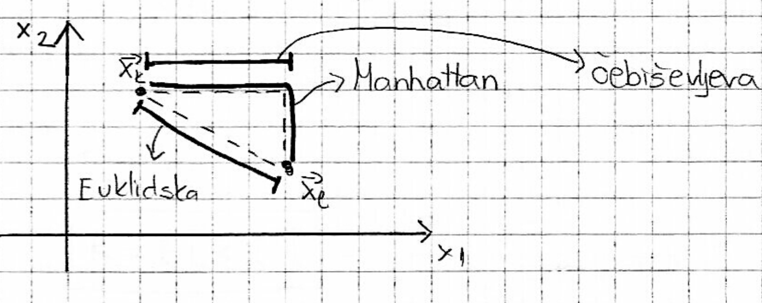
$$D(\vec{x}_k, \vec{x}_e) = \sum_{j=1}^{n} |x_{kj} - x_{ej}|$$ → Manhattan ili city-block udaljenost

$s = 2$

$$D(\vec{x}_k, \vec{x}_e) = \left(\sum_{j=1}^{n} |x_{kj} - x_{ej}|^2 \right)^{\frac{1}{2}}$$ → Euklidska udaljenost

$s = \infty$

$$D(\vec{x}_k, \vec{x}_e) = \max\{|x_{kj} - x_{ej}|\}$$ → Čebiševljeva udaljenost



Težinska udaljenost Minovskog

$$D(\vec{x}_k, \vec{x}_l) = \left(\sum_{j=1}^{n} w_j |x_{kj} - x_{lj}|^s \right)^{\frac{1}{s}}$$

$w_j \rightarrow$ težina pojedine značajke

$w_j \geq 1$

Mahalanobisova udaljenost

$$D(\vec{x}_k, \vec{x}_l) = (\vec{x}_k - \vec{x}_l)^T \, C^{-1} \, (\vec{x}_k - \vec{x}_l)$$

$C \rightarrow$ kovarijacijska matrica dobivena iz skupa za učenje