Strojno učenje – domaća zadaća 5

UNIZG FER, ak. god. 2016./2017.

Zadano: 15. 11. 2016. Rok: 18. 11. 2016.

Napomena: Zadatke možete rješavati samostalno ili u grupi. Ako zadatke rješavate u grupi, pobrinite se da svi članovi grupe pridonose rješenju i da ga naposlijetku svi razumiju. Po potrebi konzultirajte sve dostupne izvore informacija. Rješenja zadataka ponesite na iduće auditorne vježbe. Zabilježite sve nejasnoće i nedoumice, kako bismo ih prodiskutirali.

1. [***Svrha: Razumjeti izvod modela stroja potpornih vektora.***]

   (a) Definirajte, korak po korak, problem maksimalne margine (tvrda margina).

   (b) Definirajte, korak po korak, dualni problem maksimalne margine te pripadne KKT-uvjete koji vrijede u točci rješenja.

   (c) Koje su prednosti formulacije problema kao dualnoga optimizacijskog problema?

   (d) Napišite primarnu i dualnu formulaciju modela SVM.

   (e) Zašto kažemo da je SVM, uz dualnu formulaciju problema, neparametarski model? Koja je prednost takvog modela u smislu računalne učinkovitosti? Dolazi li taj prednost do izražaja kod treniranja ili kod uporabe modela (predikcije)?

2. [***Svrha: Isprobati izračuna modela potpornih vektora na konkretnom numeričkom primjeru i tako bolje razumjeti formule. Razumjeti povezanost primarne i dualne formulacije problema.***] Raspolažemo sljedećim primjerima za učenje:

$$
\begin{aligned}
\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_i = \\
\big\{((0,0),-1),((2,4),-1),((4,2),-1),((6,4),+1),((6,8),+1),((8,8),+1)\big\}
\end{aligned}
$$

   (a) Skicirajte primjere u ulaznom prostoru $\mathbb{R}^2$ i granicu maksimalne margine za te primjere. Napišite izraz za linearni model $h(\mathbf{x})$ koji odgovara toj granici.

   (b) Odredite širinu margine.

   (c) U ovom slučaju potporni vektori su $\mathbf{x}^{(3)} = (4,2)$ i $\mathbf{x}^{(4)} = (6,4)$. Odredite vektor Lagrangeovih koeficijenata $\boldsymbol{\alpha}$ temeljem izraza (7.5) iz skripte.

   (d) Upoznajte se s formulom (7.8) iz skripte (taj dio smo na predavanjima preskočili) te izračunajte pomak w_0.

   (e) Odredite klasifikaciju novog primjera $\mathbf{x}^{(7)} = (5,6)$ temeljem dualne reprezentacije granice.

3. [***Svrha: Razumjeti potrebu za mekom marginom. Znati izvesti problem meke margine SVM-a preko Lagrangeove dualnosti.***]

   (a) Objasnite motivaciju za uvođenje meke margine. Skicirajte primjer prenaučenosti uslijed meke margine, i to za linearno odvojiv i linearno neodvojiv slučaj.

(b) Formulirajte problem optimizacije meke margine.

(c) Definirajte dualni kvadratni problem za meku marginu.

(d) Krenuvši od KKT-uvjeta, dokažite da potporni vektori za koje vrijedi $\alpha_i \leqslant C$ leže na margini, a vektori za koje $\alpha_i = C$ unutar margine.

4. [***Svrha: Znati izvesti formulaciju algoritma SVM preko gubitka zglobnice. Razumijeti funkciju pogreške SVM-a.***]

   (a) Krenuvši od problema meke margine, izvedite gubitak zglobnice.

   (b) Pretpostavite da je ispravna klasifikacija primjera $\mathbf{x}^{(7)} = (5, 6)$ iz zadatka 2 negativna. Koliko iznosi gubitak koji primjer $\mathbf{x}^{(7)}$ nanosi modelu treniranom u zadatku 2?

   (c) Skicirajte pogrešku učenja i pogrešku ispitivanja kao funkciju od C. Kojem području odgovara prenaučenost a kojem podnaučenost?

   (d) Pokušajte odgovoriti: zašto SVM rezultira rijetkim modelima, unatoč tome što zapravo koristi L2-regularizaciju, za koju je poznato da ne rezultira rijetkim modelima? (Pomoć: usporedite gubitak zglobnice i gubitak logističke regresije.)

5 DOMAĆA ZADAĆA

1 DEF, KORAK PO KORAK, PROBLEM MAKSIMALNE (TVRDE) MARGINE.

a) MODEL: $h(x) = w^T \phi(x) + w_0$ ⇒ linearan model bez aktivacijske f-je

OZNAKE PR. ZA UČENJE: $y \in \{-1, 1\}$ (radi jednostavnosti).

PREDIKCIJA ZA NOVI PRIMJER x: $y = \mathrm{sgn}(h(x))$
↳ (PREDIKCIJA KLASE)

⇒ Pretpostavimo da su primjeri iz D linearno odvojivi:

↳ POSTOJE $\vec{w}$ i w_0 TAKVI DA: $h(x^{(i)}) \geq 0$ za $y^{(i)} = +1$, $h(x^{(i)}) < 0$ za $y^{(i)} = -1$ $\Big\}$ $\boxed{\forall (x^{(i)}, y^{(i)}) \in D \quad y^{(i)} h(x^{(i)}) \geq 0}$

↳ Nas zanima rješenje MAKSIMALNE MARGINE, odnosno NAJVEĆA UDALJENOST HIPERRAVNINE OD NAJBLIŽEG PRIMJERA.

→ PREDZNAČNA UDALJENOST PR. OD HIPERRAVNINE = $\boxed{d = \frac{h(x)}{\|\vec{w}\|}}$

→ Nas zanimaju samo rješenja kod kojih su sve točke ispravno klasificirane. U tom slučaju udaljenost od hiperravnine jednaka:

$$\boxed{d = \frac{h(x)}{\|\vec{w}\|} = \frac{y^{(i)} h(x^{(i)})}{\|\vec{w}\|} = \frac{y^{(i)}(\vec{w}^T \phi(x) + w_0)}{\|\vec{w}\|}}$$ → $y^{(i)} h(x^{(i)}) \geq \emptyset$ pa je $d \geq \emptyset$!

⇒ MARGINA - UDALJENOST DO NAJBLIŽEG PRIMJERA! (iz D)

↳ $$\boxed{d_{marg} = \frac{1}{\|\vec{w}\|} \min_i \{ y^{(i)} (\vec{w}^T \phi(x^{(i)}) + w_0) \}}$$

↳ Nas zanimaju vrijednosti parametara $\vec{w}$ i w_0 za koje će margina biti max! Odnosno tu udaljenost želimo maksimizirati.
Ako su primjeri iz D linearno odvojivi tad postoji samo jedna takva margina:

$$\boxed{\underset{\vec{w}, w_0}{\mathrm{argmax}} \left\{ \frac{1}{\|\vec{w}\|} \min_i \{ y^{(i)} (\vec{w}^T \phi(x^{(i)}) + w_0) \} \right\}}$$

→ Zbog pojednostavljenja problema, pretpostavimo da za primjer koji je najbliži margini vrijedi $\boxed{y^{(i)} (\vec{w}^T \phi(x^{(i)}) + w_0) = 1}$. Uvrštavanjem u gornji izraz pišemo:

$$\boxed{\underset{\vec{w}, w_0}{\mathrm{argmax}} \left\{ \frac{1}{\|\vec{w}\|} \right\}}$$, uz ograničenje $y^{(i)} (\vec{w}^T \phi(x^{(i)}) + w_0) \geq 1$

↳ ekvivalentno:

$$\boxed{\underset{\vec{w}, w_0}{\mathrm{argmax}} \left\{ \frac{1}{2} \|\vec{w}\|^2 \right\}}$$

⇒ ŠIRINA MAX MARGINE

$$\boxed{\frac{2}{\|\vec{w}\|}}$$

b) DEF. KORAK PO KORAK DUALNI PROBLEM MAX MARGINE TE PRIPADNE KKT UVJETE.

Problem iz a) podzadatka možemo riješiti uvođenjem LAGRANGEOVIH MULTIPLIKATORA.

⇒ LAGRANGEOVA FUNKCIJA (PRIMARNI PROBLEM)
$$\boxed{L(\vec{w}, w_0, \vec{\alpha}) = \frac{1}{2}\|\vec{w}\|^2 - \sum_{i=1}^{N} \alpha_i \left\{ y^{(i)} (\vec{w}^T \phi(x^{(i)}) + w_0) - 1 \right\}}$$

↳ $\alpha = (\alpha_1, \dots, \alpha_N)$ → vektor Lagrangeovih multiplikatora

→ PRELAZAK U DUALNI PROBLEM $\boxed{\partial L / \partial \vec{w} = \emptyset}$ i $\boxed{\partial L / \partial w_0 = \emptyset}$

$$\frac{\partial L}{\partial w} = \frac{1}{2}(2\vec{w}^T) - \sum_{i=1}^{N} \alpha_i \left\{ y^{(i)} \phi^T(x^{(i)}) \right\} = \emptyset \qquad \vec{w}^T = \sum_{i=1}^{N} \alpha_i y^{(i)} \phi^T(x^{(i)})$$

$\|w\|^2 = w^T w$

$= w^T w + w^T w$

$$\frac{\partial L}{\partial w_0} = -\sum_{i=1}^{N} \alpha_i y^{(i)} = \emptyset \qquad \boxed{\sum_{i=0}^{N} \alpha_i y^{(i)} = \emptyset} \quad \boxed{\vec{w} = \sum_{i=1}^{N} \alpha_i y^{(i)} \phi(x^{(i)})}$$

u $L(\vec{w}, w_0, \vec{\alpha})$

$$\Rightarrow \tilde{L}(\alpha) = \frac{1}{2}\overbrace{\|\vec{w}\|^2}^{w^T w} - \sum_{i=1}^{N} \alpha_i \left\{ y^{(i)} (\vec{w}^T \phi(x^{(i)}) + w_0) - 1 \right\} =$$

$$= \frac{1}{2} \sum_{i=1}^{N} \alpha_i y^{(i)} \phi^T(x^{(i)}) \sum_{i=1}^{N} \alpha_i y^{(i)} \phi(x^{(i)}) - \sum_{i=1}^{N} \alpha_i \left\{ y^{(i)} \left(\sum \alpha_i y^{(i)} \phi^T(x^{(i)}) \phi(x^{(i)}) + w_0 \right) - 1 \right\} =$$

$$= \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y^{(i)} y^{(j)} \phi^T(x^{(i)}) \phi(x^{(j)}) - \sum_{i=1}^{N} \alpha_i y^{(i)} \sum \alpha_i y^{(i)} \phi^T(x^{(i)}) \phi(x^{(i)}) - \overbrace{\sum \alpha_i y_i w_0}^{\emptyset} + \sum \alpha_i$$

$$\boxed{\tilde{L}(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y^{(i)} y^{(j)} \phi^T(x^{(i)}) \phi(x^{(j)})}$$

↳ DUALNI PROBLEM ZAHTJEVA MAKSIMIZACIJU OVOG IZRAZA! uz KKT uvjete:

⇒ i) DUALNI OPT. PROBLEM
$$\boxed{\tilde{L}(\alpha) = \underset{\alpha}{\arg\max} \left\{ \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y^{(i)} y^{(j)} \phi^T(x^{(i)}) \phi(x^{(j)}) \right\}}$$

ii) OGRANIČENJA NA α: $\alpha_i \geq \emptyset$ i $\sum_{i=1}^{N} \alpha_i y^{(i)} = \emptyset$

iii) KKT UVJETI: 1) $\boxed{\alpha_i \geq \emptyset}$ 2) $\boxed{y^{(i)} h(x^{(i)}) - 1 \geq \emptyset}$ 3) $\boxed{\alpha_i \left(y^{(i)} h(x^{(i)}) - 1 \right) = \emptyset}$

↳ ovo je ispunjeno za $\alpha_i = \emptyset$ ili $y^{(i)} h(x^{(i)}) = 1$

↳ ovi se vektori zanemaruju jer ne utječu na izlaz modela

↳ POTPORNI VEKTORI!

LIN. KOMB. POTP. VEKTORA DEFINIRA HIPERRAVNINU

SAMO POTPORNE VEKTORE ZADRŽAVAMO NAKON UČENJA MODELA!

c) PREDNOST FORMULACIJE PROBLEMA KAO DUALNOG OPTIMIZACIJSKOG PROBLEMA?

- Od $(n+1)$ primarnih varijabli problem smo sveli na N DUALNIH VAR.
  ↳ Smanjena računska složenost u slučaju većeg broja značajki od broja broja primjera za učenje (npr. klasifikacija teksta).

- Moguće korištenje JEZGRENOG TRIKA! Ovo omogućuje jednostavno preslikavanje u visokodimenzijski prostor značajki.

- Moguće korištenje posebnog optimizacijskog postupka SMO, čime smanjujemo računsku složenost s $\underline{O(n^3)}$ na $\underline{O(N^2)}$!

d) NAPIŠI PRIMARNU I DUALNU FORMULACIJU MODELA SVM

$$\boxed{h(x) = \underbrace{w^T\Phi(x) + w_0}_{\text{PRIMARNA FORM.}} = \underbrace{\sum_{i=1}^{N} \alpha_i y^{(i)} \Phi^T(x^{(i)})\Phi(x) + w_0}_{\text{DUALNA FORMULACIJA}}}$$

↳ RAČUNAMO SKAL. PRODUKT $\Phi^T(x^{(i)})\Phi(x)$

PR. X USPOREĐUJEMO SA SVIM PR. $X^{(i)}$ IZ SKUPA D

→ USPOREĐUJEMO KOLIKO JE X SLIČAN PR. $X^{(i)}$ U PROSTORU ZNAČAJKI

e) ZAŠTO KAŽEMO DA JE SVM NEPARAM. MODEL? PREDNOST TAKVOG MODELA U SMISLU RAČ. UČINKOVITOSTI? DOLAZI LI TA PREDNOST DO IZRAŽAJA KOD TRENIRANJA ILI UPORABE?

→ U dualnoj formulaciji, umjesto da pohranjujemo težine $\vec{w}$, moramo pohraniti primjere i njihove oznake. → EFEKTIVNO SMO DOBILI NEPARAMETARSKI MODEL.

$(x^{(i)}, y^{(i)})$ pohranjujemo

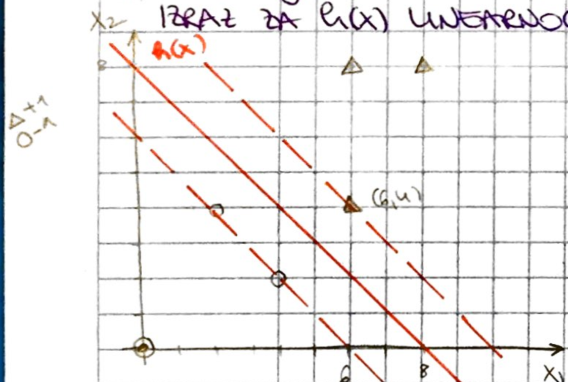
↳ U dualnoj formulaciji broj parametara ovisi o broju primjera.

→ PREDNOST: Obavljamo MANJE RAČ. OPERACIJA → manje +/* zbog uporabe Kernela

→ Ta prednost dolazi do izražaja kod PREDIKCIJE!

(2) $D = \{(x^{(i)}, y^{(i)})\}_i = \{((0,0),-1), ((2,4),-1), ((4,2),-1), ((6,4),+1), ((6,8),+1), ((8,8),+1)\}$

a) SKICIRAJTE PR. U UL. PROSTORU $\mathbb{R}^2$ I GRANICU MAX MARGINE ZA TE PR. NAPIŠI IZRAZ ZA h(x) LINEARNOG MODELA KOJI ODGOVARA TOJ GRANICI.



HIPERRAVNINA $x_2 = -x_1 + 8 \Rightarrow x_2 + x_1 - 8 = \emptyset$

h(x) za točke na margini mora biti ±1

$h(6,4) = 2 \neq 1$ → Treba podijeliti s 2!

$\boxed{h(x) = \frac{1}{2}(x_1 + x_2) - 4} \Rightarrow \vec{w} = \begin{bmatrix} -4 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$

↳ $h(x) = \underbrace{\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}}_{\vec{w}^T} \underbrace{\vec{X}}_{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}} + \underbrace{\begin{bmatrix} -4 \end{bmatrix}}_{w_0}$

b) ŠIRINA MARGINE? $= \boxed{\frac{2}{\|\vec{w}\|}}$ gdje je $\|\vec{w}\| = \sqrt{0.5^2 + 0.5^2} = 2\sqrt{2}$

$\vec{w} = [0.5 \; 0.5]$

c) POTP. VEKTORI su $x^{(3)} = (4,2)$, $x^{(4)} = (6,4)$. $\alpha = ?$ TEOREMOM (7.5)

$\boxed{\vec{w} = \sum_{i=1}^{N} \alpha_i y^{(i)} \Phi(x^{(i)})}$ $\qquad \vec{w} = \alpha_3 y^{(1)} \Phi(x^{(1)}) + \alpha_4 y^{(2)} \Phi(x^{(2)})$

$\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} = -\alpha_3 \begin{bmatrix} 4 \\ 2 \end{bmatrix} + \alpha_4 \begin{bmatrix} 6 \\ 4 \end{bmatrix}$

$\vec{w} = [0.5 \; 0.5]^T$

$\Phi = \begin{bmatrix} 1 & 4 & 2 \\ 1 & 6 & 4 \end{bmatrix}$

$y = [-1 \; 1]^T$

$0.5 = -4\alpha_3 + 6\alpha_4$

$0.5 = -2\alpha_3 + 4\alpha_4 \quad / \cdot -2$

$-0.5 = -8\alpha_4 + 6\alpha_4$

$-\frac{1}{2} = -2\alpha_4$

$\alpha_4 = 0.25$

$\alpha_3 = 0.25$

i $\boxed{\sum_{i=1}^{N} \alpha_i y^{(i)} = \emptyset}$

↳ $-\alpha_1 + \alpha_2 = \emptyset$

$\boxed{\alpha_1 = \alpha_2}$

$\alpha = [0.25 \; 0.25]$

d) IZR. POMAK w_0 i UPOZNAJ SE s (7.8)

$\alpha = [0.25 \ 0.25]$

$y = [-1 \ 1]$

$\Phi = \begin{bmatrix} 4 & 2 \\ 6 & 4 \end{bmatrix}$

$\Phi^3 = [4 \ 2] \quad \Phi^4 = [6 \ 4]$

$$\boxed{w_0 = \frac{1}{|S|} \sum_{i \in S} \left(y^{(i)} - \sum_{j \in S}^{N} \alpha_j y^{(j)} \Phi^T(x^{(j)}) \Phi(x^{(i)}) \right)}$$

S - SKUP INDEKSA POTPORNIH VEKTORA

$$w_0 = \frac{1}{2} \Big(y^{(3)} - \alpha_3 y^3 \Phi^T(x^{(3)}) \Phi(x^{(3)}) - \alpha_4 y^{(4)} \Phi^T(x^{(3)}) \Phi(x^{(4)}) + y^{(4)} -$$

$$- \alpha_3 y^{(3)} \Phi^T(x^{(4)}) \Phi(x^{(3)}) - \alpha_4 y^{(4)} \Phi^T(x^{(4)}) \Phi(x^{(4)}) \Big) =$$

$$= \frac{1}{2} \left(\cancel{-1} + 5 - \cancel{8} + \cancel{1} + \cancel{8} - 13 \right) \quad \Rightarrow \quad w_0 = -4$$

e) $x^{(7)} = (5, 6)$ KLASIF ? → TEMELJEM DUALNE REPREZENTACIJE GRANICE

$$h(x) = w^T \Phi(x) + w_0 = \sum_{i=1}^{N} \alpha_i y^{(i)} \Phi^T(x^{(i)}) \Phi(x) + w_0$$

$$h(x^{(7)}) = \alpha_3 y^{(3)} \Phi^T(x^{(3)}) \Phi(x^{(7)}) + \alpha_4 y^{(4)} \Phi^T(x^{(4)}) \Phi(x^{(7)}) + w_0$$

$$= (0.25)(-1)[4 \ 2]\begin{bmatrix} 5 \\ 6 \end{bmatrix} + (0.25)(1)[6 \ 4]\begin{bmatrix} 5 \\ 6 \end{bmatrix} - 4$$

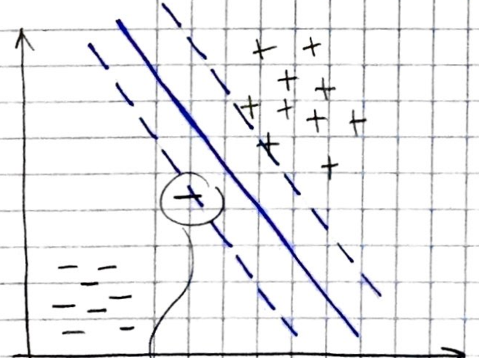
$$= -8 + 13.5 - 4 = 1.5 \qquad h(x^{(7)}) = 1.5 \Rightarrow \boxed{y(x^{(7)}) = +1}$$

3 a) MOTIVACIJA ZA UVOĐENJE ~~MEKE~~ MARGINE? SKICIRAJ PR. PRENAUČENOSTI USLIJED ~~MEKE~~ TVRDE MARG.) ZA LIN. ODV. I LIN. NEODV. SL.

- SUKLADNO COVEROVOM TM puno je vjerojatnije da će problem biti linearno neodvojiv nego da će biti linearno odvojiv.
- Ne želimo da model bude suviše nelinearan jer bi to moglo dovesti do prenaučenosti odnosno do prilagodbe šumu u podacima.
- Ne želimo da model izgubi svojstvo GENERALIZACIJE, stoga proširujemo prethodni model tako da on dozvoljava da neki primjeri ipak budu pogrešno klasificirani ⇒ MEKA MARGINA!

⇒ Kod meke margine dozvoljavamo da pr. budu na pogrešnoj strani granice, ali to kažnjavamo tim više što se pr. nalaze dublje na pogrešnoj strani granice. !

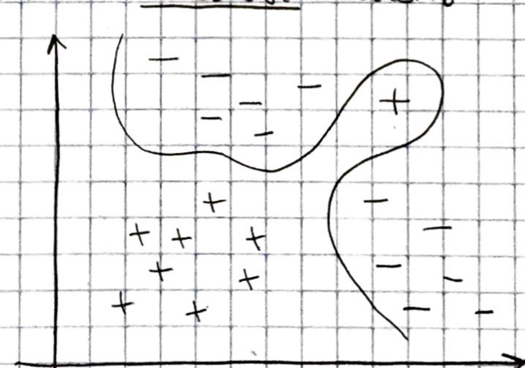
i) LIN. ODVOJIVI PRIMJERI



OUTLIER → ŠUM

UZROKUJE NEPOTREBNO USKU MARGINU! MEKA MARGINA BI BILA BLIŽE MINUSIMA, A OUTLIER BI KAZNILA

ii) LIN. NEODVOJIV SLUČAJ



MODEL PREVIŠE SLOŽEN! PRENAUČEN, PREVIŠE SMO SE PRILAGODILI PODACIMA (I ŠUMU)

b) FORMULIRAJ PROBLEM OPTIMIZACIJE MEKE MARGINE

$\boxed{\underset{\vec{w}, w_0}{\text{argmin}}\left\{\frac{1}{2}\|\vec{w}\|^2 + C\sum_{i=1}^{N} \xi_i\right\}}$ uz $\xi_i \geq 0$ i $y^{(i)}(\vec{w}^T\phi(x^{(i)})+w_0) \geq 1-\xi_i$

↳ $C = 1/\lambda > 0$ → C određuje kompromis između veličine margine i ukupne kazne. $\boxed{C\uparrow \text{ KAZNA} \uparrow \text{ SLOŽENOST MODELA} \uparrow}$

c) DEF. DUALNI KVADRATNI PROBLEM ZA MEKU MARGINU

$$\boxed{\underset{\alpha}{\text{argmax}} \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y^{(i)} y^{(j)} \phi^T(x^{(i)}) \phi(x^{(j)})}$$

↳ ISTO KAO ZA TVRDU, ali se razlikuju OGRANIČENJA: i) $0 \leq \alpha_i \leq C$ ii) $\sum_{i=1}^{N} \alpha_i y^{(i)} = 0$ $\quad i \in (1, N)$

d) NE ULAZI U ISPIT

4 a) IZVOD ZA GUBITAK ZGLOBNICE, KRENUVŠI OD PROBL. MEKE MARGINE

→ ZA PR. KOJI SU NA ISPRAVNOJ STRANI MARGINE VRIJEDI: $\underbrace{y^{(i)} h(x^{(i)}) \geq 1}_{\text{OVE PR. NE KAŽNJAVAMO!}}$

→ Pr. koji su na POGR. STR. MARGINE ili unutar nje kažnjavamo s $\boxed{\xi_i = |y^{(i)} - h(x^{(i)})|}$

↳ $\xi_i = |y^{(i)} - h(x^{(i)})|$ je jednako = $\boxed{1 - y^{(i)} h(x^{(i)})}$

↳ za $y = +1$ kažnjavamo pr. za koje je $h(x) < 1$

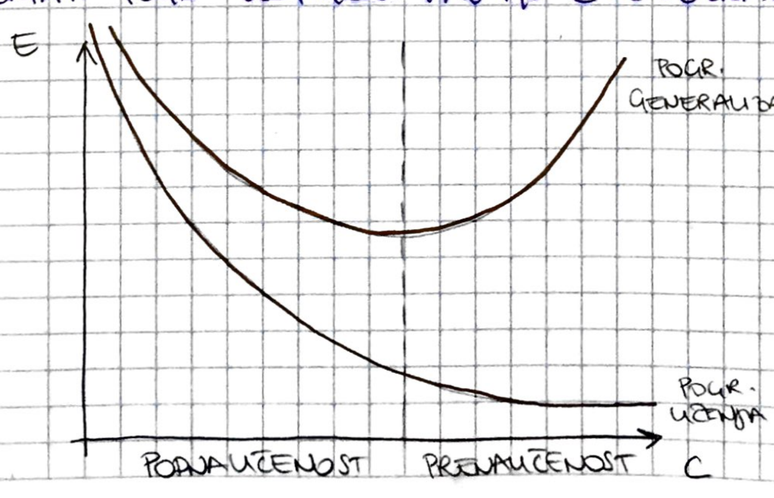
$y = -1$ -||- $h(x) > 1$

PREMA TOME, F-ja GUBITKA JE $\boxed{L(h(x), y) = \max(0, 1 - y h(x))}$

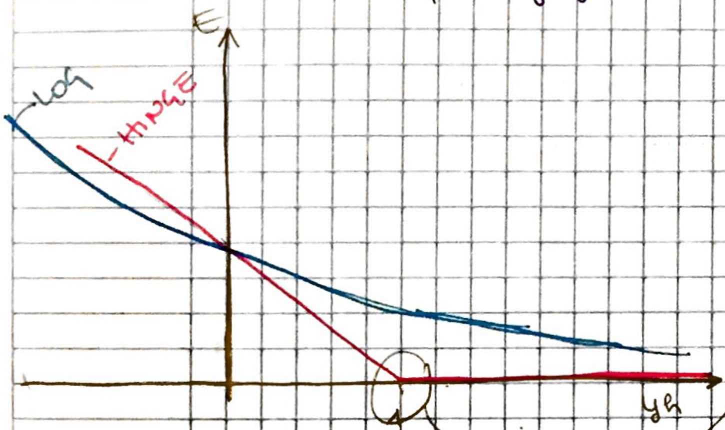
b) $x^{(7)} = (5, 6)$, $y = -1$ (a w=1 kao u prethodnom zad). GUBITAK KOJI x^7 nanosi?

$y(x^{(7)}) = -1 \quad h(x^{(7)}) = 1.5 \quad L(h(x), y) = \max(0, 1 - (-1)(1.5)) = \underline{\underline{2.5}}$

c) NACRTAJ POGR. UČ/GEN. KAO F-JU C-a. OZNAČI PRENAUČ/PODNAUČ.



d) ZAŠTO SVM REZULTIRA RIJETKIM MODELIMA UNATOČ TOME ŠTO ZAPRAVO KORISTI L2-REG, ZA KOJU JE POZNATO DA NE REZULTIRA RJ. MOD?



Iz slike vidimo kako je za ispravno klasificirane primjere funkcija zglobnice jednaka Ø, s obzirom da je više težina pritegnuto na Ø rezultirajući rijetkim modelima.

U drugu ruku imamo LOG. REG. kod koje se i ispravno klasificirani pr. kažnjavaju, stoga iako greška teži u Ø, ona neće nikada biti jednaka Ø, te stoga neće rezultirati rijetkim mod