$\vec{W}^T \vec{X}_i > 0$ za $i = 1, 2, \ldots, N$

$\vec{W}_{[1\times3]}^T \vec{X}_{i[3\times1]} > 0_{[1\times1]}$ za $i = 1, 2, \ldots, N$

gdje je $\vec{X}_i$ i-ti (transponirani) redak matrice [X] dimenzija $N \times (n+1)$

OSNOVNA ZAMISAO:

IZABRATI NEKU FUNKCIJU KOJA ĆE DOSTIĆI MINIMUM KAD JE ISPUNJEN UVJET: $\vec{W}^T \vec{X}_i > 0$ za $i = 1, 2, \ldots, N$

-Zahtijevamo da izabrana funkcija ima samo jedan minimum (i naravno funkcija je funkcija vektorskog argumenta ($\vec{W}$))

UZMIMO FUNKCIJU : $J(\vec{W}, \vec{X}) = (\overbrace{\left|\vec{W}^T \vec{X}\right|}^{\text{apsolutna vrijednost!}} - \vec{W}^T \vec{X})$

-minimum funkcije $J(\vec{W}, \vec{X}) = 0$

-trivijalan slučaj $\vec{W} = 0$

-netrivijalan slučaj:

MINIMUM (KRITERIJSKE) FUNKCIJE $J(\vec{W}, \vec{X})$ POSTIŽE SE PRI ISPUNJENJU UVJETA $\vec{W}^T \cdot \vec{X} > 0$

GRADIJENTNI POSTUPAK: KORAK PO KORAK POVEĆAVAMO ARGUMENT $\vec{W}$ U SMJERU NEGATIVNOG GRADIJENTA FUNKCIJE $J(\vec{W}, \vec{X})$ SVE DOK NE POSTIGNEMO MINIMUM (KRITERIJSKE) FUNKCIJE $J(\vec{W}, \vec{X})$.

$\vec{W}(k)$ - vrijednost $\vec{W}$ u k-tom koraku

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

$\vec{W}(k+1)$ - vrijednost "novog" vektora $\vec{W}$ (u k+1 koraku)

c - pozitivna konstanta različita od 0, koja određuje količinu korekcije

/korekcija se više ne izvodi kada je $\frac{\partial J(\vec{W}, \vec{X})}{\partial \vec{W}} = \vec{0}$; što je uvjet za minimum!/

Funkcija	Derivacija
C (konst)	0
x^n	nx^{n-1}
$\frac{1}{x}$	$-\frac{1}{x^2}$
$\frac{1}{x^n}$	$-\frac{n}{x^{n+1}}$
$\sqrt{x}$	$\frac{1}{2\sqrt{x}}$
$\sqrt[n]{x}$	$\frac{1}{n\sqrt[n]{x^{n-1}}}$
e^x	e^x
a^x	$a^x \ln a$
$\ln x$	$\frac{1}{x}$

Derivacija funkcije s konst. faktorom

$(cu)' = c \cdot u'$

Derivacija produkta dvije ili nekoliko funkcija

$(uv)' = uv' + u'v$

$(uvw)' = uvw' + uv'w + u'vw$

Derivacija razlomaka

$(\frac{u}{v})' = \frac{u \cdot v' - u'v}{v^2}$

Derivacija funkcije od funkcije (složene funkcije)

$y = f(u); u = \varphi(x)$ tada $\frac{dy}{dx} = f'(u) \cdot \varphi'(x)$

Derivacija složene funkcije od nekoliko varijbali: $u = f(x, y, \ldots, t)$ gdje je $x = \varphi(\xi) y = \psi(\xi) \ldots t = \chi(\xi)$

$\frac{du}{d\xi} = \frac{\partial u}{\partial x}\frac{dx}{d\xi} + \frac{\partial u}{\partial y}\frac{dy}{d\xi} + \ldots + \frac{\partial u}{\partial t}\frac{dt}{d\xi}$

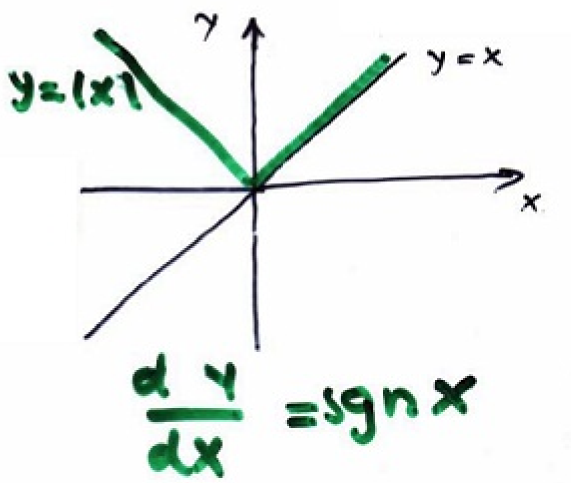
Deriviranje lineranih funkcija:
$\frac{d}{d\vec{X}}(\vec{X}^T A) = A$
$\frac{d}{d\vec{X}}(\vec{X}^T) = I$
$\frac{d}{d\vec{X}}(\vec{X}^T \vec{a}) = \frac{d}{d\vec{X}}(\vec{a}^T \vec{X}) = \vec{a}$
$\frac{d}{dX}(\vec{a}^T X \vec{b}) = \vec{a}\vec{b}^T$
$\frac{d}{dX}(\vec{a}^T X \vec{a}) = \frac{d}{dX}(\vec{a}^T X^T \vec{a}) = \vec{a}\vec{a}^T$
$\frac{d}{d\vec{X}}(\vec{X}^T C \vec{X}) = (C + C^T)\vec{X}$, ako je $C = C^T$ onda $\frac{d}{d\vec{X}}(\vec{X}^T C \vec{X}) = 2C\vec{X}$
$\frac{d}{d\vec{X}}(\vec{X}^T \vec{X}) = 2\vec{X}$
$\frac{d}{d\vec{X}}(A\vec{X} + b)^T (D\vec{X} + e) = A^T(D\vec{X} + e) + D^T(A\vec{X} + b)$

a) Postupak učenja za kriterijsku funkciju $J(\vec{W}, \vec{X}) = \frac{1}{2}(\left|\vec{W}^T \vec{X}\right| - \vec{W}^T \vec{X})$

-parcijalna derivacija funkcije:
$\frac{\partial J}{\partial \vec{W}} = \frac{1}{2}\left[\vec{X} sgn(\vec{W}^T \vec{X}) - \vec{X}\right]$

$sgn(\vec{W}^T \vec{X}) = \begin{cases} 1 \text{ ako je } \vec{W}^T \vec{X} > 0 \\ -1 \text{ ako je } \vec{W}^T \vec{X} \leq 0 \end{cases}$

$\frac{d}{d\vec{X}}(\vec{X}^T \vec{a}) = \vec{a}$, $sgn(x) = \begin{cases} 1 \text{ ako je } x > 0 \\ -1 \text{ ako je } x \leq 0 \end{cases}$



Slika 1: Funkcija sgn

UVRSTIMO $\frac{\partial J}{\partial \vec{W}}$ u $\vec{W}(k+1) = \vec{W}(k) - c\left\{\frac{\partial J(\vec{W},\vec{X})}{\partial \vec{W}}\right\}$
$\vec{W}(k+1) = \vec{W}(k) + \frac{c}{2}\left\{\vec{X}(k) - \vec{X}(k) sgn\left[\vec{W}^T(k)\vec{X}(k)\right]\right\}$
$\vec{X}(k)$ - uzorak iz skupa za učenje koji se primjenjuje u k-tom koraku učenja
<u>POSTUPAK PERCEPTRONA SA STALNIM PRIRASTOM</u> (Rosenblatt,1962.):

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

PRIMJER:
$\omega_1 = \left\{(0,0)^T, (0,1)^T\right\}$ i
$\omega_2 = \left\{(1,0)^T, (1,1)^T\right\}$
-povećajmo dimenzionalnost prostora značajki na dimenziju n+1
$\left\{(0,0,1)^T, (0,1,1)^T\right\}$
$\left\{(1,0,1)^T, (1,1,1)^T\right\}$

-pomnožimo sve uzorke iz razreda ω_2 sa -1
$\left\{(0,0,1)^T, (0,1,1)^T\right\}$
$\left\{(-1,0,-1)^T, (-1,-1,-1)^T\right\}$

-pretpostavimo da su skupovi uzoraka za učenje razdvojivi linearnom funkcijom odlučivanja
$d(\vec{X}) = w_1x_1 + w_2x_2 + w_3 = \vec{W}^T\vec{X}$
-postupak perceptrona sa stalnim prirastom
Neka je c=1: neka je početni vektor koeficijenata $\vec{W}(1)$ proizvoljan, npr. $\vec{W}(1) = (-1,0,0)^T$
-u prvom koraku učenja uzimamo uzorak $\vec{X}(1) = (0,0,1)^T$

Računamo:
$\vec{W}^T(1)\vec{X}(1) = (-1,0,0)\begin{pmatrix}0\\0\\1\end{pmatrix} = 0,$
zbog toga $\vec{W}(2) = \vec{W}(1) + \vec{X}(1) = \begin{pmatrix}-1\\0\\1\end{pmatrix}$
U drugom koraku uzimamo uzorak $\vec{X}(2) = (0,1,1)^T$
Računamo:
$\vec{W}^T(2)\vec{X}(2) = (-1,0,1)\begin{pmatrix}0\\1\\1\end{pmatrix} = 1,$
Zato što je $\vec{W}^T(2)\vec{X}(2) > 0$, u skladu s postupkom učenja $\vec{W}(3) = \vec{W}(2)$
U trećem koraku upotrebljavamo uzorak $\vec{X}(3) = (-1,0,-1)^T$
Računamo:
$\vec{W}^T(3)\vec{X}(3) = (-1,0,1)\begin{pmatrix}-1\\0\\-1\end{pmatrix} = 0,$
$\vec{W}(4) = \vec{W}(3) + \vec{X}(3) = \begin{pmatrix}-2\\0\\0\end{pmatrix}$
U četvrtom koraku upotrebljavamo uzorak $\vec{X}(4) = (-1,-1,-1)^T$
Računamo: $\vec{W}^T(4)\vec{X}(4) = (-2,0,0)\begin{pmatrix}-1\\-1\\-1\end{pmatrix} = 2,$

Budući da je $\vec{W}^T(4)\vec{X}(4) > 0$ vrijedi $\vec{W}(5) = \vec{W}(4)$
RJEŠENJE ĆEMO DOBITI KADA NA TEMELJU IZRAČUNATOG $\vec{W}$ PRAVILNO RAZVRSTAMO SVE UZORKE
-u ovom primjeru došlo je do pogrešnog razvrstavanja dva uzorka - POSLJEDICA $\rightarrow$ popravljanje vektora koef. funkcije odlučivanja $\vec{W}$. Postupak učenja nastavljamo tako da pretpostavimo $\vec{X}(5) = \vec{X}(1)$, $\vec{X}(6) = \vec{X}(2)$, $\vec{X}(7) = \vec{X}(3)$ i $\vec{X}(8) = \vec{X}(4)$
Dobivamo:

$\vec{W}^T(5)\vec{X}(5) = 0$ slijedi $\vec{W}(6) = \vec{W}(5) + \vec{X}(5) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$

$\vec{W}^T(6)\vec{X}(6) = 1$ slijedi $\vec{W}(7) = \vec{W}(6)$

$\vec{W}^T(7)\vec{X}(7) = 1$ slijedi $\vec{W}(8) = \vec{W}(7) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$

$\vec{W}^T(8)\vec{X}(8) = 1$ slijedi $\vec{W}(9) = \vec{W}(8) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$

I u drugom ponavljanju nisu bili razvrstani svi uzorci pravilno. Treća iteracija :
$\vec{X}(9) = \vec{X}(1)$, $\vec{X}(10) = \vec{X}(2)$, $\vec{X}(11) = \vec{X}(3)$, $\vec{X}(12) = \vec{X}(4)$

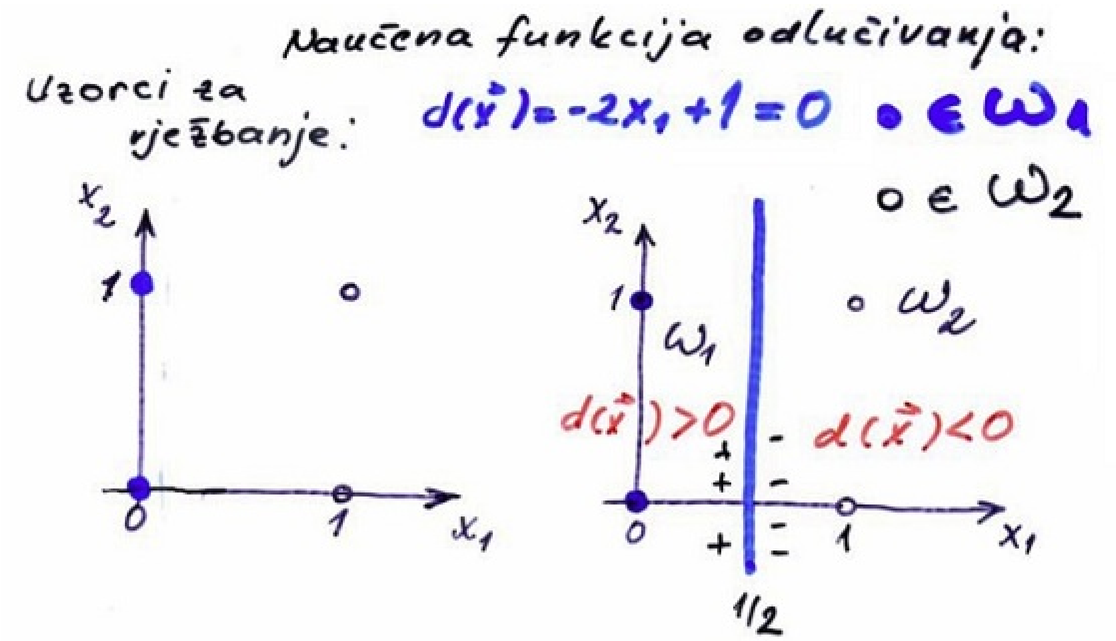
$\vec{W}^T(9)\vec{X}(9) = (-2, 0, 1)\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 1 \;;\rightarrow \vec{W}(10) = \vec{W}(9)$

$\vec{W}^T(10)\vec{X}(10) = (-2, 0, 1)\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = 1 \;;\rightarrow \vec{W}(11) = \vec{W}(10)$

$\vec{W}^T(11)\vec{X}(11) = (-2, 0, 1)\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = 3 \;;\rightarrow \vec{W}(12) = \vec{W}(11)$

$\vec{W}^T(12)\vec{X}(12) = (-2, 0, 1)\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} = 1 \;;\rightarrow \vec{W}(13) = \vec{W}(12)$

$\vec{W}^T = (-1, 0, 1)^T$
Decizijska funkcija $d(\vec{X}) = \underline{\vec{W}^T\vec{X} = -2x_1 + 1}$



Slika 2: Primjer perceptrona

b) Postupak učenja za kriterijsku funkciju:

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

Parcijalna derivacija kriterijske funkcije:

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

$\vec{W}^T\vec{X} \cdot \vec{X}sgn(\vec{W}^T\vec{X}) = \left|\vec{W}^T\vec{X}\right| \cdot \vec{X}$

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

$$\frac{\partial J}{\partial \vec{W}} = \frac{1}{2\vec{X}^T\vec{X}}\left[\left|\vec{W}^T\vec{X}\right| \cdot \vec{X}sgn(\vec{W}^T\vec{X}) - \left|\vec{W}^T\vec{X}\right| \cdot \vec{X}\right]$$

Postupak učenja :

$$\vec{W}(k+1) = \vec{W}(k) + \lambda\frac{\left|\vec{W}^T(k)\vec{X}(k)\right|}{2\vec{X}^T(k)\vec{X}(k)}\left[\vec{X}(k) - \vec{X}(k)sgn(\vec{W}^T(k)\vec{X}(k))\right]$$

$$\vec{W}(k+1) = \vec{W}(k) + \lambda\frac{\left|\vec{W}^T(k)\vec{X}(k)\right|}{2\vec{X}^T(k)\vec{X}(k)}\begin{cases} \vec{0} \text{ ako je } vecX^T(k)\vec{X}(k) > 0 \\ \vec{X}(k) \text{ ako je } vecX^T(k)\vec{X}(k) \leq 0 \end{cases}$$

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 linerano separatibilne razrede) za $0 < \lambda < 2$;

Varijacije algoritma perceptrona

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

varijacija u zavisnosti od izbora korekcijskog inkrementa c:

1) algoritam sa stalnim (čvrstim) prirastom (fixed-increment)
c=konst.>0

2) algoritam s apsolutnom korekcijom (absolute-correction)
c izabran tako da je dovoljno velik da jamči da će uzorak biti ispravno klasificiran nakon ugađanja težinskog faktora $\vec{W}$
Ako je $\vec{X}^T(k)\vec{X}(k) \leq 0$ izabire se c tako da vrijedi:
$\vec{X}^T(k+1)\vec{X}(k) = \left[\vec{X}(k) + c\vec{X}(k)\right]^T \vec{X}(k) > 0$, c je najmanji cijeli broj veći od $\frac{\left|\vec{W}^T(k)\vec{X}(k)\right|}{\vec{X}^T(k)\vec{X}(k)}$

3) algoritam s djelomičnom korekcijom (fractional-correction)
c je izabran tako da je $\left|\vec{W}^T(k)\vec{X}(k) - \vec{W}^T(k+1)\vec{X}(k)\right|$ dio λ od $\left|\vec{W}^T(k)\vec{X}(k)\right|$:

$$\left|\vec{W}^T(k)\vec{X}(k) - \vec{W}^T(k+1)\vec{X}(k)\right| = \lambda\left|\vec{W}^T(k)\vec{X}(k)\right|$$

Uvrstimo $\vec{W}(k+1) = \vec{W}(k) + c\vec{X}(k)$

$$\left|\vec{W}^T(k)\vec{X}(k) - (\vec{W}^T(k))\vec{X}(k) + c\vec{X}^T(k))\vec{X}(k)\right| = \lambda\left|\vec{W}^T(k)\vec{X}(k)\right|$$

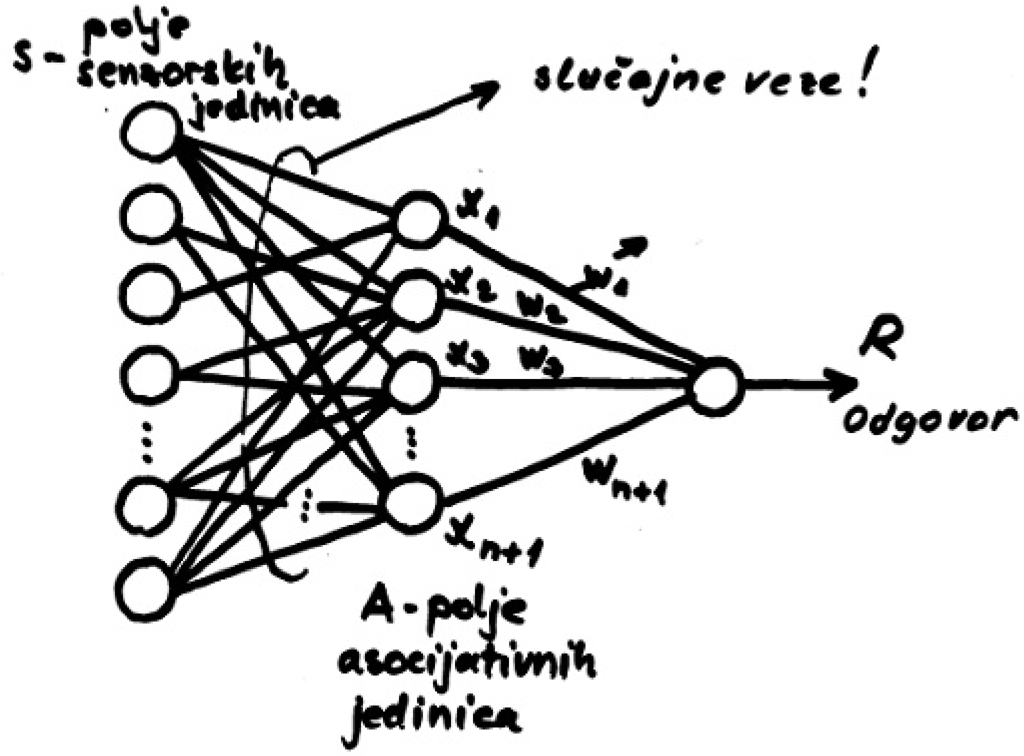
Dobivamo: $c = \lambda\frac{\left|\vec{W}^T(k)\vec{X}(k)\right|}{\vec{X}^T(k)\vec{X}(k)}$

-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



Slika 3: Shema perceptrona

Jedinica 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:
$R = \sum_{i=1}^{n+1} w_i x_i = \vec{W}^T \vec{X}$
Pravilo klasifikacije:

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

- ako je $R \leq 0$ nepoznati se uzorak razvrstava u ω_2

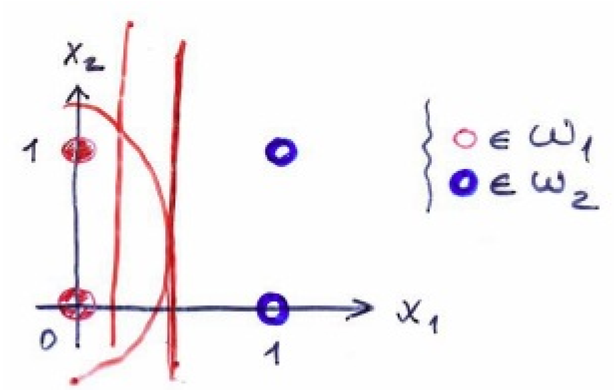
Perceptron - klasifikator uzoraka u dva razreda - ako su razredi linearno separatibilni
- modifikacija perceptrona za klasifikaciju uzoraka u M>2 razreda:
M - jedinica R
$R_1, R_2, \ldots, R_M$
Pravilo: uzorak se razvrstava u razred ω_i akko $R_i > R_j$ za sve $i \neq j$.
Osnovni model perceptrona može se raširiti na nelinearne decizijkse 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-proizvoljnoo izabran
- k-ti korak učenja:
  - Ako $\vec{X}(k) \in \omega_1$ i $\vec{W}^T(k)\vec{X}(k) \leq 0$ zamijeni $\vec{W}(k)$ sa $\vec{W}(k+1) = \vec{W}(k) + c\vec{X}(k)$, gdje je c korekcijski faktor
  - Ako $\vec{X}(k) \in \omega_2$ i $\vec{W}^T(k)\vec{X}(k) \geq 0$ zamijeni $\vec{W}(k)$ sa $\vec{W}(k+1) = \vec{W}(k) - c\vec{X}(k)$, gdje je c korekcijski faktor; c mora biti pozitivan (i konstanta)
  - U drugim slučajevima ostavi $\vec{W}(k) \rightarrow \vec{W}(k+1) = \vec{W}(k)$

PRIMJER:



Slika 4: Prikaz primjera

$\omega_1 : (0,0,1)^T, (0,1,1)^T$
$\omega_2 : (1,0,1)^T, (1,1,1)^T$

c=1 i $\vec{W}(1) = \vec{0}$

1. korak:

$$\vec{W}^T(1) \cdot \vec{X}(1) = (0,0,0) \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 0$$

$$\vec{W}(2) = \vec{W}(1) + \vec{X}(1) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

2. korak:

$$\vec{W}^T(2) \cdot \vec{X}(2) = (0,0,1) \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = 1$$

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

3. korak:

$$\vec{W}^T(3) \cdot \vec{X}(3) = (0,0,1) \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = 1$$

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

4. korak:

$$\vec{W}^T(4) \cdot \vec{X}(4) = (-1,0,0) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = -1$$

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

Zbog korekcija u 1. i 3. koraku postupak se ponavlja!

$\big[RJEŠENJESEDOBIVASAMOKADALGORITAMDAJEPOTPUNOERROR - FREEITERACIJ$

(5)=(1),(6)=(2),(7)=(3)i(8)=(4)

$DRUGAITERACIJADAJE:$

5.$korak:$

$${}^T(5) \cdot \vec{X}(5) = (-1,0,0) \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 0$$

$$\vec{W}(6) = \vec{W}(5) + \vec{X}(5) = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

6. korak:

$$\vec{W}^T(6) \cdot \vec{X}(6) = (-1,0,1) \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = 1$$

$$\vec{W}(7) = \vec{W}(6) = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

7. korak:

$$\vec{W}^T(7) \cdot \vec{X}(7) = (-1,0,1) \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = 0$$

$\vec{W}(8) = \vec{W}(7) - \vec{X}(7) = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -2 \\ 0 \\ 0 \end{pmatrix}$

8. korak:

$\vec{W}^T(8) \cdot \vec{X}(8) = (-2, 0, 0) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = -2$

$\vec{W}(9) = \vec{W}(8) = \begin{pmatrix} -2 \\ 0 \\ 0 \end{pmatrix}$

Dogodile su se dvije pogreške!

Nova iteracija

$\vec{X}(9) = \vec{X}(1)$, $\vec{X}(10) = \vec{X}(2)$, $\vec{X}(11) = \vec{X}(3)$ i $\vec{X}(12) = \vec{X}(4)$

$\vec{W}^T(9) \cdot \vec{X}(9) = 0 \quad \vec{W}(10) = \vec{W}(9) + \vec{X}(9) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$

$\vec{W}^T(10) \cdot \vec{X}(10) = 1 \quad \vec{W}(11) = \vec{W}(10) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$

$\vec{W}^T(11) \cdot \vec{X}(11) = -1 \quad \vec{W}(12) = \vec{W}(11) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$

$\vec{W}^T(12) \cdot \vec{X}(12) = -1 \quad \vec{W}(13) = \vec{W}(12) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$

Dogodila se pogreška!

Nova iteracija:

$\vec{X}(13) = \vec{X}(1)$, $\vec{X}(14) = \vec{X}(2)$, $\vec{X}(15) = \vec{X}(3)$ i $\vec{X}(16) = \vec{X}(4)$

$\vec{W}^T(13) \cdot \vec{X}(13) = (-2, 0, 1) \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 1 \quad \vec{W}(14) = \vec{W}(13) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$

$\vec{W}^T(14) \cdot \vec{X}(14) = (-2, 0, 1) \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 1 \quad \vec{W}(15) = \vec{W}(14) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$

$\vec{W}^T(15) \cdot \vec{X}(15) = (-2, 0, 1) \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = -1 \quad \vec{W}(16) = \vec{W}(15) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$

$\vec{W}^T(16) \cdot \vec{X}(16) = (-2, 0, 1) \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = -1 \quad \vec{W}(17) = \vec{W}(17) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$

$d(\vec{X}) = -2x_1 + 1 \leftarrow \vec{W} = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$

Treća metoda:

LMSE(Least-Mean-Square-Error) (Ho-Kashyap algorithm,1965)

- 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 separatibilni
- 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:
$J(\vec{W}, \vec{X}, \vec{b}) = \frac{1}{2}\sum_{j=1}^{N}(\vec{W}^T\vec{X}_j - b_j)^2 = \frac{1}{2}\left\|[X]\vec{W} - \vec{b}\right\|^2$, gdje $\left\|[X]\vec{W} - \vec{b}\right\|$ 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 varijbale $\vec{W}$ i $\vec{b}$ upotrijebiti u minimizacijskoj proceduri
- Očekujemo poboljšani postupak (brži) konvergencije
- Gradijenti:

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

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

Vektor $\vec{W}$ nije ograničen i možemo postaviti:
$\frac{\partial J}{\partial \vec{W}} = 0$
$[X]^T([X]\vec{W} - \vec{b}) = \vec{0}$
$[X]^T[X]\vec{W} - [X]^T\vec{b} = \vec{0}$
$[X]^T[X]\vec{W} = [X]^T\vec{b} \Big/ ([X]^T[X])^{-1}$
$\vec{W} = ([X]^T[X])^{-1}[X]^T\vec{b}$
$\vec{W} = [X]^{\#}\vec{b}$ (**)
$[X]^{\#} = ([X]^T[X])^{-1}[X]^T$; generalized inverse of [X]
Vektor $\vec{b}$ je ograničen - to je vektor sa svim pozitivnim komponentama
Vektor $\vec{b}$ treba tako mijenjati da uvijek budu zadovoljeno: $b_i > 0$ za $i = 1, 2, \ldots, N$
To se postiže na ovaj način:
$\vec{b}(k+1) = \vec{b}(k) + \delta\vec{b}(k)$,   (***)

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

$\frac{\partial J}{\partial \vec{b}} = -([X]\vec{W} - \vec{b})$
Jednadžbu (*) zapišemo u vektorskom obliku:

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

gdje $\Big|[X]\vec{W}(k) - \vec{b}(k)\Big|$ označava apsolutnu vrijednost svake komponente vektora $\big[[X]\vec{W}(k) - \vec{b}(k)\big]$
Iz jednadnžbi (**) i (***) dobivamo:
$\vec{W}(k+1) = [X]^{\#}\vec{b}(k+1) = [X]^{\#}[\vec{b}(k) + \delta\vec{b}(k)] = [X]^{\#}\vec{b}(k) + [X]^{\#}\delta\vec{b}(k) = \vec{W}(k) + [X]^{\#}\delta\vec{b}(k)$
Označimo $\vec{e}(k) = [X]\vec{W}(k) - \vec{b}(k)$

Ho-kashyap / LMSE Algoritam:

$\vec{W}(1) = [X]^{\#}\vec{b}(1)$, $\vec{b}(1)$ proizvoljan uz uvjet $b_i > 0$ za sve i
$\vec{e}(k) = [X]\vec{W}(k) - \vec{b}(k)$
$\vec{W}(k+1) = \vec{W}(k) + c[X][\vec{e}(k) - |\vec{e}(k)|]$
$\vec{b}(k+1) = \vec{b}(k) + c[\vec{e}(k) - |\vec{e}(k)|]$
$\vec{W}(k+1)$ 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 sve komponente vektora $\vec{e}(k)$ ne pozitivne (ali ne jednake 0) razredi su linearno neseparatibilni!
- Kada je $\vec{e}(k) = \vec{0}$ $\quad \vec{W}(k)$ je rješenje!

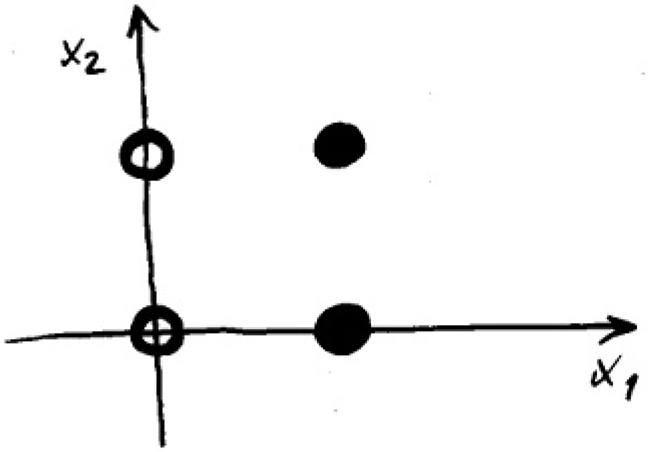
Osnovna značajka algoritma:

- 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 $[X]\vec{W}(k) = \vec{b}(k)$, tj. $\vec{W}(k)$ je rješenje
- Postupak brže konvergira negoli postupak perceptrona i nalazi "bolje" granice između razreda

PRIMJER 1:
$\omega_1 = (0,0)^T, (0,1)^T$
$\omega_2 = (1,0)^T, (1,1)^T$

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

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



Slika 5: Prikaz primjera 1

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

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

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 2:

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

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

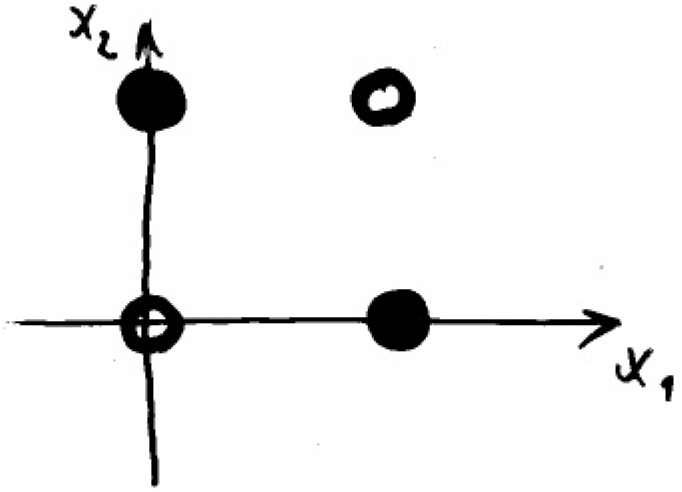
Razredi nisu linearno separatibilni (XOR problem)

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

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

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

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



Slika 6: Prikaz primjera 2

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

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

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

UČENJE ZA SLUČAJ M>2 RAZREDA

- Problem učenja koeficijenata linearne decizijske funkcije za M>2 razreda skupa za učenje može se riješiti poopćenim postupkom percetprona ali i postupkom Ho-Kashyapa.
  Podsjetimo se: 3. slučaja

  1. slučaj: svaki od M>2 razreda separatibilan 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 decizijskih funkcija)

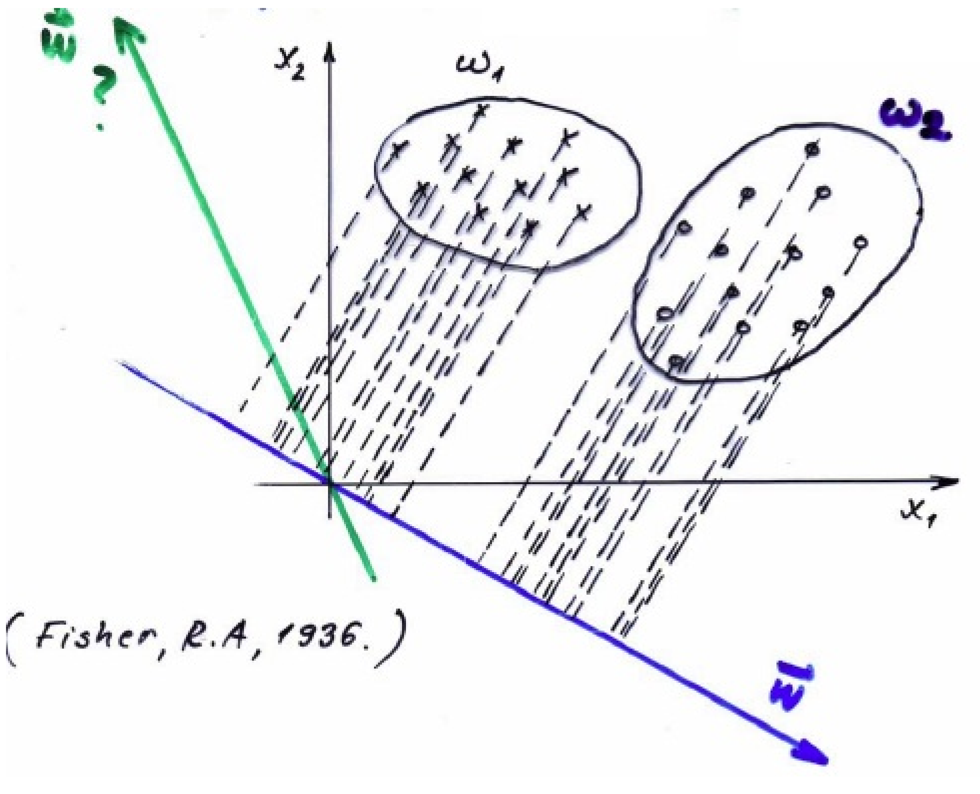
  3. slučaj: postoji M decizijskih funkcija: $d_1, d_2, \ldots, d_M$; $\vec{X} \in \omega_i$ ako $d_i(\vec{X}) > d_j(\vec{X})$; $i \neq j$ i=1,2,. . . ,M

  3. slučaj → Poopćeni algoritam perceptrona -istodobno određivanje 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 $d_i(\vec{X}(k)) > d_j(\vec{X}(k))$; $j = 1, 2, \ldots, M$, $j \neq i$ težinski vektori se ne "popravljaju":
$\vec{W}_j(k+1) = \vec{W}_j(k)$, $j = 1, 2, \ldots, M$
Ako je $d_i(\vec{X}(k)) \leq d_l(\vec{X}(k))$ popravljaju se vrijednosti težinskih vektora:
$\vec{W}_i(k+1) = \vec{W}_i(k) + c\vec{X}(k)$
$\vec{W}_l(k+1) = \vec{W}_l(k) - c\vec{X}(k)$
$\vec{W}_j(k+1) = \vec{W}_j(k)$ za $j = 1, 2, \ldots, M$, $j \neq i$, $j \neq l$, 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.

Fisherova linearna diskriminanta

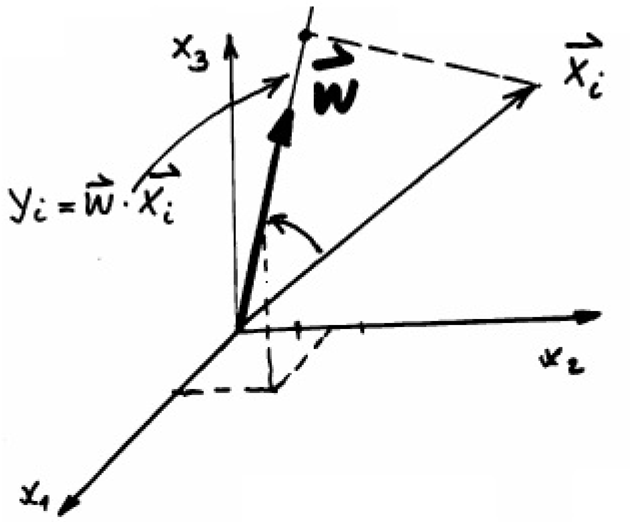
Jedan od pristupa linearnoj klasifikaciji:
d-dimenzionalan vektor značajki $\vec{X}$ reducirati na jednu dimenziju (R^{d1}) i tada ga upotrijebiti za klasifikaciju!
Prvo: M=2; ω_1 i ω_2



Slika 7: Prikaz FLD

Naći orijentaciju pravca na koji se projiciraju d-dimenzionalni uzorci $\vec{X}_i$ $i = 1, 2, \ldots, n$, ali tako da su projicirani uzorci separatibilni → CILJ "KLASIČNE" DISKRIMINANTNE ANALIZE

Skup od n d-dimenzionalnih uzoraka: $\vec{X_1}, \vec{X_2}, , \vec{X_n}$ od toga n_1 uzoraka čini podskup $\mathcal{D}_1$ (označeni s ω_1) i n_2 uzoraka u podskupu $\mathcal{D}_2$ (označeni s ω_2); $n = n_1 + n_2$
Tvorimo linearnu kombinaciju komponenti $\vec{X}$: $y_i = \vec{W}^T \vec{X}_i$, $i = 1, 2, \dots, n$
Dobivamo odgovarajući skup od n uzoraka $y_1, y_2, \dots, y_n$ koji su podijeljeni u podskupove $\mathcal{Y}_1$ i $\mathcal{Y}_2$.
Ako je $\left\|\vec{W}\right\| = 1$, svaki y_i je projekcija odgovarajućeg vektora $\vec{X}_i$ na pravac (u smjeru $\vec{W}$)

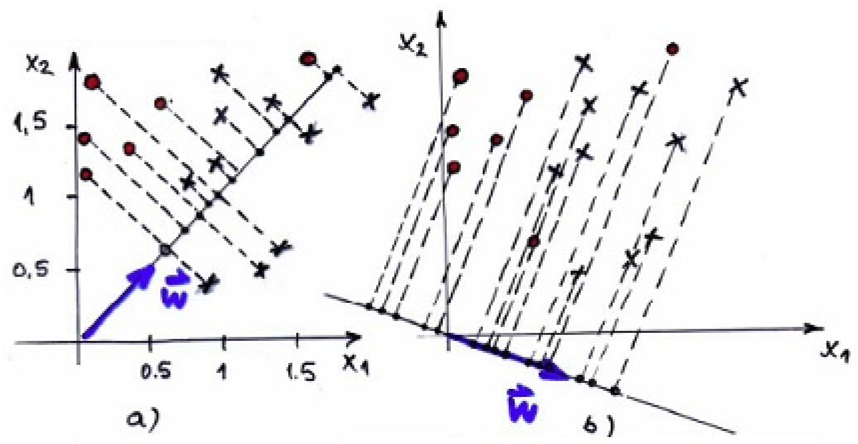


Slika 8: Prikaz projekcije $\vec{X}_i$ na $\vec{W}$

Veličina vektora $\vec{W}$ nema posebno značenje (ona samo skalira y_i) važan je smjer od $\vec{W}$!
- Ako zamislimo da uzorci iz ω_1 čine nakupinu, te da uzorci iz ω_2 čine drugu nakupinu ŽELIMO DA PROJEKCIJE NA PRAVAC ODREĐEN S $\vec{W}$ BUDU TAKVE DA SU DOBRO ODVOJIVE!

NAĆI NAJBOLJI SMJER $\vec{W}$!
Mjera odvojivosti između projciranih točaka: razlika srednjih vrijednosti projiciranih uzoraka;
Neka je $\vec{m}_i = \frac{1}{n_i} \sum_{\vec{X} \in \mathcal{D}_i} \vec{X}$; i=1,2
Srednja vrijednost projiciranih uzoraka $\tilde{m}_i = \frac{1}{n_i} \sum_{y \in \mathcal{Y}_i} y$
Da li je udaljesnot srednjih vrijednosti projiciranih uzoraka "dobra mjera"?

$\tilde{m}_i = \frac{1}{n_i} \sum_{\vec{X} \in \mathcal{D}_i} \vec{W}^T \vec{X} = \vec{W}^T \vec{m}_i$
$\tilde{m}_i$ je projekcija vektora $\vec{m}_i$
Razlika između srednjih vrijednosti projiciranih uzoraka:
$|\tilde{m}_1 - \tilde{m}_2| = \left|\vec{W}^T (\vec{m}_1 - \vec{m}_2)\right|$
Razliku možemo načiniti proizvoljno velkom skaliranjem $\vec{W}$!
Da bismo dobili dobro odvajanje projiciranih podataka želimo da "udaljenost" između srednjih vrijednosti projiciranih točaka bude relativno velika u odnosu

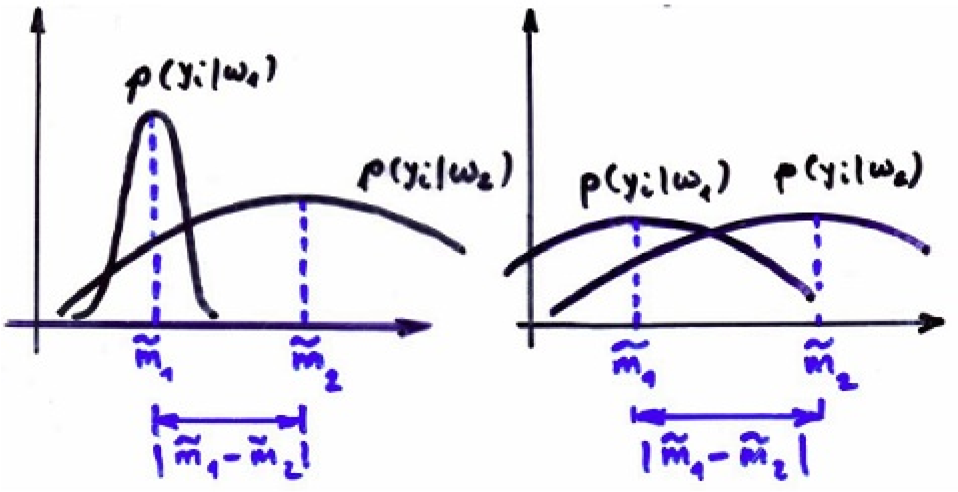


Slika 9: Odvojivost s obzirom na smjer $\vec{W}$

na neku mjeru standardne devijacije za svaki razred!
$J(\vec{W}) = \frac{(\tilde{m}_1 - \tilde{m}_2)^2}{\hat{G}^2_{y_1} + \hat{G}^2_{y_2}} = \frac{(\text{razlika srednjih vrijednosti})^2}{(\text{varijanca uzoraka unutar razreda})}$
Fisherova linearna diskriminanta određuje da linearna funkcija $\vec{W}^T\vec{X}$ za koju je kriterijska funkcija $J(\vec{W})$ maksimalna vodi najboljem razdvajanju između projiciranih skupova.
$J(\vec{W}) = \frac{(\tilde{m}_1 - \tilde{m}_2)^2}{\hat{G}^2_1 + \hat{G}^2_2}$, gdje je $\hat{G}^2_i$ mjera raspršenosti podataka (projiciranih) unutar razreda i.
Umjesto varijanci $\hat{G}^2_i$ možemo definirati raspršenost projiciranih podataka unutar razreda i kao:
$\tilde{s_i}^2 = \sum_{y \in \mathcal{Y}_i} (y - \tilde{m}_i)^2$
$\frac{1}{n}(\tilde{s_1}^2 + \tilde{s_2}^2)$ je procjena varijance podataka i $\tilde{s_1}^2 + \tilde{s_2}^2$ je ukupna mjera raspršenosti projiciranih podataka unutar razreda (total within-class scatter).
$J(\vec{W}) = \frac{(\tilde{m}_1 - \tilde{m}_2)^2}{\tilde{s_1}^2 + \tilde{s_2}^2}$
želimo dobiti J(.) kao eksplicitnu funkciju od $\vec{W}$;

Definirajmo matricu S_i:
$S_i = \sum_{\vec{X} \in \mathcal{D}_i} (\vec{X} - \vec{m}_i)(\vec{X} - \vec{m}_i)^T$ (raspršenost unutar razreda ω_i)
$S_W = S_1 + S_2$

$\tilde{s_i}^2 = \sum_{\vec{X} \in \mathcal{D}_i} (\vec{W}^T\vec{X} - \vec{W}^T\vec{m}_i)^2 = \sum_{\vec{X} \in \mathcal{D}_i} \vec{W}^T \overbrace{(\vec{X} - \vec{m}_i)(\vec{X} - \vec{m}_i)^T}^{S_i} \vec{W}$
$\tilde{s_i}^2 = \vec{W}^T S_i \vec{W}$
suma: $\tilde{s_1}^2 + \tilde{s_2}^2 = \vec{W}^T S_W \vec{W}$ (*)
slično:
$(\tilde{m}_1 + \tilde{m}_2)^2 = (\vec{W}^T\vec{m}_1 - \vec{W}^T\vec{m}_2)^2 = \vec{W}^T(\vec{m}_1 - \vec{m}_2)(\vec{m}_1 - \vec{m}_2)^T\vec{W} = \vec{W}^T S_B \vec{W}$ (**)
$S_B = (\vec{m}_1 - \vec{m}_2)(\vec{m}_1 - \vec{m}_2)^T$
$S_W = S_1 + S_2$ - within-class scatter matrix
S_B - between-class scatter matrix
S_W i S_B $\rightarrow$ simetrična i pozitivno semidefinitna i občino nesingularna ako je



Slika 10: Usporedba srednjih vrijednosti

n>d d-dimenzionalnost podataka

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

Uzorci iz skupova podataka $\mathcal{D}_1$ i $\mathcal{D}_2$ se koriste za određivanje S_W i S_B.

Tražimo maksimum $J(\vec{W})$!

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

$\frac{\partial J(\vec{W})}{\partial \vec{W}} = \frac{\partial}{\partial \vec{W}}(\frac{\vec{W}^T S_B \vec{W}}{\vec{W}^T S_W \vec{W}}) = \vec{0} \qquad (\frac{u}{v})' = \frac{vu' - uv'}{v^2}$

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

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

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

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

$\lambda S_W \hat{\vec{W}} = S_B \hat{\vec{W}}$

<u>Generalizirani problem svojstvenog vektora</u>

Ako S_W^{-1} postoji, onda je <u>smjer</u>(pazi: samo smjer) $\vec{W}$-a

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

Za naš slučaj nije potrebno rješavati to na takav način da tražimo svojstvene vrijednosti i svojstvene vektore za $S_W^{-1} S_B$ zato što <u>$S_B\vec{W}$ je uvijek usmjerena kao $\vec{m}_1 - \vec{m}_2$</u>

Budući da nas faktor skaliranja za vektor $\vec{W}$ ne zanima (zanima nas samo njegov smjer!) možemo napisati rješenje za $\hat{\vec{W}}$:

$$\vec{m}_1 = \begin{bmatrix} m_{11} \\ m_{12} \\ m_{13} \end{bmatrix}_{[3\times 1]} \qquad \vec{m}_2 = \begin{bmatrix} m_{21} \\ m_{22} \\ m_{23} \end{bmatrix}_{[3\times 1]}$$

$S_B = (\vec{m}_1 - \vec{m}_2)(\vec{m}_1 - \vec{m}_2)^T$

$S_B = (\vec{m}_1 - \vec{m}_2)_{[3\times1]} \cdot (\vec{m}_1 - \vec{m}_2)^T_{[1\times3]} = S_{B[3\times3]}$
smjer $S_B \cdot \vec{W}$?
$S_B \cdot \vec{W} = (\vec{m}_1 - \vec{m}_2)_{[3\times1]} \cdot \underbrace{(\vec{m}_1 - \vec{m}_2)^T_{[1\times3]} \cdot \vec{W}_{[3\times1]}}_{k_{[1\times1]}}$
$S_B \cdot \vec{W} = (\vec{m}_1 - \vec{m}_2) \cdot k$; očito je $S_B\vec{W}$ usmjeren s $(\vec{m}_1 - \vec{m}_2)$
$\hat{\vec{W}} = S_W^{-1}(\vec{m}_1 - \vec{m}_2)$
$\hat{\vec{W}}$ u skladu s FLD nam određuje linearnu funkciju koja maksimizira omjer između raspresenja između razreda i raspršenja unutar razreda.

PROBLEM SVOJSTVENIH VRIJEDNOSTI I SVOJSTVENIH VEKTORA (EIGENVALUE/EIGENVECTOR PROBLEM)

Osnovno:
-Kvadratna matrica A $n \times n$
Da li postoji skalar λ i vektor $\vec{X}$ tako da vrijedi:
$A\vec{X} = \lambda\vec{X} \quad ?$
Drugim riječima, $A\vec{X}$ ima isti smjer u R^n kao $\vec{X}$ i skaliran je s faktorom λ.
Trivijalno rješenje: $\vec{X} = \vec{0}$
$(A - \lambda I)\vec{X} = \vec{0}$
$\vec{X} = (A - \lambda I)^{-1}\vec{0}$
Podsjetimo se $(A-\lambda I)^{-1}$ postoji ako je $|A-\lambda I| \neq 0$, $|\ |$ označava determinantu.
Netrivijalno rješenje za $\vec{X}$ zahtijeva $|A-\lambda I| = 0$ za $n\times n$ matricu to daje polinom po λ stupnja n $|A - \lambda I| = \lambda^n + a_1\lambda^{n-1} + \ldots + a_{n-1}\lambda + a_n = 0$; karakteristična jednadžba (polinom) matrice A $P(\lambda)$
n rješenja za λ - svojstvene vrijednosti od A (e-values)
a odgovarajući vektori $\vec{X}$ su - svojstveni vektori (e-vectors)

PRIMJER:

Naći svojstvene vrijednosti i njima pridružene svojstvene vektore za matricu $A = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}$.

$A = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}$

Tražimo λ i vektor $\vec{X}$ razlčit od $\vec{0}$ tako da vrijedi $A\vec{X} = \lambda\vec{X}$

$\begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$\begin{matrix} x_1 + 2x_2 = \lambda x_1 \\ 3x_1 + 2x_2 = \lambda x_2 \end{matrix}$ ili $\begin{matrix} (\lambda - 1)x_1 - 2x_2 = 0 \\ -3x_1 + (\lambda - 2)x_2 = 0 \end{matrix}$

Homogen sustav ima rješenje različito od 0 akko je determinanta matrica koeficijena jednaka 0.

$\begin{vmatrix} \lambda - 1 & -2 \\ -3 & \lambda - 2 \end{vmatrix} = \lambda^2 - 3\lambda - 4 = (\lambda - 4)(\lambda + 1)$

λ je svojstvena vrijednost matrice A akko je $\lambda = 4$ ili $\lambda = -1$

Uzmimo $\lambda = 4$

$3x_1 - 2x_2 = 0$

$-3x_1 + 2x_2 = 0$ ili $3x_1 - 2x_2 = 0$

$x_1 = \frac{2}{3}x_2$ uzmimo: $x_2 = 3$ onda je $x_1 = 2$

Dobivamo svojstveni vektor $\vec{X} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ koji pripada svojstvenoj vrijednosti $\lambda = 4$.

Svaki drugi vektor$\vec{X}$ pomnožen s k je također e-vektor

$\begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = 4 \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

$\begin{bmatrix} 1 \cdot 2 + 2 \cdot 3 \\ 3 \cdot 2 + 2 \cdot 3 \end{bmatrix} = \begin{bmatrix} 8 \\ 12 \end{bmatrix} = 4 \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

Uzmimo $\lambda = -1$

$\begin{matrix} (\lambda - 1)x_1 - 2x_2 = 0 \\ -3x_1 + (\lambda - 2)x_2 = 0 \end{matrix} \rightarrow \begin{matrix} -2x_1 - 2x_2 = 0 \\ -3x_1 - 3x_2 = 0 \end{matrix}$

$x + y = 0 \rightarrow x = -y$

$\vec{X} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ svojstveni vektor koji pripada e-vrijednosti $\lambda = -1$