b) Postupak učenja za kriterijsku funkciju:

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

Parcijalna derivacija kriterijske funkcije:

$D_x(u \cdot v) = u \cdot v' + v \cdot u'$

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

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

$$(\vec{w}^T\vec{x}) \cdot \vec{x} \cdot \mathrm{sgn}(\vec{w}^T\vec{x}) = |\vec{w}^T\vec{x}| \cdot \vec{x}$$

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

Postupak učenja:

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

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

λ – korekcijski faktor

- početna vrijednost $\vec{w}$ različita od $\vec{0}$
- ako je $\lambda > 1$ promatrani uzorak se pravilno razvrstava nakon (svake) korekcije vektora $\vec{w}$
- postoji dokaz da postupak učenja konvergira (za lin. separatibilne razrede) za $0 < \lambda < 2$;

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

varijacija u zavisnosti od izbora korekcijskog intrementa c

1) algoritam sa stalnim (čvrstim) prirastom (fixed-increment)
2) algoritam s apsolutnom korekcijom (absolute-correction)
3) algoritam s djelomičnom korekcijom (fractional-correction)

1) $c = \text{konst.} > 0$

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

Ako je $\vec{w}^T(k)\cdot\vec{x}(k) \leq 0$ izabire se c tako da vrijedi:

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

c je najmanji cijeli broj veći od

3)

c je izabran tako da je

$|\vec{w}^T(k)\vec{x}(k) - \vec{w}(k+1)\vec{x}(k)|$ dio λ

od $|\vec{w}^T(k)\vec{x}(k)|$:

$$|\vec{w}^T(k)\vec{x}(k) - \vec{w}^T(k+1)\vec{x}(k)| = \lambda\, |\vec{w}^T(k)\cdot\vec{x}(k)|$$

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

$$|\vec{w}^T(k)\vec{x}(k) - (\vec{w}^T(k) + c\vec{x}^T(k))\vec{x}(k)| = \lambda\, |\vec{w}^T(k)\vec{x}(k)|$$

Dobivamo:

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

- zahtijeva se da početna vrijednost vektora $\vec{w}$ bude različita od $\vec{0}$

- za $\lambda > 1$ uzorak se pravilno razvrstava nakon svakog ugađanja tež. vektora

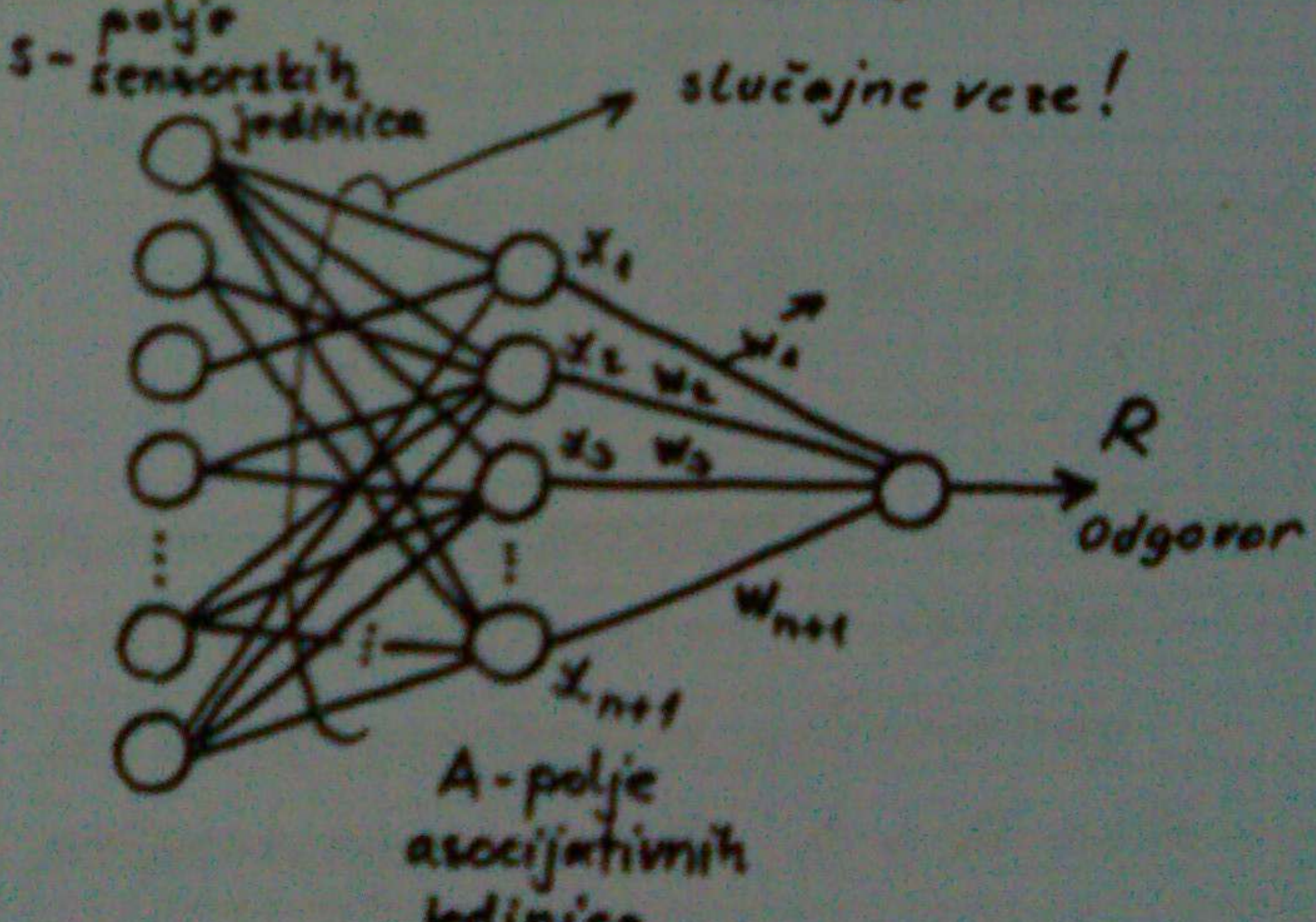
- za $0 < \lambda < 2$ postupak konvergira

PERCEPTRON (Rosenblatt, 1957)

- bionika (biologija – elektronika): grana znanosti koja traži polazne točke za rješenje tehničkih problema u uzorima što ih čovjeku pruža sama priroda

  : primjena bioloških koncepata za izgradnju elektroničkih naprava

Perceptron – klasificira uzorke u jedan od dva razreda



edinica u polju A generira izlaz različit od 0 ako je dovoljan broj senzorskih jedinica priključenih na A jedinicu pobuđen (aktivan).

x_i - odgovor i-te asocijativne jedinice

W_i - težinska vrijednost

Odgovor R proporcionalan je sumi odgovora A jedinica; odgovori su pomnoženi težinskim vrijednostima W_i :

$$R = \sum_{i=1}^{n+1} W_i X_i = \vec{W}^T \cdot \vec{X}$$

Pravilo klasifikacije:

- ako je $R > 0$ nepoznati se uzorak razvrstava u ω_1
- ako je $R < 0$ nepoznati se uzorak razvrstava u ω_2

Perceptron - klasifikator uzoraka u dva razreda - ako su razredi linearno separatibilni

- modifikacija perceptrona za klasifikaciju uzoraka u $M>2$ razreda:

M - jedinica R

$R_1, R_2, \dots, R_M$

Pravilo:

uzorak se razvrstava u razred ω_i akko $R_i > R_j$ za sve $i \neq j$.

Osnovni model perceptrona može se raširiti na nelinearne decizijske funkcije umetanjem nelinearnog pretprocesora između A i R polja

- višeslojni perceptron MLP

ALGORITAM PERCEPTRONA

(Reward - Punishment Concept)

Algoritam učenja perceptrona:

- Zadana su dva skupa uzoraka za učenje koji pripadaju razredu ω_1 i ω_2

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

- k-ti korak učenja :

• Ako $\vec{x}(k) \in \omega_1$ i

$\vec{w}^T(k)\vec{x}(k) \leq 0$ zamijeni $\vec{w}(k)$ s

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

gdje je c korekcijski faktor

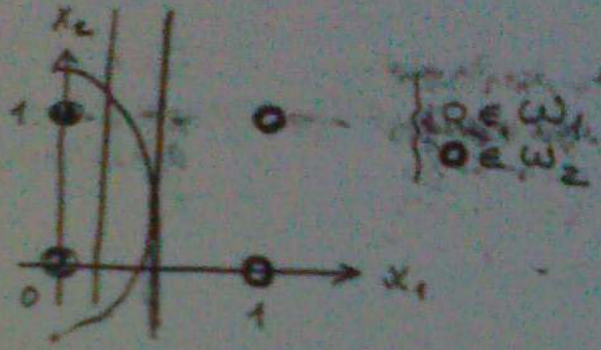
• Ako $\vec{x}(k) \in \omega_2$ i

$\vec{w}^T(k)\vec{x}(k) \geq 0$ zamijeni

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

• U drugim slučajevima ostavi $\vec{w}(k) \Rightarrow \vec{w}(k+1) = \vec{w}(k)$

PRIMJER



$\omega_1 : \{ (0,0,1)', (0,1,1)' \}$

$\omega_2 : \{ (1,0,1)', (1,1,1)' \}$

$c = 1 \qquad ; \qquad \vec{w}(1) = \vec{0}$

1.k. $\vec{w}^T(1) \cdot \vec{x}(1) = (0,0,0) \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 0$

$\vec{w}(2) = \vec{w}(1) + \vec{x}(1) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

2.k. $\vec{w}^T(2) \cdot \vec{x}(2) = (0,0,1) \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = 1$

$\vec{w}(3) = \vec{w}(2) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

3.k. $\vec{w}^T(3) \cdot \vec{x}(3) = (0,0,1) \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = 1$

$\vec{w}(4) = \vec{w}(3) - \vec{x}(3) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}$

4.k. $\vec{w}^T(4) \vec{x}(4) = (-1,0,0) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = -1$

LMSE (Least-Mean-Square-Error)
(Ho-Kashyap algoritam, 1966.)

- Algoritam perceptrona (i njegove varijacije) konvergira kada su razredi linearno razdvojivi
- Ne može se sa sigurnosti utvrditi da li duga sekvenca učenja znači i da su razredi linearno seperatibilni

- Algoritam koji brže konvergira i ima ugrađen mehanizam za detekciju slučaja kad razredi nisu linearno separabilni

Umjesto nalaženja vektora $\vec{w}$ tako da vrijedi $[X]\vec{w} > \vec{0}$

tražit ćemo vektore $\vec{w}$ i $\vec{b}$ tako da vrijedi

$$[X]\vec{w} = \vec{b},$$

gdje je $\vec{b} = (b_1, b_2, \ldots, b_N)^T$ takav da su sve njegove komponente b_i, $i = 1, 2, \ldots N$ pozitivne.

Kriterijska funkcija:

$$J(\vec{w}, \vec{x}, \vec{b}) = \frac{1}{2} \sum_{j=1}^{N} (\vec{w}^T \vec{x}_j - b_j)^2 = \frac{1}{2} \| [X]\vec{w} - \vec{b} \|^2,$$

gdje $\|[X]\vec{w} - \vec{b}\|$ označava veličinu vektora $([X]\vec{w} - \vec{b})$.

Kriterijska funkcija $J(\vec{w}, \vec{x}, \vec{b})$ postiže minimum kada je $[X]\vec{w} = \vec{b}$

Umjesto traženja vektora $\vec{w}$ koji zadovoljava nejednadžbu tražimo vektore $\vec{w}$ i $\vec{b}$ tako da je zadovoljena jednadžba:

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

- Možemo obje varijable $\vec{w}$ i $\vec{b}$ upotrijebiti u minimizacijskoj proceduri
- Očekujemo poboljšani postupak (brži) konvergencije
- Gradijenti:

$$\frac{\partial J}{\partial \vec{w}} = [X]^T([X]\vec{w} - \vec{b})$$

$$\frac{\partial J}{\partial \vec{b}} = -([X]\vec{w} - \vec{b})$$

...ktor $\vec{w}$ nije ograničen i možemo postaviti:

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

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

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

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

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

$$\boxed{\vec{w} = [X]^{\#}\vec{b}} \qquad (**)$$

$[X]^{\#}$ - generalized inverse of $[X]$

$$[X]^{\#} = ([X]^T[X])^{-1}[X]^T$$

Vektor $\vec{b}$ je ograničen - to je vektor sa svim pozitivnim komponentama

vektor $\vec{b}$ treba tako mijenjati da uvijek bude zadovoljeno: $b_i > 0$ za $i = 1, 2, \ldots, N$

To se postiže na ovaj način:

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

gdje je

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

converges for $0 < \lambda < 2$.

5.3 DERIVATION OF PATTERN CLASSIFICATION ALGORITHMS

$$\frac{\partial J}{\partial \vec{b}} = -([X]\vec{w} - \vec{b})$$

Jednadžbu (*) zapišimo u vektorskom obliku:

$$\delta \vec{b}(k) = c\,[\overbrace{[X]\vec{w}(k) - \vec{b}(k)}^{\vec{e}(k)} + |\overbrace{[X]\vec{w}(k) - \vec{b}(k)}^{\vec{e}(k)}|]$$

gdje

$|[X]\vec{w}(k) - \vec{b}(k)|$ označava apsolutnu vrijednost svake komponente vektora $[X]\vec{w}(k) - \vec{b}(k)$

Iz jednadžbi (**) i (***) dobivamo:

$$\begin{aligned}
\vec{w}(k+1) &= [X]^{\#}\vec{b}(k+1) \\
&= [X]^{\#}[\vec{b}(k) + \delta\vec{b}(k)] \\
&= [X]^{\#}\vec{b}(k) + [X]^{\#}\delta\vec{b}(k) \\
&= \vec{w}(k) + [X]^{\#}\delta\vec{b}(k)
\end{aligned}$$

Označimo

$$\vec{e}(k) = [X]w(k) - \vec{b}(k)$$

Algoritam:

$$\vec{w}(1) = [X]^{\#}\,\vec{b}(1)$$

$\vec{b}(1)$ proizvoljan uz uvjet

$b_i > 0$ za sve i

$$\vec{e}(k) = [X]\,\vec{w}(k) - \vec{b}(k)$$

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

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

$\vec{w}(k+1)$ može se izračunati i kao:

$$\vec{w}(k+1) = [X]^{\#}\,\vec{b}(k+1)$$

Napomena: $|\vec{e}(k)|$ označava vektor čije su komponente apsolutne vrijednosti komponenti vektora $\vec{e}(k)$

Važno svojstvo algoritma:

- Ako su u bilo kojem iteracijskom koraku <u>sve</u> komponente vektora $\vec{e}(k)$ ne pozitivne (ali ne jednake 0) razredi su linearno neseparabilni!
- Kada je $\vec{e}(k) = \vec{0}$ $\vec{w}(k)$ je rješenje!

Osnovna značajka algoritma:

- Ako rješenje za nejednadžbu

$[X]\vec{w} > \vec{0}$ postoji,

postupak konvergira za

$0 < c \leq 1$

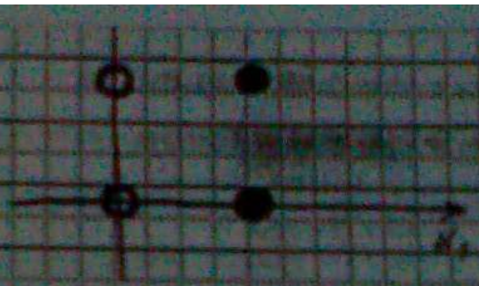
- Razredi nisu separatibilni s linearnom decizijskom funkcijom ako u bilo kojem koraku postupka nisu sve komponente vektora $\vec{e}(k)$ pozitivne ili jednake nuli.

- Ako je $\vec{e}(k) = \vec{0}$ znači da vrijedi $[X]\vec{w}(k) = \vec{b}(k)$, t.j. $\vec{w}(k)$ je rješenje

- Postupak brže konvergira negoli postupak perceptrona i nalazi "bolje" granice između razreda.

$\omega_1 : \{ (0,0)^T, (0,1)^T \}$

$\omega_2 : \{ (1,0)^T, (1,1)^T \}$



+ proširujemo vektor i pomnožimo uzorke iz razreda ω_2 sa -1, te tvorimo matricu $[X]$

$$[X] = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & 0 & -1 \\ -1 & -1 & -1 \end{bmatrix}_{4\times 3}$$

Potražimo

$$[X]^{\#} = \left([X]^T [X] \right)^{-1} [X]^T$$

$$[X]^{\#} = \frac{1}{2} \begin{bmatrix} -1 & -1 & -1 & -1 \\ -1 & 1 & 1 & -1 \\ 3/2 & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Neka je $\vec{b}(1) = (1,1,1,1)^T$ i $c = 1$; $b_i > 0$

$$\vec{w}(1) = [X]^{\#} \vec{b}(1) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$$

$$\vec{e}(1) = [X] \vec{w}(1) - \vec{b}(1) = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

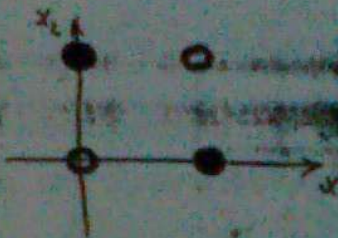
$\vec{w}(1) = (-2, 0, 1)^T$ traženo rješenje

$$d(\vec{x}) = -2x_1 + 1$$

PRIMJER:

$\omega_1 : \{(0,0)^T, (1,1)^T\}$

$\omega_2 : \{(0,1)^T, (1,0)^T\}$



Razredi nisu linearno separabilni!

(XOR problem)

$\vec{b}(1) = (1,1,1,1)^T$ i $c = 1$

$$[X] = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & -1 & -1 \\ -1 & 0 & -1 \end{bmatrix}$$

izračunajmo $[X]^{\#} = ([X]^T[X])^{-1}[X]^T$

$$[X]^{\#} = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \end{bmatrix}$$

$$\vec{w}(1) = [X]^{\#}\,\vec{b}(1) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\vec{e}(1) = [X]\cdot\vec{w}(1) - \vec{b}(1)$$

$$= \begin{pmatrix} -1 \\ -1 \\ -1 \\ -1 \end{pmatrix}$$

$\vec{e}(1)$ negativno $[X]\vec{w} > 0$ nema rješenja

...can be proved in a variety of ways. The proof ...ted above, however, is one of the most concise.

UČENJE ZA SLUČAJ M>2 RAZREDA

• Problem učenja koeficijenata linearne decizijske funkcije za $M>2$ razreda skupa za učenje može se riješiti poopćenim postupkom perceptrona ali i postupkom Ho-Kashyapa.

Podsjetimo se: 3. slučaja

<u>1. slučaj:</u> svaki od $M>2$ razreda separatibilan je od ostalih razreda jednom decizijskom (hiper)ravninom;
(Traži se $M>2$ decizijskih funkcija)

<u>2. slučaj:</u> svaki je razred separatibilan od svakog drugog razreda;
(Traži se $M(M-1)/2$ decizij. funkcija)

<u>3. slučaj:</u> postoji M decizijskih funkcija: $d_1, d_2, \ldots, d_M$;
$\vec{x} \in \omega_i$ ako
$d_i(\vec{x}) > d_j(\vec{x})$; $i \neq j$

3. slučaj → **Poopćeni algoritam perceptrona**

- istodobno određujemo težinskih koeficijenata: $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_M$

- M razreda uzoraka za učenje:

$$\omega_1, \omega_2, \ldots, \omega_M$$

U k-tom koraku postupka:

$\vec{x}(k)$ – uzorak iz skupa za učenje

$$\vec{x}(k) \in \omega_i$$

Računamo M decizijskih funkcija:

$$d_j(\vec{x}(k)) = \vec{w}_j^T(k)\vec{x}(k); \quad j = 1, 2, \ldots, M$$

Ako je

$$d_i(\vec{x}(k)) > d_j(\vec{x}(k)); \quad j = 1, 2, \ldots, M \quad j \neq i$$

težinski vektori se ne "popravljaju":

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

Ako je

$$d_i(\vec{x}(k)) \leq d_l(\vec{x}(k))$$

popravljaju se vrijednosti težinskih vektora :

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

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

$$\vec{w}_j(k+1) = \vec{w}_j(k) \quad \text{za}$$

$$j = 1, 2, \dots, M \quad j \neq i, \; j \neq l$$

c - pozitivna konstanta

Postupak konvergira u konačnom broju ponavljanja za proizvoljno izabrane početne vrijednosti vektora težinskih koeficijenata

$$\vec{w}_j(1); \; j = 1, 2, \dots, M$$

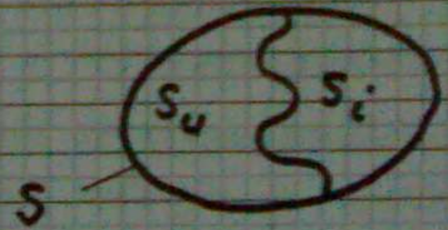
ako su razredi linearno separatibilni.

Skup uzoraka za učenje i skup uzoraka za ispitivanje - metode ispitivanja

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

Važna pretpostavka: 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

postupak učenja decizijskih funkcija potencijalnim funkcijama

Dosadašnji postupci → unaprijed smo pretpostavili oblik decizijske funkcije

Potencijalne funkcije dopuštaju učenje decizijskih funkcija bez pretpostavke o konačnom obliku decizijske funkcije

Decizijske funkcije mogu se generirati uporabom potencijalnih funkcija na temelju uzoraka $\vec{X}_k$, $k=1,2,\dots$,

Potencijalna funkcija za neki uzorak $\vec{X}$ može se opisati kao:

$$K(\vec{X},\vec{X}_k)=\sum_{i=1}^{\infty}\lambda_i^2\varphi_i(\vec{X})\varphi_i(\vec{X}_k)$$

gdje su $\varphi_i(\vec{X})$ $i=1,2,\dots$ ortonormalne funkcije i λ_i realni brojevi $\neq 0$ izabrani tako da je $K(\vec{X},\vec{X})$ ograničena

za $\vec{X}_k \in \omega_1 \cup \omega_2$.

Decizijska funkcija $d(\vec{X})$ konstruira se iz sekvence potencijalnih funkcija $K(\vec{X},\vec{X}_1)$, $K(\vec{X},\vec{X}_2),\dots$ koje odgovaraju slijedu uzoraka za učenje $\vec{X}_1, \vec{X}_2,\dots$ i koji su predstavljeni stroju tijekom procesa učenja.

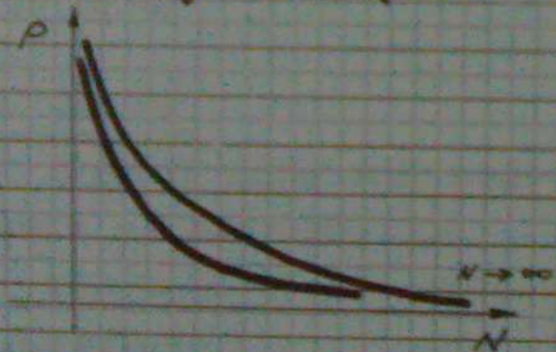
Potencijalne funkcije:

Kontinuirane funkcije koje zadovoljavaju sljedeće uvjete:

1. $K(\vec{X},\vec{X}_k)=K(\vec{X}_k,\vec{X})$
   - simetrična
2. $K(\vec{X},\vec{X}_k)$ ima maksimum pri $\vec{X}=\vec{X}_k$
3. $K(\vec{X},\vec{X}_k)\rightarrow 0$ ako udaljenost $\|\vec{X}-\vec{X}_k\|\rightarrow\infty$

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.

... 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 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 $N \to \infty$

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

$$N \doteq 5 * 110 * 580 = 319\,000 \text{ !!!}$$

slika lica

nih 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

Poopćene (Linearne) decizijske funkcije

- Složenost granica :

Linearne $\longrightarrow$ vrlo nelinearne!

- vrlo nelinearne ?

Rješenje :

Poopćeni oblik (linearne) decizijske funkcije :

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

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

$\{f_i(\vec{x})\}$, $i = 1, 2, \ldots, k$

$f_{k+1} = 1$

Pazi: $f_i(\vec{x})$

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

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

$d(\vec{x})$ je linearna funkcija po $\vec{x}^*$

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

$$\vec{x}^* = \begin{bmatrix} f_1(\vec{x}) \\ f_2(\vec{x}) \\ \vdots \\ f_k(\vec{x}) \\ 1 \end{bmatrix} \qquad \vec{w} = (w_1, w_2, \ldots w_k, w_{k+1})$$

Decizijska funkcija $d(\vec{x})$ može se promatrati kao linearna funkcija u k -dimenzionalnom prostoru!

Transformacija :

$$\vec{x} \Big| \longrightarrow \vec{x}^* \Big|$$

n-dimenzija $\qquad$ k-dimenzija

Vrijedi :

$$k > n$$

$d(\vec{x})$ se često naziva i "virtualno linearna"

Kako izabrati $\{f_i(\vec{x})\}_{i=1}^{k}$?

Mogućnost:

$\{f_i(\vec{x})\}$ $f_i(\vec{x})$ u obliku polinoma

- Najjednostavniji slučaj: linearna funkcija

$$\vec{x} = (x_1, x_2, \dots, x_n)^T$$

$$f_i(\vec{x}) = x_i \qquad \vec{x} = \vec{x}^*$$

$$k = n$$

- Polinom drugog stupnja:

Npr. $n=2$ $\vec{x} = (x_1, x_2)^T$

$$d(\vec{x}) = w_{11}x_1^2 + w_{12}x_1x_2 + w_{22}x_2^2 + w_1x_1 + w_2x_2 + w_3$$

$$d(\vec{x}) = \vec{W}^T\vec{X}^*$$

$$\vec{X}^* = (x_1^2, x_1x_2, x_2^2, x_1, x_2, 1)^T$$

Opći slučaj kvadratnog oblika:

$$d(\vec{x}) = \underbrace{\sum_{j=1}^{n} w_{jj}x_j^2}_{n\text{ izraza}} + \underbrace{\sum_{j=1}^{n-1}\sum_{k=j+1}^{n} w_{jk}x_jx_k}_{\frac{n(n-1)}{2}\text{ izraza}} + \underbrace{\sum_{j=1}^{n} w_jx_j}_{n\text{ izraza}} + \underbrace{w_{n+1}}_{1\text{ izraz}}$$

Kvadratnu decizijsku funkciju možemo promatrati kao linearnu ("linear by virtue") funkciju s

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

$$\{f_1(\vec{x}) \dots f_n(\vec{x})\} = \{x_1^2, x_2^2, \dots, x_n^2\}$$

$$\{f_{n+1}(\vec{x}), \dots, f_{2n}(\vec{x})\} = \{x_1, x_2, \dots, x_n\}$$

$$\{f_{2n+1}(\vec{x}), \dots, f_k(\vec{x})\} = \{x_1x_2, x_1x_3, \dots, x_{n-1}x_n\}$$

- Decizijske funkcije višeg reda (polinomi stupnja $r > 2$) možemo promatrati kao linearne funkcije

$$\frac{(n+r)!}{r!\,n!} \text{ varijabli}$$

n - dimenzionalnost vektora $\vec{x}$

r - stupanj polinoma

...unkcije iz skupa $\{f_i(\vec{x})\}$ određene su

s:

$$f_i(\vec{x}) = x_{p_1}^{s_1} x_{p_2}^{s_2} \cdots x_{p_r}^{s_r}$$

$$p_1, p_2, \dots p_r = 1, 2, \dots n$$

$$s_1, s_2, \dots s_r = 0, 1$$

Decizijsku funkciju u obliku polinoma r-tog stupnja možemo zapisati u rekurzivnom obliku:

$$\boxed{d^r(\vec{x}) = \left(\sum_{p_1=1}^{n} \sum_{p_2=p_1}^{n} \cdots \sum_{p_r=p_{r-1}}^{n} w_{p_1 p_2 \cdots p_r} x_{p_1} x_{p_2} \cdots x_{p_r} \right) + d^{r-1}(\vec{x}) \quad ; \quad d^0(\vec{x}) = w_{n+1}}$$

Primjer:

za $r=2$; $n=2$:

$$d^2(\vec{x}) = \sum_{p_1=1}^{2} \sum_{p_2=p_1}^{2} w_{p_1 p_2} x_{p_1} x_{p_2} + d^1(\vec{x})$$

$$d^1(\vec{x}) = \sum_{p_1=1}^{2} w_{p_1} x_{p_1} + d^0(\vec{x})$$

$$d^1(\vec{x}) = w_1 x_1 + w_2 x_2 + w_3$$

$$d^2(\vec{x}) = w_{11} x_1^2 + w_{12} x_1 x_2 + w_{22} x_2^2 + d^1(\vec{x})$$

$$d^2(\vec{x}) = w_{11} x_1^2 + w_{12} x_1 x_2 + w_{22} x_2^2 + w_1 x_1 + w_2 x_2 + w_3$$

$$\vec{W}^T = \begin{bmatrix} w_{11} \\ w_{12} \\ w_{22} \\ w_1 \\ w_2 \\ w_3 \end{bmatrix} \qquad \vec{x}^{*} = [x_1^2, x_1 x_2, x_2^2, x_1, x_2, 1]^T$$

Primjer: $r=3$; $n=2$

$$d^3(\vec{x}) = \left(\sum_{p_1=1}^{2} \sum_{p_2=p_1}^{2} \sum_{p_3=p_2}^{2} w_{p_1 p_2 p_3} x_{p_1} x_{p_2} x_{p_3} \right) + d^2(\vec{x}) ;$$

$$d(\vec{x}) = w_{111} x_1^3 + w_{112} x_1^2 x_2 + w_{122} x_1 x_2^2 + w_{222} x_2^3 + d^2(\vec{x})$$

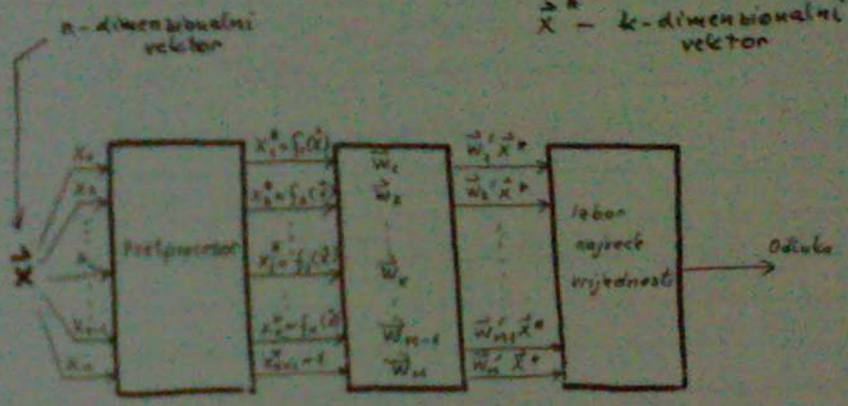
$$d(\vec{x}) = w_{111} x_1^3 + w_{112} x_1^2 x_2 + w_{122} x_1 x_2^2 + w_{222} x_2^3 + w_{11} x_1^2 + w_{12} x_1 x_2 + w_{22} x_2^2 + w_1 x_1 + w_2 x_2 + w_3$$

Cijena koju plaćamo za linearizaciju:

- broj koeficijenata (težinskih faktora) za funkciju r-tog stupnja i različite n:

n \ r	1	2	3	...	6	10
1	2	3	4		7	11
2	3	6	10		28	66
3	4	10	20		84	286
⋮						
9	10	55	220		5005	92378
10	11	66	286		8008	184756

Dimenzionalnost "lineariziranog prostora"!



$k \gg n$

Blok-shema sustava za raspoznavanje

$$\vec{X}^* = \begin{bmatrix} x_1^2 \\ x_1 x_2 \\ x_2^2 \\ x_1 \\ x_2 \\ 1 \end{bmatrix}$$

$$d(\vec{x}^*) = \vec{W}^T \vec{X}^*$$

$$\vec{W} = ?$$

M razreda $\omega_1, \omega_2, \ldots, \omega_M$

Pretpostavimo da u k-tom koraku tijekom učenja uzorak $\vec{x}(k)$ pripada razredu ω_i.

Računamo M decizijskih funkcija.

Ako je

$$d_i[\vec{x}(k)] > d_j[\vec{x}(k)] \quad j = 1, 2, \ldots, M; \ j \neq i$$

težinski vektor se ne ugađa:

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

Pretpostavimo da je za neki l

$$d_i[\vec{x}(k)] \leq d_l[\vec{x}(k)]$$

težinski vektori se sada ugađaju:

$$\vec{w}_i(k+1) = \vec{w}_i(k) + c\,\vec{x}(k)$$

$$\vec{w}_l(k+1) = \vec{w}_l(k) - c\,\vec{x}(k)$$

$$\vec{w}_j(k+1) = \vec{w}_j(k); \quad j = 1, 2, \ldots, M, \ j \neq i, \ j \neq l.$$

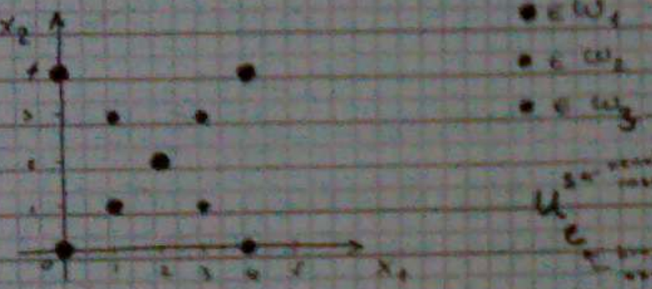
c - pozitivna konstanta

$\vec{w}_i(1)$ su proizvoljni početni vektori $i = 1, 2, \ldots, M$.

Primjer

Zadan je skup uzoraka za učenje:

- $U_4^1 = \{(0,0)^T, (0,4)^T, (4,0)^T, (4,4)^T\}$
- $U_4^2 = \{(1,1)^T, (1,3)^T, (3,1)^T, (3,3)^T\}$
- $U_1^3 = \{(2,2)^T\}$



Zadane uzorke ne možemo odijeliti linearnim dec. funkcijama!

Pokušajmo kvadratnim! $\Rightarrow$ $r=2$, $n=2$

$$K = \frac{(n+1)(n+2)}{2} = 6 \text{ — dimenzionalni prostor}$$

$n=2$ 2-dim $\longrightarrow$ 6-dim.

Preslikavanje 2- u 6-dimenz. prostor funkcijama

$$f_1 = x_1^2 \quad f_2 = x_1 x_2 \quad f_3 = x_2^2 \quad f_4 = x_1 \quad f_5 = x_2 \quad f_6 = 1$$

Preslikan skup za učenje je:

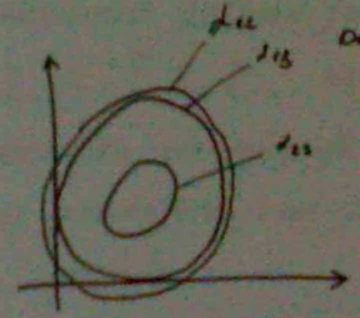
$U_4^1 = \{(0,0,0,0,0,1)^T, (0,0,16,0,4,1)^T, (16,0,0,4,0,1)^T,$
$(16,16,16,4,4,1)^T\}$

$U_4^2 = \{(1,1,1,1,1,1)^T, (1,3,9,1,3,1)^T, (9,3,1,3,1,1)^T,$
$(9,9,9,3,3,1)^T\}$

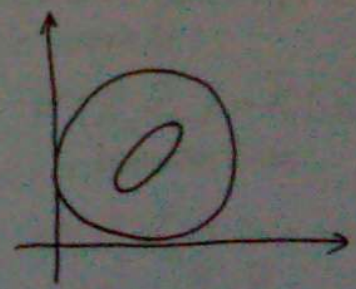
$U^3 \ldots$

Upotrijebimo poopćeni algoritam perceptrona. (M=3).

Dva od mogućih rezultata (u zavisnosti od izbora početnih vrijednosti:



Decizijske se granice dobivaju nakon nešto više od 2000 koraka poopćenog algoritma perceptrona



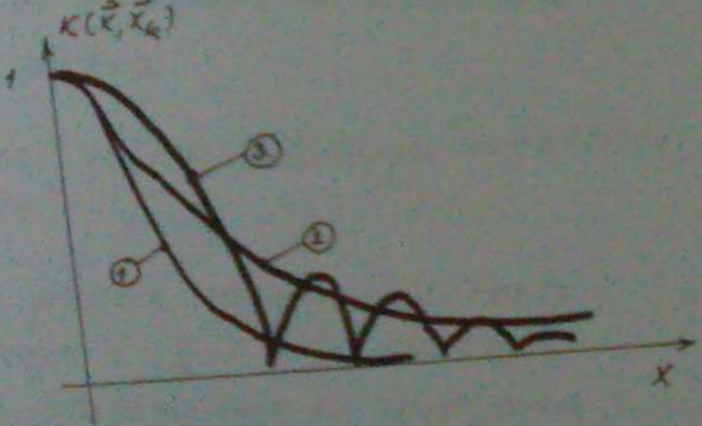
mjeri potencijalnih funkcija:

1. $K(\vec{x}, \vec{x}_k) = e^{-\alpha \|\vec{x} - \vec{x}_k\|^2}$; $\alpha > 0$

2. $K(\vec{x}, \vec{x}_k) = \dfrac{1}{1 + \alpha \|\vec{x} - \vec{x}_k\|^2}$

3. $K(\vec{x}, \vec{x}_k) = \left| \dfrac{\sin \alpha \|\vec{x} - \vec{x}_k\|^2}{\alpha \|\vec{x} - \vec{x}_k\|^2} \right|$

Ilustracija potencijalnih funkcija za $n = 1$



- Potencijalne funkcije opisuju potencijal točke $\vec{x}$ (u prostoru značajki) zbog naboja $\vec{x}_k$ (u točki prostora)

- Ako je potencijal točke $\vec{x}$ zbog jediničnog naboja u točki $\vec{x}_k$ jednak $K(\vec{x}, \vec{x}_k)$ onda je u toj točki prosječni potencijal zbog jediničnih naboja N_i uzoraka iz razreda ω_i jednak

$$\frac{1}{N_i} \sum_{k=1}^{N_i} K(\vec{x}, \vec{x}_k)$$

- Uzorci iz razreda ω_1 i ω_2
- Uzorci su u prostoru predočeni točkama kojima je dodijeljen jedinični naboj

  uzorcima iz $\omega_1 \rightarrow$ pozitivan jedinični naboj

  uzorcima iz $\omega_2 \rightarrow$ negativan jedinični naboj

$+q$ svakom uzorku $\vec{x}_k \in \omega_1$

$-q$ svakom uzorku $\vec{x}_k \in \omega_2$ $k = 1, 2, \ldots, N$

...ci zbog svog naboja prouzrokuju potencijalno polje – potencijal

- Potencijal ima vršnu vrijednost u samoj točki $\vec{X}_k$ i pada s udaljenosti od $\vec{X}_k$
- Uzorci koji pripadaju razredu ω_1 oblikuju "pozitivni plato" s točkama koje su smještene na vrhovima tog platoa
- Uzorci iz ω_2 oblikuju "negativni plato"
- Dva platoa zbog različitosti naboja ($+q$ i $-q$) oblikuju dva platoa koja su odijeljena udolinom u kojoj je potencijal jednak nuli
- Udolina → granica između razreda → decizijska funkcija

Postupak učenja:

N uzoraka u skupu za učenje

$\vec{X}_i \in \omega_1 \cup \omega_2$

Pretpostavimo da se decizijska funkcija oblikuje ovako:

$$d_k(\vec{X}) = d_{k-1}(\vec{X}) + r_k K(\vec{X}, \vec{X}_k) \text{, gdje je}$$

$$d_0(\vec{X}) = 0$$

r_k je korekcijski faktor

$$r_k = \begin{cases} 0 & \text{ako je } \vec{X}_k \in \omega_1 \text{ i } d_{k-1}(\vec{X}_k) > 0 \\ 0 & \text{ako je } \vec{X}_k \in \omega_2 \text{ i } d_{k-1}(\vec{X}_k) < 0 \\ 1 & \text{ako je } \vec{X}_k \in \omega_1 \text{ i } d_{k-1}(\vec{X}_k) \leq 0 \\ -1 & \text{ako je } \vec{X}_k \in \omega_2 \text{ i } d_{k-1}(\vec{X}_k) \geq 0 \end{cases}$$

za $k=1$ (uzimamo prvi uzorak iz skupa za učenje)

$$d_1(\vec{X}_1) = 0 + K(\vec{X}, \vec{X}_1) \quad \text{ako je } \vec{X}_1 \text{ iz } \omega_1$$

$$d_1(\vec{X}_1) = 0 - K(\vec{X}, \vec{X}_1) \quad \text{ako je } \vec{X}_1 \text{ iz } \omega_2$$

k+1 korak:

- Ako je $\vec{X}_{k+1} \in \omega_1$ i

  $d_k(\vec{X}_{k+1}) > 0$ ili

  $\vec{X}_{k+1} \in \omega_2$ i

  $d_k(\vec{X}_{k+1}) < 0$ tada

  $$d_{k+1}(\vec{x}) = d_k(\vec{x})$$

- Ako je $\vec{X}_{k+1} \in \omega_1$ i

  $d_k(\vec{X}_{k+1}) \leq 0$ onda

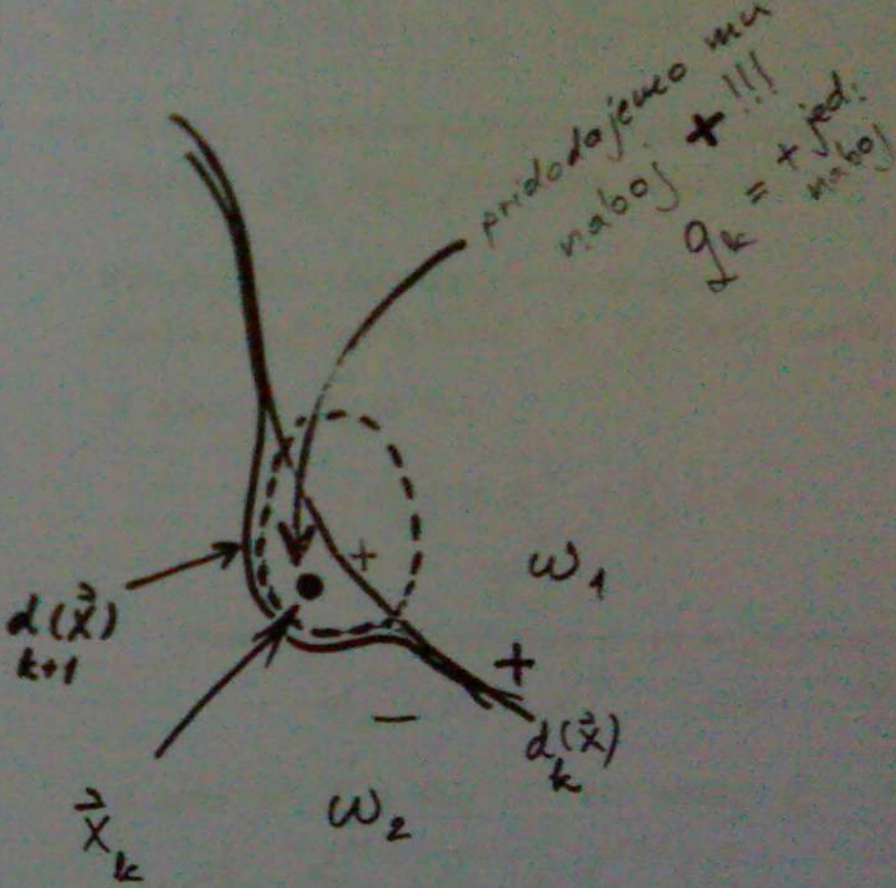
  $$d_{k+1}(\vec{x}) = d_k(\vec{x}) + K(\vec{x}, \vec{X}_{k+1})$$

- Ako je $\vec{X}_{k+1} \in \omega_2$ i $d_k(\vec{X}_{k+1}) \geq 0$

  onda

  $$d_{k+1}(\vec{x}) = d_k(\vec{x}) - K(\vec{x}, \vec{X}_{k+1})$$

Grafička interpretacija.



$\vec{X}_k \in \omega_1$ ima pozitivan naboj ali $d(\vec{x}) = 0$ ga svrstava u ω_2

POVEĆATI NABOJ $\vec{X}_k$ TAKO DA JE $d(\vec{x})$||

...enje postupka za $M>2$ razreda:

razred ω_i:

svaka točka područja P_i u prostoru značajki koji odgovara uzorcima iz ω_i ima prosječni potencijal (zbog naboja uzoraka iz ω_i) veći od potencijala zbog naboja uzoraka iz ostalih razreda

$$d_i(\vec{x}) = \frac{1}{N_i}\sum_{k=1}^{N_i} K(\vec{x},\vec{x}_k)$$

W. S. Meisel (1969. godine) je pokazao da se uvijek može naći takva potencijalna funkcija $K(\vec{x},\vec{x}_k)$ da granicom

$$d(\vec{x}) = d_i(\vec{x}) - d_j(\vec{x}) = 0$$

pravilno razgraničavamo uzorke iz skupa za učenje iz razreda ω_i i ω_j.

Postupak učenja za $M>2$

$$d_0^1(\vec{x}),\ d_0^2(\vec{x}) \ldots\ d_0^M(\vec{x}) = 0$$

- u k-tom koraku učenja obrađujemo uzorak $\vec{x}_k$ koji pripada razredu ω_i
- decizijsku funkciju u k-tom koraku određujemo na sljedeći način:
  - ako je $\vec{x}_k \in \omega_i$ i $d_{k-1}^{(i)}(\vec{x}_k) > d_{k-1}^{(j)}(\vec{x}_k)$ za $\forall j \neq i$

    funkciju odlučivanja ne mijenjamo

    $$d_k^{(i)}(\vec{x}) = d_{k-1}^{(i)} \quad i=1,2,\ldots,M$$

  - ako je $\vec{x}_k \in \omega_i$ i za neki l je $d_{k-1}^{(i)}(\vec{x}_k) \leq d_{k-1}^{(l)}(\vec{x}_k)$

tada MORAMO POPRAVITI DEC. FUNKCIJE

$$d_k^{(i)}(\vec{x}) = d_{k-1}^{(i)} + K(\vec{x},\vec{x}_k)$$

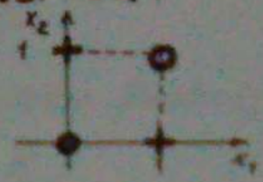
$$d_k^{(l)}(\vec{x}) = d_{k-1}^{(l)} - K(\vec{x},\vec{x}_k)$$

$$d_k^{(j)}(\vec{x}) = d_{k-1}^{(j)}(\vec{x}) \quad j=1,2,\ldots,M,\ j\neq i;\ j\neq l$$

Zadana su dva razreda uzoraka:

$\omega_1 = \{(0,0)^T, (1,1)^T\}$

$\omega_2 = \{(0,1)^T, (1,0)^T\}$



$+ \in \omega_2$

$o \in \omega_1$

Izaberimo potencijalnu funkciju $K(\vec{x}, \vec{x}_k) = e^{-\alpha \|\vec{x} - \vec{x}_k\|^2}$

$\alpha = 1$

Za $n=2$ (dimenzionalnost uzorka) potencijalna funkcija ima oblik:

$$K(\vec{x}, \vec{x}_k) = e^{-[(x_1 - x_{k1})^2 + (x_2 - x_{k2})^2]}$$

Započnimo postupkom učenja:

- $\vec{x}_1 = (0,0)^T \in \omega_1$ prvi uzorak za učenje

$$d_1(\vec{x}) = d_0(\vec{x}) + K(\vec{x}, \vec{x}_1) \; ; \; d_0(\vec{x}) = 0$$

$$d_1(\vec{x}) = e^{-[(x_1-0)^2 + (x_2-0)^2]}$$

$$d_1(\vec{x}) = e^{-(x_1^2 + x_2^2)}$$

- uzorak $\vec{x}_2 = (1,1) \in \omega_1$ vrijedi

$$d_1(\vec{x}_2) = e^{-(1+1)} = e^{-2} > 0$$

zato je: $d_2(\vec{x}) = d_1(\vec{x}) = e^{-(x_1^2 + x_2^2)}$

- za uzorak $x_3 = (0,1)^T \in \omega_2$ je

$$d_2(\vec{x}_3) = e^{-(x_1^2 + x_2^2)} = e^{-1} > 0$$

PAZI $\vec{x}_3$ je iz ω_2!

$$d_3(\vec{x}) = d_2(\vec{x}) - K(\vec{x}, \vec{x}_3)$$

$$d_3(\vec{x}) = e^{-(x_1^2 + x_2^2)} - e^{-[(x_1-0)^2 + (x_2-1)^2]}$$

- za sljedeći uzorak $\vec{x}_4 = (1,0)^T \in \omega_2$ dobivamo:

$$d_3(\vec{x}_4) = e^{-(1^2)} - e^{-[1^2 + (-1)^2]}$$

$$= e^{-1} - e^{-2} > 0$$

Moramo popraviti funkciju odlučivanja

$$d_4(\vec{x}) = d_3(\vec{x}) - K(\vec{x}, \vec{x}_4)$$

$$d_4(\vec{x}) = e^{-(x_1^2 + x_2^2)} - e^{-[x_1^2 + (x_2-1)^2]} - e^{-[(x_1-1)^2 + x_2^2]}$$

Provjeravamo da li se funkcijom $d_4(\vec{x})$ može pravilno razvrstati sve uzorke iz skupa za učenje!

$\vec{x}_5 = \vec{x}_1 = (0,0)^T \in \omega_1$

$$d_4(\vec{x}_5) = e^{-0} - e^{-1} - e^{-1} > 0$$

$$d_5(\vec{x}) = d_4(\vec{x})$$

za $X_6 = X_2 = (1,1)' \in \omega_1$

$$d_5(\vec{X}_6) = e^{-2} - e^{-1} - e^{-1} < 0$$

Treba popraviti decizijsku funkciju!

$$d_6(\vec{x}) = d_4(\vec{x}) + K(\vec{x}, \vec{X}_6)$$

$$d_6(\vec{x}) = e^{-(x_1^2+x_2^2)} - e^{-[x_1^2+(x_2-1)^2]} - e^{-[(x_1-1)^2+x_2^2]} + e^{-[(x_1-1)^2+(x_2-1)^2]}$$

- za $\vec{X}_7 = \vec{X}_3 = (0,1)^T \in \omega_2$ je

$$d_6(\vec{X}_7) = e^{-1} - e^{0} - e^{-2} + e^{-1} < 0 \quad \text{zato je}$$

$$d_7(\vec{x}) = d_6(\vec{x})$$

- za $\vec{X}_8 = \vec{X}_4 = (1,0)^T \in \omega_2$ je

$$d_7(\vec{X}_8) = e^{-1} - e^{-2} - e^{0} + e^{-1} < 0 \quad \text{zato je}$$

$$d_8(\vec{x}) = d_7(\vec{x})$$

- za $\vec{X}_9 = \vec{X}_1 = (0,0)^T \in \omega_1$ je

$$d_8(\vec{X}_9) = e^{0} - e^{-1} - e^{-1} + e^{-2} > 0 \quad \text{zato je}$$

$$d_9(\vec{x}) = d_8(\vec{x})$$

- za $X_{10} = \vec{X}_2 = (1,1)' \in \omega_1$

$$d_9(\vec{X}_{10}) = e^{-2} - e^{-1} - e^{-1} + e^{0} > 0$$

$$d_{10}(\vec{x}) = d_9(\vec{x})$$

Sve smo uzorke iz skupa za učenje pravilno razvrstali funkcijom odlučivanja:

$$d(\vec{x}) = e^{-(x_1^2+x_2^2)} - e^{-[x_1^2+(x_2-1)^2]} - e^{-[(x_1-1)^2+x_2^2]} + e^{-[(x_1-1)^2+(x_2-1)^2]}$$

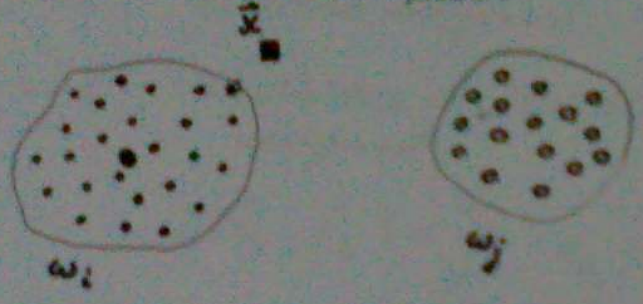
Funkcija odlučivanja ima onoliko eksponencijalnih članova koliko je bilo KOREKCIJA u postupku učenja.

KLASIFIKACIJA UZORAKA POMOĆU FUNKCIJA UDALJENOSTI

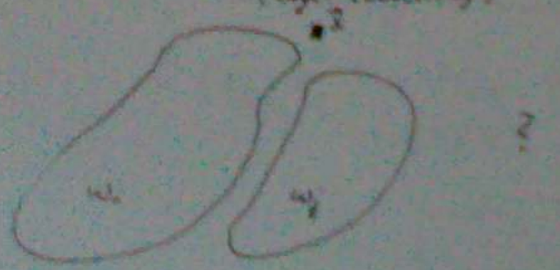
- jedna od najjednostavnijih metoda
- osnovana na intuiciji

- mjera sličnosti između vektora uzoraka

vektor — točka u Euklidovom prostoru



- djelotvorna metoda kada razredi uzoraka pokazuju svojstva grupiranja (engl. clustering)



- KLASIFIKACIJA POMOĆU NAJMANJE UDALJENOSTI

(MINIMUM-DISTANCE PATTERN CLASSIFICATION)

MJERA UDALJENOSTI BILO KOJA FUNKCIJA KOJA ZADOVOLJAVA UVJETE:

(a) $D(\vec{x}_k, \vec{x}_l) = 0$ za $\vec{x}_k = \vec{x}_l$

$D(\vec{x}_k, \vec{x}_l) \neq 0$ za $\vec{y}_k \neq \vec{x}_l$

(b) $D(\vec{x}_k, \vec{x}_l) = D(\vec{x}_l, \vec{x}_k)$

(c) $D(\vec{x}_k, \vec{x}_l) \leq D(\vec{x}_k, \vec{x}_j) + D(\vec{x}_j, \vec{x}_l)$

Euklidova udaljenost

$$D = \|\vec{x}_k - \vec{x}_l\| = \sqrt{\sum_{i=1}^{n} (x_{k_i} - x_{l_i})^2}$$

- invarijantnost grupa u odnosu na translaciju i rotaciju