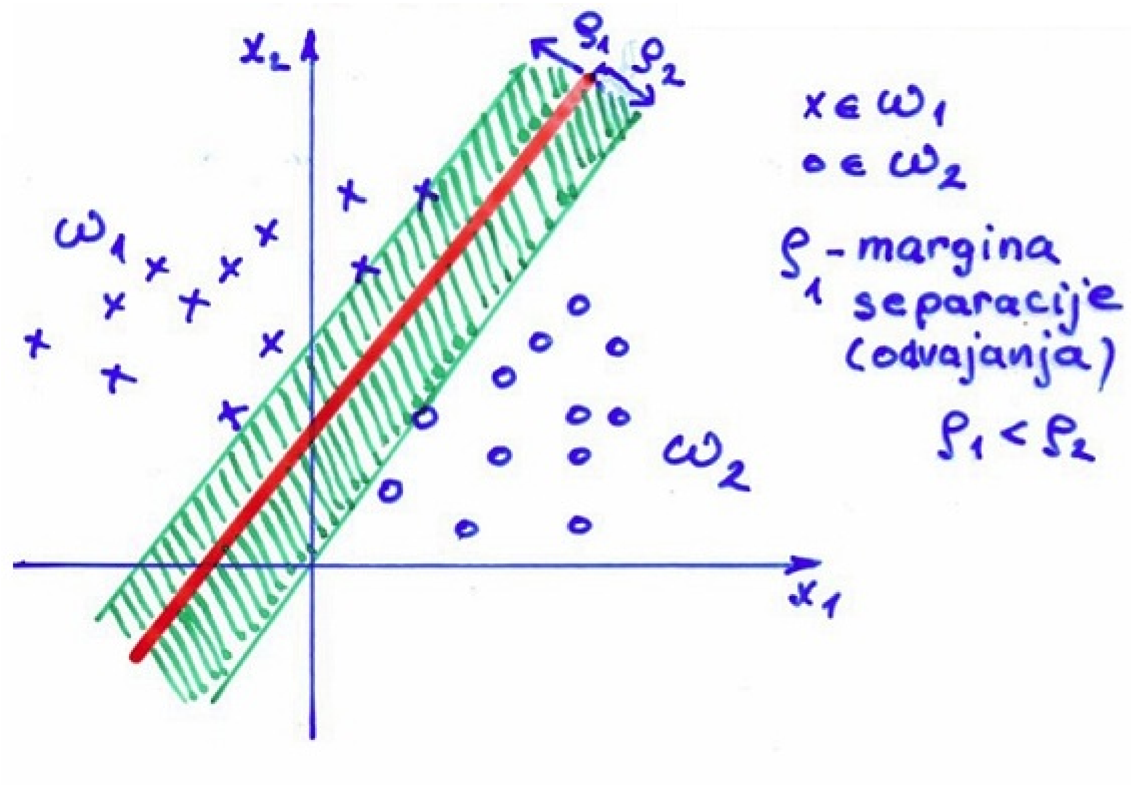
1 SVM - Support Vector Machines
Strojevi s potpornim vektorima

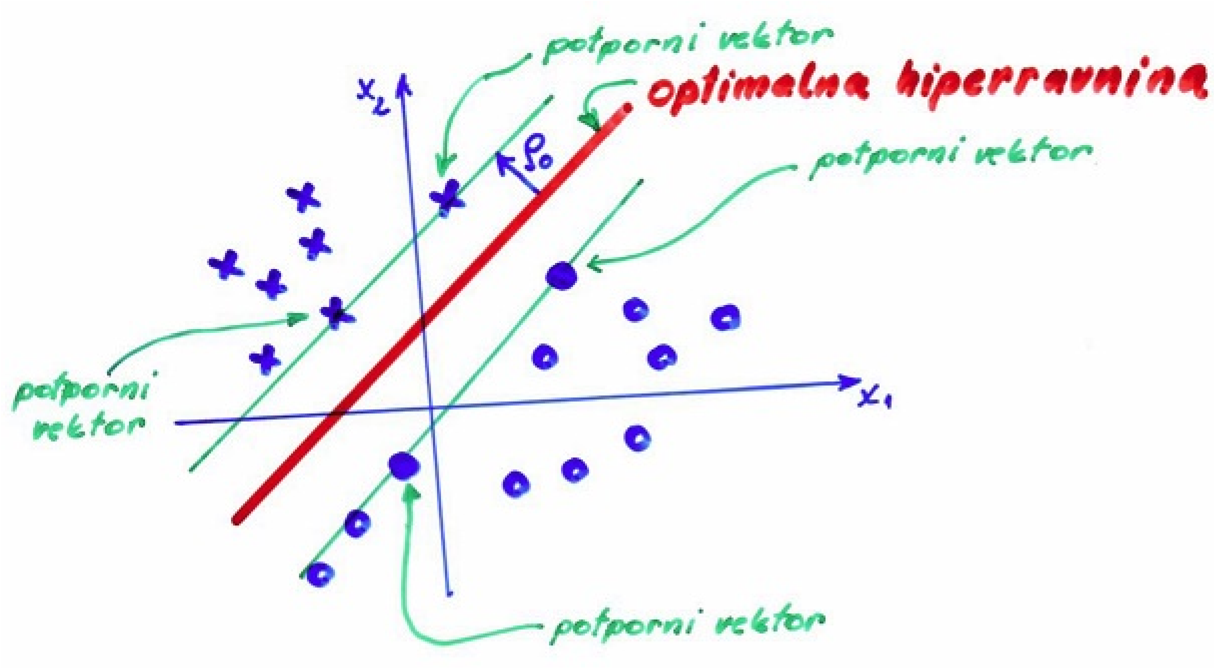
- Izvorno SVM je linearni stroj
- Osnovna zamisao SVM: konstrukcija hiperravnine kao decizijske plohe, ali tako da je margina odvajanja između "pozitivne" i "negativne" skupine uzoraka (za učenje) maksimalna.



Slika 1: Primjer razdvajanja

Imamo skup uzoraka za učenje:
$\left\{\left(\vec{X}_i, d_i\right)\right\}_{i=1}^{N}$, gdje je
$\vec{X}_i; i = 1, 2, \ldots, N$ ulazni vektor uzoraka za i-ti primjer
d_i - željeni odgovor klasifikatora
$\left.\begin{array}{l}\omega_1 \text{ -> } d_i = +1 \\ \omega_2 \text{ -> } d_i = -1\end{array}\right\}$ označeni uzorci
Pretpostavka : razredi ω_1 i ω_2 su linearno separatibilni
Jednadžba decizijske ravnine:
$\vec{W}^T\vec{X} + b = 0$
$\vec{X}$ - ulazni vektor
$\vec{W}$ - vektor težinskih koeficijenata
b - pomaknuće (w_0)

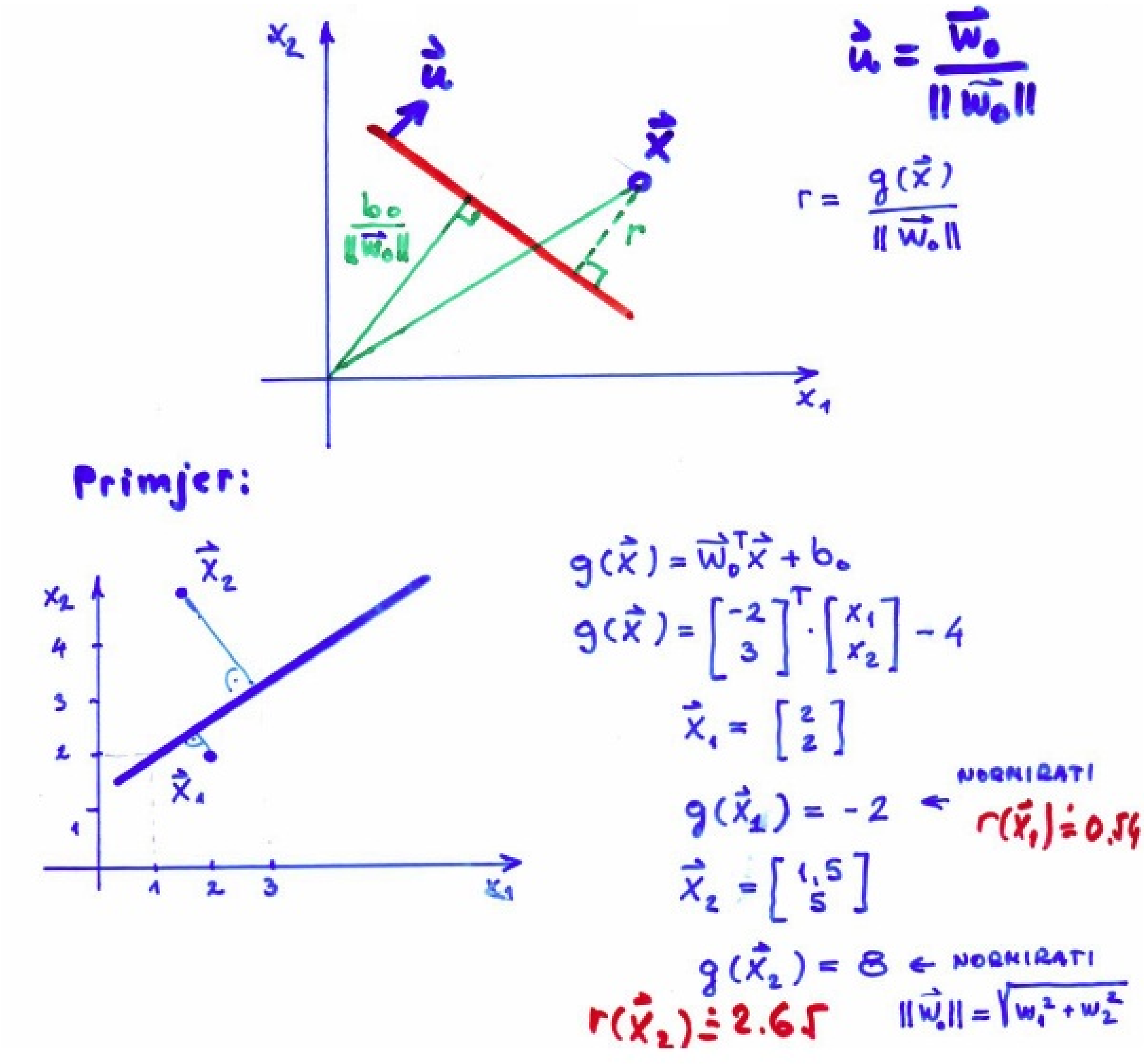
Vrijedi:
$\vec{W}^T\vec{X} + b \geq 0$ za $d_i = +1$
$\vec{W}^T\vec{X} + b < 0$ za $d_i = -1$
Margina: za zadani vektor težinskih koeficijenata $\vec{W}$ i pomaknuće b udaljenost između hiperravnine i najbliže točke (uzorka) u n-dimenzionalnom prostoru naziva se MARGINA ODVAJANJA (engl. margin of separation) i označit ćemo ju s ρ.
Cilj: Naći posebnu hiperravninu za koju je margina odvajanja ρ maksimalna. Takva hiperravnina naziva se OPTIMALNA RAVNINA.



Slika 2: Optimalna ravnina

Optimalna hiperravnina: $\underbrace{\left\{\vec{W}_0, b_0\right\}}_{\text{optimalne vrijednosti}}$

$\vec{W_0}^T\vec{X} + b_0 = 0$
Decizijska funkcija:
$g(\vec{X}) = \vec{W_0}^T\vec{X} + b_0$ daje mjeru udaljenosti $\vec{X}$-a od optimalne hiperravnine



Slika 3: Primjer decizijske funkcije

Par $(\vec{W}_0, b_0)$ mora zadovoljavati sljedeća ograničenja
$\vec{W_0}^T \vec{X}_i + b_0 \geq 1$ za $d_i = +1$ (1)
$\vec{W_0}^T \vec{X}_i + b_0 \leq -1$ za $d_i = -1$ (2)
$\vec{X}_i \in \left\{(\vec{X}_i, d_i)\right\}_{i=1}^{N}$
Naravno ovo vrijedi ako su uzorci linerano odvojivi.
Uvijek možemo skalirati $\vec{W}_0$ i b_0 tako da nejednadžbe (1) i (2) vrijede!
$r = \frac{|g(\vec{X})|}{||\vec{W}_0||}$
možemo skalirati $\vec{W}_0$ i b_0 tako da za najbliže (hiperravnini $g(\vec{X})$) uzorke iz ω_1 i ω_2 bude
$g(\vec{X}) = 1$ za ω_1

$g(\vec{X}) = -1$ za ω_2

Za uzorke (točke u n-dimenzionalnom prostoru) iz skupa za učenje i to za one za koje vrijedi
$\vec{W_0}^T \vec{X} + b_0 = 1$ za $d_i = +1$
$\vec{W_0}^T \vec{X} + b_0 = -1$ za $d_i = -1$
kažemo da su potporni vektori(support vectors).
Potporni vektori su one točke koje leže najbliže decizijskoj hiperravnini i zato se najteže klasificiraju. Zbog toga oni imaju izravan utjecaj na optimalni položaj decizijske hiperravnine.

Potporni vektor $\vec{X}^{(s)}$:
g$(^{(s)}) =^T \vec{X}^{(s)} + b_0 = \mp 1$ za $d^{(s)} = \mp 1$
Algebarska udaljenost potpornog vektora $\vec{X}^{(s)}$ od optimalne hiperravnine je
$r = \frac{\left| g(\vec{X^{(s)}}) \right|}{\left\| \vec{W_0} \right\|}$
$r = \begin{cases} \frac{1}{\left\| \vec{W_0} \right\|} & \text{ako je } \; d^{(s)} = +1 \\ -\frac{1}{\left\| \vec{W_0} \right\|} & \text{ako je } \; d^{(s)} = -1 \end{cases}$
gdje znak + označava da $\vec{X}^{(s)}$ leži na pozitivnoj strani optimalne hiperravnine a - predznak pokazuje da je $\vec{X}^{(s)}$ na negativnoj strani optimalne hiperravnine.
ρ - optimalna vrijednsot MARGINE ODVAJANJA između dva razreda koji definiraju skup uzoraka za učenje
$\rho = 2r = \frac{2}{\left\| \vec{W_o} \right\|}$
iz $\rho = 2r = \frac{2}{\left\| \vec{W_o} \right\|}$ slijedi da se maksimiziranje margine odvajanja temelji na minimizaciji norme vektora težinskih koeficijenata $\vec{W_0}$.
Optimalna hiperravnina: $\vec{W_0}^T \vec{X} + b = 0$ je jedinstvena u tom smislu da vektor $\vec{W_0}$ daje maksimalnu separaciju između pozitivnih i negativnih uzoraka iz skupa za učenje.
CILJ: Razvoj djelotvorne procedure (uporabom skupa uzoraka za učenje) tako da nađemo optimalnu hiperravninu uz zadovoljenje ograničenja:
$d_i \cdot (\vec{W}^T \vec{X}_i + b) \geq 1$ za $i = 1, 2, \ldots, N$
Formalno postavljen problem:

- Zadan je skup uzoraka za učenje $\left\{ (\vec{X}_i, d_i) \right\}_{i=1}^{N}$
- Nađi optimalnu vrijednost vektora težinskih koeficijenata $\vec{W}$ i pomaknuće b tako da su zadovoljena ograničenja
  $d_i \cdot \vec{W}^T \vec{X}_i + b \geq 1$ za $i = 1, 2, \ldots, N$
  a pri tomu vektor težniskih koeficijenata $\vec{W}$ minimizira kriterijsku funkciju
  $J(\vec{W}) = \frac{1}{2} \vec{W}^T \vec{W} \qquad \vec{W}^T \vec{W} = \left\| \vec{W} \right\|^2$ |
  -nelinearni optimizacijski zadatak sa skupom linearnih nejednadžbi

Optimizacijski problem riješiti metodom Lagrangeovih multiplikatora
Primjer:
Određivanje vezanih ekstrema funkcije $z = f(x, y)$ uz uvjet $\varphi(x, y) = 0$ svodi se na računanje slobodnih ekstrema Lagrangeove funkcije:
$F = f(x, y) + \lambda \varphi(x, y)$

$\frac{\partial F}{\partial x} = 0; \frac{\partial F}{\partial y} = 0; \varphi(x, y) = 0$
Iz tog se sustava jednadžbi određuju vrijednosti x,y i Lagrangeov multiplikator λ.

- Ako je $d^2F < 0$ u izračunatoj točki, funkcija z=f(x,y) ima <u>maksimum</u>
- Ako je $d^2F > 0$ u izračunatoj točki, funkcija z=f(x,y) ima <u>minimum</u>

Tražimo ekstrem funkcije:
$z = x + 2y$ uz uvjet $x^2 + y^2 = 5$

- Lagrangeova funkcija: $F = x + 2y + \lambda(x^2 + y^2 - 5)$
- Računamo:
  $F_x = \frac{\partial F}{\partial x} = 1 + 2x\lambda$
  $F_y = \frac{\partial F}{\partial y} = 2 + 2y\lambda$

- iz sustava jednadžbi:
  $1 + 2\lambda x = 0$
  $2 + 2\lambda y = 0$
  $x^2 + y^2 = 5$
  slijedi: $x = -\frac{1}{2\lambda}$; $y = -\frac{1}{\lambda}$
  uvrštavamo z 3. jednadžbu:
  $\frac{1}{4\lambda^2} + \frac{1}{\lambda^2} = 5 \Big/ 4\lambda^2$
  $5(1 - 4\lambda^2) = 0$
  $\lambda_{1,2} = \pm\sqrt{\frac{1}{4}}$
  $\lambda_1 = \frac{1}{2}$; $\lambda_2 = -\frac{1}{2}$
  za $\lambda_1 = \frac{1}{2}$ dobivamo : $x_1 = -1$ $y_1 = -2$
  za $\lambda_2 = -\frac{1}{2}$ dobivamo : $x_2 = 1$ $y_2 = 2$
  Računamo:
  $d^2F = \frac{\partial F^2}{\partial x^2} dx^2 + 2\frac{\partial^2 F}{\partial x \partial y} dxdy + \frac{\partial F^2}{\partial y^2} dy^2$
  $F_{xx} = 2\lambda$; $F_{yy} = 2\lambda$; $F_{xy} = 0$
  $d^2F = 2\lambda dx^2 + 2\lambda dy^2 = 2\lambda(dx^2 + dy^2)$
  za $(\lambda_1 = \frac{1}{2})$ $d^2F > 0$ minimum
  za $(\lambda_2 = -\frac{1}{2})$ $d^2F < 0$ maksimum
  funkcije f(x,y).

$J(\vec{W}) = \frac{1}{2}\vec{W}^T\vec{W}$
$d_i \cdot (\vec{W}^T\vec{X}_i + b) \geq 1$ za $i = 1, 2, \ldots, N$
Lagrangeova funkcija:
$J(\vec{W}, b, \lambda) = \frac{1}{2}\vec{W}^T\vec{W} - \sum_{i=1}^{N} \lambda_i \left[d_i \cdot (\vec{W}^T\vec{X}_i + b) - 1\right]$
λ_i - Lagrangeovi multiplikatori

a) $\frac{\partial J(\vec{W}, b, \lambda)}{\partial \vec{W}} = \vec{0}$

b) $\frac{\partial J(\vec{W}, b, \lambda)}{\partial b} = 0$

c) $\lambda_i \left[d_i \cdot (\vec{W}^T \vec{X}_i + b) - 1\right] = 0,\, i = 1, 2, \ldots, N$

d) $\lambda_i \geq 0,\, i = 1, 2, \ldots, N$

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

- Traženi vektor $\vec{W}$ određen je s $N_s \leq N$ vektora uzoraka
  $\vec{W} = \sum_{i=1}^{N} \lambda_i d_i \vec{X}_i,\, \lambda_i \neq 0$
  Vektor $\vec{W}$ je optimalno rješenje!

- Budući da je skup ograničenja
  $\lambda_i \left[d_i \cdot (\vec{W}^T \vec{X}_i + b) - 1\right] = 0,\, i = 1, 2, \ldots, N$ potporni vektori leže u dvije hiperravnine:
  $\vec{W}^T \vec{X} + b = \pm 1$
  Potporni (support) vektori su oni vektori koji leže tako da su NAJBLIŽI hiperravnini linearnog klasifikatora i određuju kritične elemente skupa za učenje.

- Vektori $\vec{X}_i$ za koje je $\lambda_i = 0$ mogu ležati izvan pojasa odvajanja, ali mogu ležati, također na jednoj od hiperravnina

- Rezultirajuća (optimalna) hiperravnina je neosjetljiva na broj i položaj takvih vektora

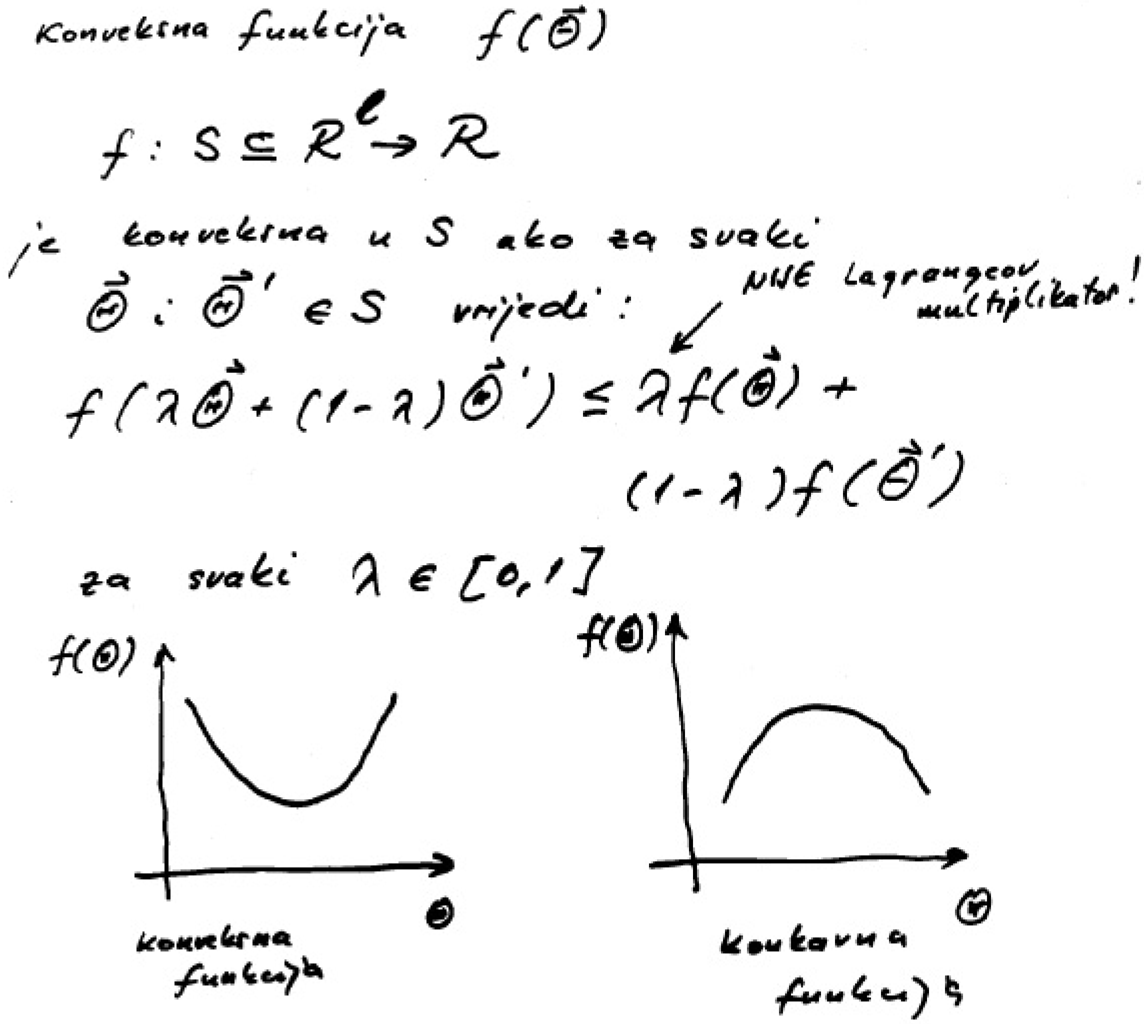
- $\vec{W}$ je eksplicitno određena, b se može dobiti iz jednog od uvjeta
  $\lambda_i \left[d_i \cdot (\vec{W}^T \vec{X}_i + b) - 1\right] = 0,\, i = 1, 2, \ldots, N$ za $\lambda_i \neq 0$ (*)
  U praksi, b se obično računa kao srednja vrijednost dobivena uporabom svih uvjeta tog tipa(*).

Optimalna hiperravnina linearnog klasifikatora je jedinstvena.

- $J(\vec{W})$ je konveksna (strogo)

- Nejednadžbe su linearne

lokalni minimum je ujedno i globalni!



Slika 4: Konveksno-konkavno

Lagrangeov dualni problem

- Optimizacijski zadatak: minimiziraj $J(\vec{W})$ uz ograničenje $\varphi_i(\vec{W}) \geq 0$, $i = 1, 2, \ldots, N$
  Lagrangeova funkcija $J(\vec{W}.\vec{\lambda}) = J(\vec{W}) - \sum_{i=1}^{N} \lambda_i \varphi_i(\vec{W})$
  Neka je $J^*(\vec{W}, \vec{\lambda}) = \max_{\vec{\lambda}}(J(\vec{W}, \vec{\lambda}))$
  Budući da je $\vec{\lambda} \geq \vec{0}$ i $\varphi_i(\vec{W}) \geq 0$
- Maksimalna vrijednost Lagrangeove funkcije onda kad je $\lambda_i = 0$; $i = 1, 2, \ldots, N$ ili kada je $\varphi_i(\vec{W}) = 0$ (ili oboje) u tom slučaju je $J^*(\vec{W}, \vec{\lambda}) = J(\vec{W})$
  Originalni problem je ekvivalentan sa:
  $\min_{\vec{W}} J(\vec{W}) = \min_{\vec{W}} \max_{\vec{\lambda}} J(\vec{W}, \vec{\lambda})$

Dualni problem:

$$\max_{\vec{\lambda}\geq\vec{0}} \underbrace{\min_{\vec{W}} J(\vec{W},\vec{\lambda})}$$

rješenje ovog dijela

$\vec{W} = \sum_{i=1}^{N} \lambda_i d_i \vec{X}_i$

i

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

Detekcija živosti ruke

$$\left.\begin{array}{l}\text{85 slika IR}\\ \text{29 živih ruku}\\ \text{56 neživih (umjetnih) ruku (2 tipa)}\end{array}\right\}\text{skup za ispitivanje}$$

SVM

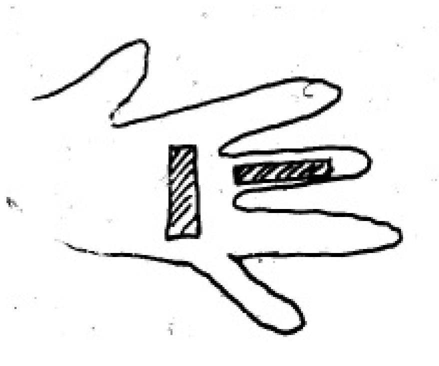
45 slika za učenje

40 slika za ispitivanje

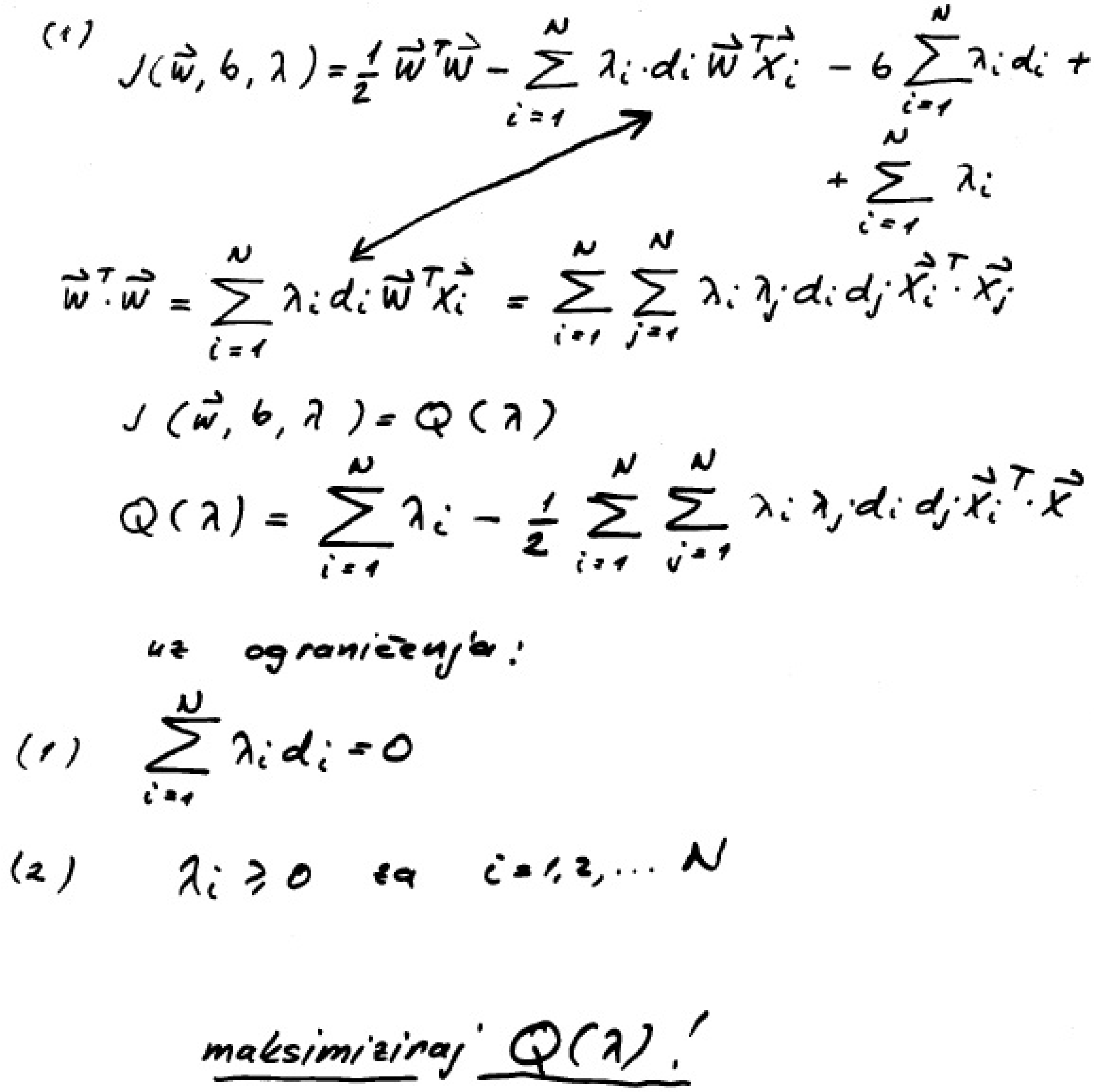
žive ruke $\notin$ skupu za učenje

umjetne ruke ne pripadaju skupu za učenje

SVM s linearnom jezgrom 5%pogreške



Slika 5:



Slika 6: Primjer SVM

Dualni problem - Lagrangeova dualnost (Wolfe dual representation)

Ako prvotni problem ima optimalno rješenje tada dualni problem ima također optimalno rješenje i odgovarajuća optimalna rješenja su jednaka.
Maksimiziraj $J(\vec{W}, b, \lambda)$ uz
$\vec{W} = \sum_{i=1}^{N} \lambda_i d_i \vec{X}_i$

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

Nalazimo Lagrangeove multiplikatore koji daju optimalno rješenje!
Neke značajke dualnog pristupa: Kriterijska funkcija koja se treba maksimizirati zavisi samo od ulaznih uzoraka u obliku skupa skalarnog produkta $\left\{\vec{X_i}^T|\vec{x_j}\right\}_{(i,j)=1}^{N}$
Optimalno rješenje $\vec{W}=\vec{W_0}$
$\vec{W_0}=\sum_{i=1}^{N}\lambda_{0,i}d_i\vec{X_i}$ gdje je $\lambda_{0,i}$ optimalni Lagrangeov multiplikator
$b_0=1-\vec{W_0}^T\vec{X}^{(s)}$ za $d^{(s)}=1$

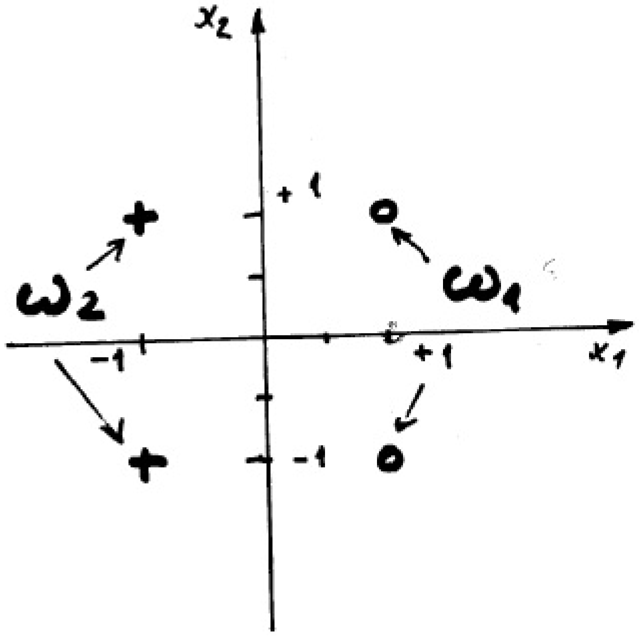
(1) $J(\vec{W},b,\lambda)=\frac{1}{2}\vec{W}^T\vec{W}-\sum_{i=1}^{N}\lambda_i\left[d_i\cdot(\vec{W}^T\vec{X_i}+b)-1\right]$
(2) $\vec{W}=\sum_{i=1}^{N}\lambda_i d_i\vec{X_i}$
(3) $\sum_{i=1}^{N}\lambda_i d_i=0,\ \lambda_i\geq 0,\quad i=1,2,\ldots,N$

Zamjenom (2) i (3) u (1) i nakon uređivanja dobiva se :
(**) $\max_\lambda(\sum_{i=1}^{N}\lambda_i-\frac{1}{2}\sum_{i,j}^{N}\lambda_i\lambda_j d_i d_j\vec{X_i}^T\vec{X_j})$ uz uvjet $\sum_{i=1}^{N}\lambda_i d_i=0$, $\lambda_i\geq 0$

Optimalni Lagrangeovi multiplikatori se računaju optimiziranjem (MAKSIMIZIRANJEM) izraza (**), a optimalna se hiperravnina dobiva $\vec{W}=\sum_{i=1}^{N}\lambda_i d_i\vec{X_i}$, gdje su λ_i optimalni Lagrangeovi multiplikatori

PRIMJER:
$\omega_1=\left\{[1,1]^T,[1,-1]^T\right\}$
$\omega_2=\left\{[-1,1]^T,[-1,-1]^T\right\}$
$g(\vec{X})=\vec{W}^T\vec{X}+b=0$
$w_1x_1+w_2x_2+b=0$

Ograničenja (linearne nejednadžbe):
$w_1+w_2+b-1\geq 0$
$w_1-w_2+b-1\geq 0$
$w_1-w_2-b-1\geq 0$
$w_1+w_2-b-1\geq 0$
$d_i(\vec{W}^T\vec{X_i}+b)\geq 1$
$J(\vec{W},b,\lambda)=\frac{1}{2}\vec{W}^T\vec{W}-\sum_{i=1}^{4}\lambda_i[d_i(\vec{W}^T\vec{X_i}+b)-1]$
$J(\vec{W},b,\lambda)=\frac{w_1^2+w_2^2}{2}-\lambda_1(w_1+w_2+b-1)-\lambda_2(w_1-w_2+b-1)-\lambda_3(w_1-w_2-b-1)-\lambda_4(w_1+w_2-b-1)$
KKT uvjeti su zadani sa:
$\frac{\partial J}{\partial w_1}=0\Rightarrow w_1=\lambda_1+\lambda_2+\lambda_3+\lambda_4$ (1)
$\frac{\partial J}{\partial w_2}=0\Rightarrow w_2=\lambda_1+\lambda_4-\lambda_2-\lambda_3$ (2)
$\frac{\partial J}{\partial b}=0\Rightarrow \lambda_1+\lambda_2-\lambda_3-\lambda_4=0$ (3)
$\lambda_1(w_1+w_2+b-1)=0$ (4)
$\lambda_2(w_1-w_2+b-1)=0$ (5)
$\lambda_3(w_1-w_2-b-1)=0$ (6)
$\lambda_4(w_1+w_2-b-1)=0$ (7)
$\lambda_1,\lambda_2,\lambda_3,\lambda,4\geq 0$



Slika 7: Primjer zadatka

7 jednadnžbi $\rightarrow$ 7 nepoznanica

Znamo rješenje s maksimalnom marginom (za ovaj jednostavan slučaj): $w_1 = 1$, $w_2 = 0$, $b = 0$

$g(\vec{X}) = X_1 = 0$

Uvrstimo $w_1 = 1$; $w_2 = 0$ i $b = 0$ u jednadžbe:

(1) $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1$

(2) $\lambda_1 + \lambda_4 - \lambda_2 - \lambda_3 = 0$

(3) $\lambda_1 + \lambda_2 - \lambda_3 - \lambda_4 = 0$

Sustav linearnih jednadžbi: 3 jednadžbe i 4 nepoznanice!

Očito $\rightarrow$ više od jednog rješenja!

Međutim svako od rješenja vodi do JEDINSTVENE (OPTIMALNE) LINIJE RAZDVAJANJA!

Na primjer:

(1)+(3) $2\lambda_1 + 2\lambda_2 = 1 \Rightarrow \lambda_1 = \frac{1-2\lambda_2}{2}$

(1)+(2) $2\lambda_1 + 2\lambda_4 = 1$

(2)+(3) $2\lambda_1 - 2\lambda_3 = 0 \Rightarrow \lambda_1 = \lambda_3$

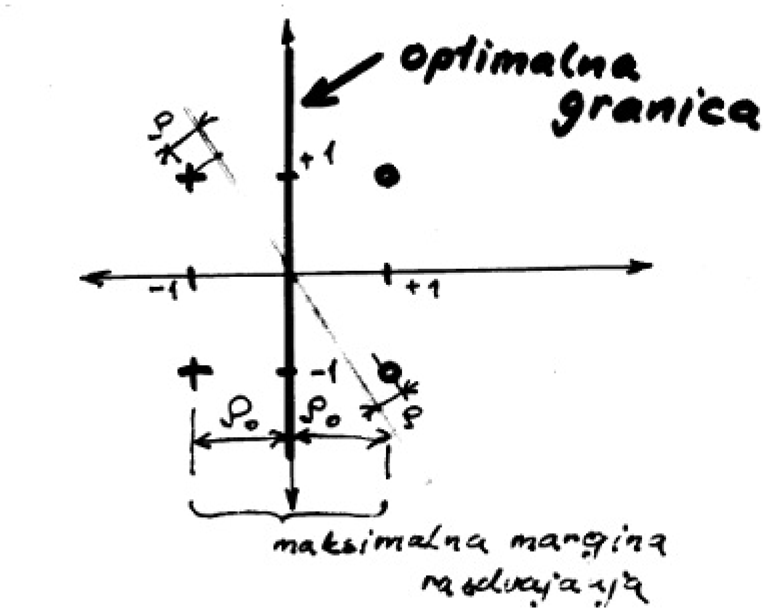
Uzmimo $\lambda_1 = \frac{1}{8} \Rightarrow \lambda_3 = \frac{1}{8}$

$\lambda_2 = \frac{1-2\lambda_1}{2} = \frac{1-2\frac{1}{8}}{2} = \frac{\frac{3}{4}}{2} = \frac{3}{8}$

$\lambda_4 = \frac{1-2\lambda_1}{2} = \frac{3}{8}$

$\vec{W} = \sum_{i=1}^{N=4} \lambda_i d_i \vec{X}_i = \lambda_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \lambda_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} - \lambda_3 \begin{bmatrix} -1 \\ 1 \end{bmatrix} - \lambda_4 \begin{bmatrix} -1 \\ -1 \end{bmatrix}$

$\vec{W} = \frac{1}{8} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \frac{3}{8} \begin{bmatrix} 1 \\ -1 \end{bmatrix} - \frac{1}{8} \begin{bmatrix} -1 \\ 1 \end{bmatrix} - \frac{3}{8} \begin{bmatrix} -1 \\ -1 \end{bmatrix}$



Slika 8: Prikaz margine zadatka

$\vec{W} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$\lambda_1 = \frac{1}{4}$, $\lambda_2 = \frac{1}{4}$, $\lambda_3 = \frac{1}{4}$, $\lambda_4 = \frac{1}{4}$

$\vec{W} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad g(\vec{X}) = x_1 = 0$

SVM za M>2 razreda?

Podsjetimo se :

- za svaki od razreda tražimo optimalnu decizijsku funkciju $g_i(\vec{X}), \quad i = 1, 2, \ldots, M$ tako da $g_i(\vec{X}) > g_j(\vec{X}), \quad \forall j \neq i$ ako je $\vec{X} \in \omega_i$

<u>za SVM tražimo</u> decizijsku funkciju $g_i(\vec{X}) = 0$ takva da bude optimalna hiperravnina koja odvaja razred ω_i od svih ostalih $g_i(\vec{X}) > 0$, za $\vec{X} \in \omega_i$, $g_i(\vec{X}) < 0$ inače

Klasifikacijsko pravilo: dodijeli $\vec{X}$ u ω_i ako $i = argmax_k \left\{ {}_k(\vec{X}) right \right\}.$