a) Postupak učenja za kriterijsku funkciju

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

- parcijalna derivacija funkcije

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

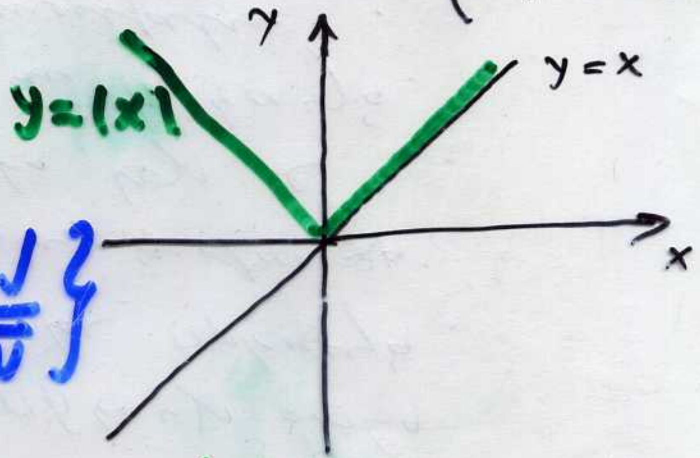
$$\mathrm{sgn}(\vec{w}^T\vec{x})=\begin{cases}1 & \text{ako je } \vec{w}^T\vec{x}>0\\ -1 & \text{ako je } \vec{w}^T\vec{x}\leq 0\end{cases}\qquad \frac{d}{d\vec{x}}(\vec{x}^T\cdot\vec{a})=\vec{a}$$

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

UVRSTIMO $\frac{\partial J}{\partial \vec{w}}$

u:

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



$$\frac{dy}{dx}=\mathrm{sgn}\,x$$

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

$\vec{x}(k)$ – uzorak iz skupa za učenje koji se primjenjuje u k-tom koraku učenja

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

$\uparrow$

POSTUPAK PERCEPTRONA

SA STALNIM PRIRASTOM

(Rosenblatt, 1962.)

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 :

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

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

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

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

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

Postupak učenja:

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

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

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 $0 < \lambda < 2$;

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

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

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

3. slučaj: postoji M decizijskih funkcija: $d_1, d_2, \ldots, d_M$;
$\vec{x} \in \omega_i$ ako
$d_i(\vec{x}) > d_j(\vec{x})$; $i \neq j$
$i = 1, 2, \ldots, M$

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:

$$d_j(\vec{x}(k)) = \vec{w}_j^T(k)\vec{x}(k);$$

$$j = 1, 2, \ldots, M$$

Ako je

$$d_i(\vec{x}(k)) > d_j(\vec{x}(k));$$

$$j = 1, 2, \ldots, M \quad j \neq i$$

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 :

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

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

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

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

c - pozitivna konstanta

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

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

ako su razredi linearno separatibilni.

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 $\vec{x}^*$

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

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

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 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

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

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

$k = n$

- Polinom drugog stupnja:

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

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

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

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

Opći slučaj kvadratnog oblika:

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

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

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

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

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

$$\{f_{2n+1}(\vec{x}), \dots, f_k(\vec{x})\} = \{x_1 x_2, x_1 x_3, \dots, x_{n-1} x_n\}$$

- 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

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

s :

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

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

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

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

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

Primjer:

za $r=2$ i $n=2$:

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

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

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

$$d^2(\vec{x}) = w_{11}x_1^2 + w_{12}x_1x_2 + w_{22}x_2^2 + d^1(\vec{x})$$

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

$$\vec{W}^T = \begin{bmatrix} w_{11} \\ w_{12} \\ w_{22} \\ w_1 \\ w_2 \\ w_3 \end{bmatrix} \qquad \vec{x}^* = [x_1^2, x_1x_2, x_2^2, x_1, x_2, 1]^T$$

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

$$d^3(\vec{x}) = \left(\sum_{p_1=1}^{2} \sum_{p_2=p_1}^{2} \sum_{p_3=p_2}^{2} w_{p_1p_2p_3} x_{p_1} x_{p_2} x_{p_3} \right)$$
$$+ d^2(\vec{x}) ;$$

$$d(\vec{x}) = w_{111}x_1^3 + w_{112}x_1^2x_2 + w_{122}x_1x_2^2 +$$
$$w_{222}x_2^3 + d^2(\vec{x})$$

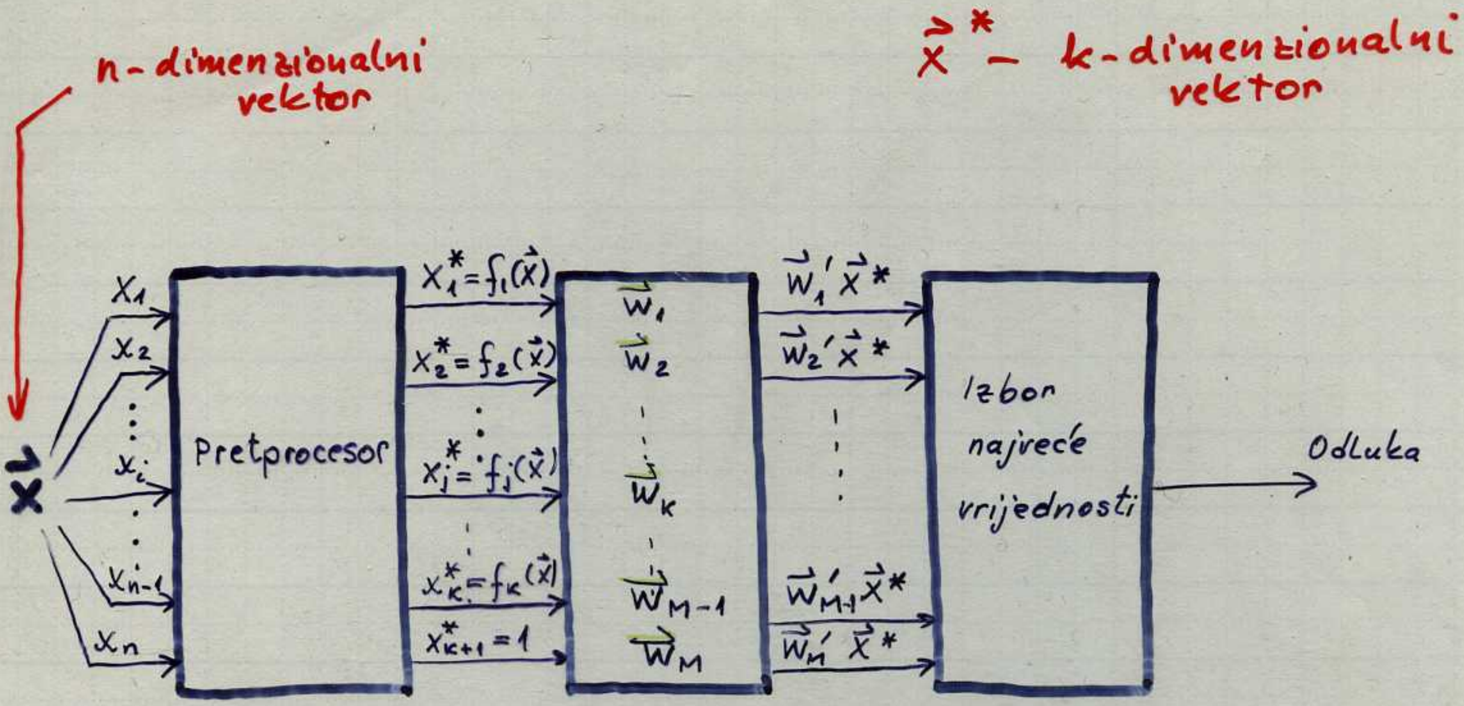
$$d(\vec{x}) = w_{111}x_1^3 + w_{112}x_1^2x_2 + w_{122}x_1x_2^2 +$$
$$w_{222}x_2^3 + w_{11}x_1^2 + w_{12}x_1x_2 + w_{22}x_2^2 +$$
$$w_1x_1 + w_2x_2 + w_3$$

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

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

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

$$\vec{W} = ?$$

Poopćeni algoritam perceptrona

M razreda : $\omega_1, \omega_2, \dots, \omega_M$

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

Računamo M decizijskih funkcija.

Ako je

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

težinski vektor se ne ugađa:

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

Pretpostavimo da je za neki ℓ

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

težinski vektori se sada ugađaju:

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

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

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

c - pozitivna konstanta

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

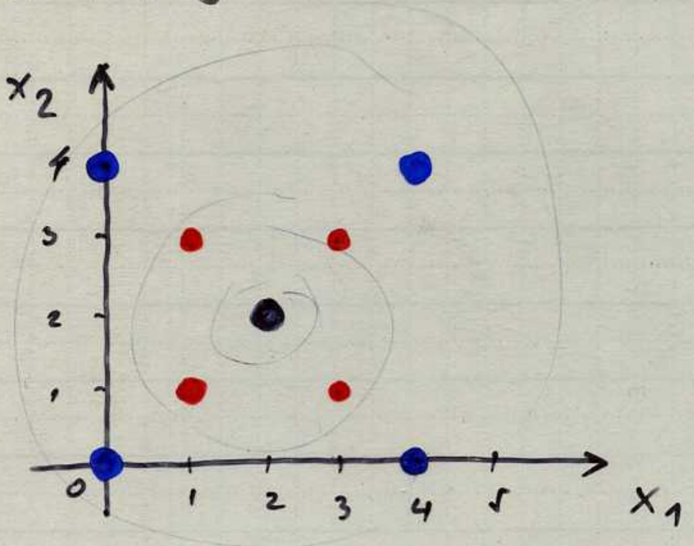
PRIMJER

Zadan je skup uzoraka za učenje:

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

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

• $U_1^3 = \{(2,2)^T\}$



● $\in \omega_1$

● $\in \omega_2$

● $\in \omega_3$

U_e^s ← oznaka razreda; e ← broj uzoraka u raz.

Zadane uzorke ne možemo odijeliti linearnim dec. funkcijama!

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

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

$\underline{n=2}$ 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:

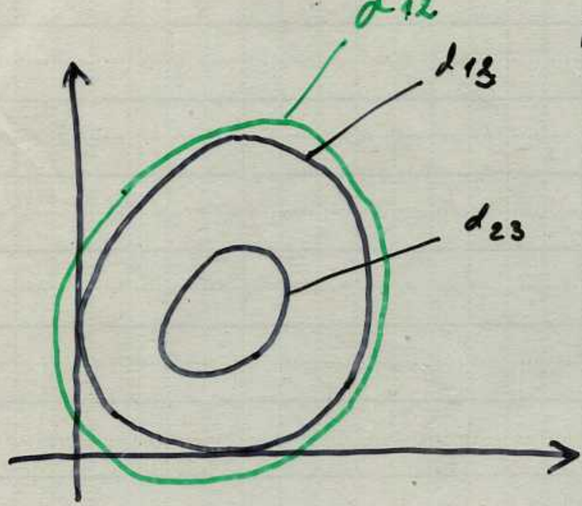
$U_4^1 = \{(0,0,0,0,0,1)^T, (0,0,16,0,4,1)^T, (16,0,0,4,0,1)^T,$
$\underline{(16,16,16,4,4,1)^T\}}$

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

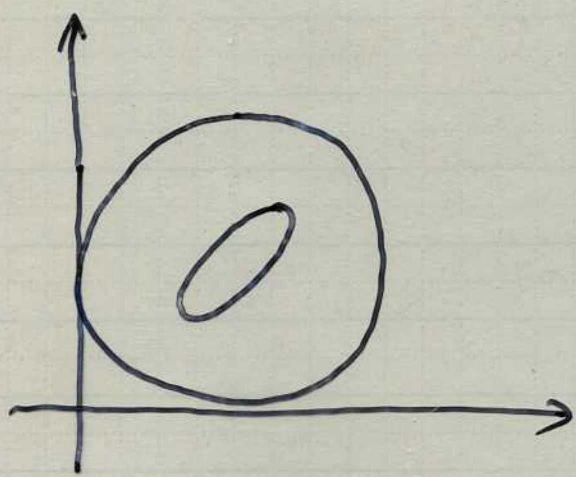
$U_1^3 = \{(4,4,4,2,2,1)^T\}.$

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

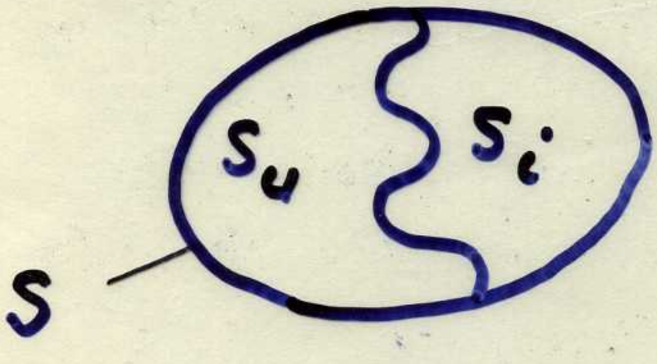


SKUP UZORAKA ZA UČENJE I SKUP UZORAKA ZA ISPITIVANJE - METODE ISPITIVANJA

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

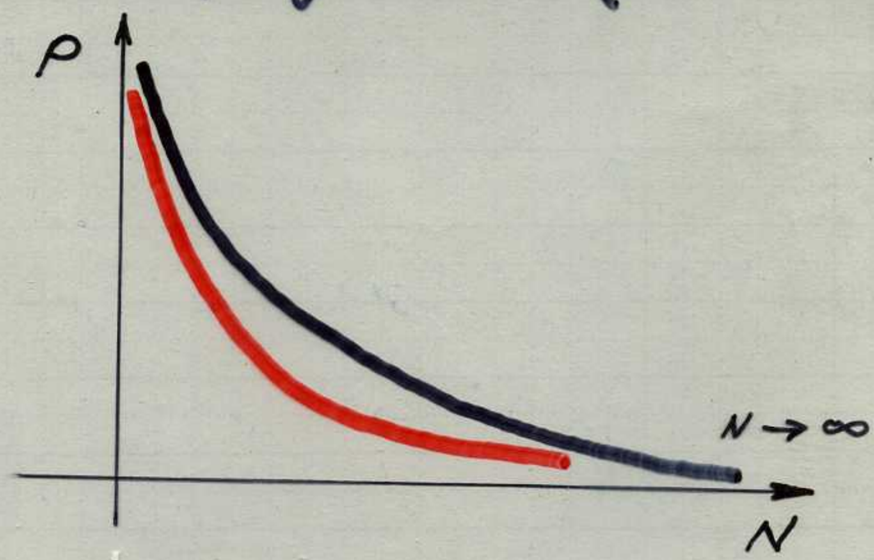
S_i - skup uzoraka za ispitivanje

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

S - skup uzoraka s poznatom klasifikacijom

Nedostaci Holdout metode:

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



2. Leave-One-Out metoda

- metoda pokušava "zaobići" problem podjele skupa označenih uzoraka

Učenje se obavlja uporabom $N-1$ uzoraka a ispitivanje se izvodi uporabom onog jednog preostalog uzorka.

Ako je taj uzorak pogrešno razvrstan → inkrementira se brojilo pogreške;

Postupak se ponavlja N puta ali tako da je svaki put isključen drugi uzorak.

Ukupan broj pogrešaka nas upućuje na procjenjenu vjerojatnost pogreške klasifikatora

Nedostatak metode: velika računska složenost

3. Resubstitution metoda
(Metoda ponovne zamjene)

Isti se skup podataka koristi, prvo za učenje a zatim za ispitivanje.

- optimistička procjena vjerojatnosti pogreške klasifikatora

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

- dovoljnost informacije
- postojanost značajki
- geometrijska postojanost

  (mala udaljenost među uzorcima u prostoru značajki znači i malu razliku u svojstvima objekta)

N?

Idealno $N \to \infty$

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

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

slika lica

Primjer: Klasifikacija brojčano-slovčanih znakova

$M = 30 + 10 = 40$

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

Dimenzionalnost vektora značajki

$n = 18.$

$N = 5 * 18 * 40 = 3600$

slika brojčano-slovčanih znakova