[redacted]

3. Domaća zadaća

SVM I.

[redacted]

STROJ POTPORNIH VEKTORA

ZADACI ZA UČENJE

1.

a) želimo hiperravninu tako da maksimizira marginu (udaljenost hiperravnine do najbližeg primjera). Pretpostavimo da su primjeri lin. odvojivi

$\Rightarrow$ ovo iskoristimo u

$y^{(i)}(\vec{w}^T\vec{x}+w_0) \geq 0$ $\qquad d = \frac{h(x)}{\|w\|}$

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

$\Rightarrow$ to znamo da dobijemo najbliže primjere i sad oćemo naći takve parametre da dobijemo hiperravninu da ovo maximizira.

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

$\rightarrow$ skaliramo težine t.d. je za primjere na margini $h(\vec{x}) = 1$ ili -1.

↳ t.d. imamo da za sve primjere:

$$\boxed{y^{(i)}(\vec{w}^T\vec{x}+w_0) \geq 1}$$

$$\underset{\vec{w}, w_0}{\arg\max} \frac{1}{\|w\|} \sim \underset{\vec{w}, w_0}{\arg\min} \|w\|$$

Konačno: $\boxed{\underset{\vec{w}, w_0}{\arg\min} \frac{1}{2}\|w\|^2}$ uz ogr: $y^{(i)}(\vec{w}^T\vec{x}^{(i)}+w_0) \geq 1,$ $i = 1 \dots N$

[redacted]

b) Imamo ciljnu fju koju želimo minimizirat, ali uz neko ograničenje. Koristimo kvadratno programiranje kako bismo napravili optimizaciju ?!

min. $f(\vec{x})$

ograničenja $\quad g_i(\vec{x}) \leq 0 \quad i = 1 \dots m$

$\quad h_i(\vec{x}) = 0 \quad i = 1 \dots p$

$$L(\vec{x}, \vec{\alpha}, \vec{\beta}) = f(x) + \sum_{i=1}^{m} \alpha_i \, g_i(\vec{x}) + \sum_{i=1}^{p} \beta_i \, h_i(\vec{x})$$

↳ u točki rj. vrijede sljedeći KKT.:

$g_i(\vec{x}) \leq 0 \quad i = 1 \dots m$ — ograničenje nejednakosti

$h_i(\vec{x}) = 0 \quad i = 1 \dots p$ — ograničenje jednakosti

$\alpha_i \geq 0 \quad i = 1 \dots m$ — uvjet Lagrangeovog multiplikatora

$\alpha_i \cdot g_i(\vec{x}) = 0 \quad i = 1 \dots m$ — komplementarna labavost.

↳ ogr. ili je neaktivno ($\alpha_i = 0$) ili je aktivno i ispoštovano ($g_i(\vec{x}) = 0$)

Kako bi dobili dualnu fju, trebamo minimizirat Lagrangeovu fju po primarnom parametru $\vec{x}$.

$$\tilde{L}(\vec{\alpha}, \vec{\beta}) = \min_{\vec{x}} L(\vec{x}, \vec{\alpha}, \vec{\beta})$$

↳ maksimiziraj $\tilde{L}(\vec{\alpha}, \vec{\beta})$

uz ogr.

$\alpha_i \geq 0 \quad i = 1 \dots m$

[redacted] [redacted]

c)

$$L(\vec{w}, w_0, \vec{\mathcal{L}}) = \frac{1}{2}\|\vec{w}\|^2 - \sum_{i=1}^{N} \mathcal{L}_i \left(y^{(i)}(\vec{w}^T\vec{x} + w_0) - 1 \right)$$

↳ ovo želimo minimizirat po primarnim parametrima

$$\frac{\partial L}{\partial \vec{w}} = 0 \qquad \boxed{\vec{w} = \sum_{i=1}^{N} \mathcal{L}_i \, y^{(i)} x^{(i)}}$$

$$\frac{\partial L}{w_0} = 0 \qquad \boxed{\sum_{i=1}^{N} \mathcal{L}_i \, y^{(i)} = 0}$$

Nakon uvrštanja dobivamo:

$$\boxed{\tilde{L}(\vec{\mathcal{L}}) = \sum_{i=1}^{N} \mathcal{L}_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \mathcal{L}_i \mathcal{L}_2 \, y^{(i)} y^{(j)} (x^{(i)})^T x^{(j)}}$$

ovo maksimiziramo

uz ograničenja:

$$\mathcal{L}_i \geq 0 \qquad i = 1 \dots N$$

$$\sum_{i=1}^{N} \mathcal{L}_i \, y^{(i)} = 0$$

KKT uvjeti u točki rj.:

$$y^{(i)}(\vec{w}^T \vec{x}^{(i)} + w_0) \geq 1 \qquad i = 1 \dots N$$

$$\mathcal{L}_i \geq 0 \qquad i = 1 \dots N$$

$$\mathcal{L}_i \left(y^{(i)}(\vec{w}^T \vec{x}^{(i)} + w_0) - 1 \right) = 0 \qquad i = 1 \dots N$$

[redacted]

d)

Predmosti su što sada možemo primijeniti algoritam (SMO) koji lako iskorištava uvjete

Primarni problem imao je $n+1$ parametara, a dualni ima N parametara.

e)

$$h(\vec{x}; \vec{w}) = \vec{w}^T \vec{x} + w_0 \qquad \text{PRIMARNO}$$

$$h(\vec{x}; \vec{\alpha}) = \sum_{i=1}^{N} \alpha_i y^{(i)} x^T x^{(i)} + w_0 \qquad \text{DUALNO}$$

f)

Potporni vektori su vektori koji leže na margini i na temelju njih radimo predikciju.

Znamo da leže na rubu margine iz uvjeta:

$$\alpha_i \left(y^{(i)} \cdot h(x^{(i)}) - 1 \right) = 0$$

za njih je $\alpha_i > 0$, a onda mora vrijediti

$y^{(i)} h(x^{(i)}) = 1$ → a to smo definirali da vrijedi za primjere koji su najbliže margini !!

↳ tj. leže na njoj i to su potporni vektori !

g) Potrebno je radi sličnosti $\vec{x}^{(i)T}\vec{x}$. Naprosto ako neka značajka ima veće vrijednosti, dobit ćemo ogromnu sličnost

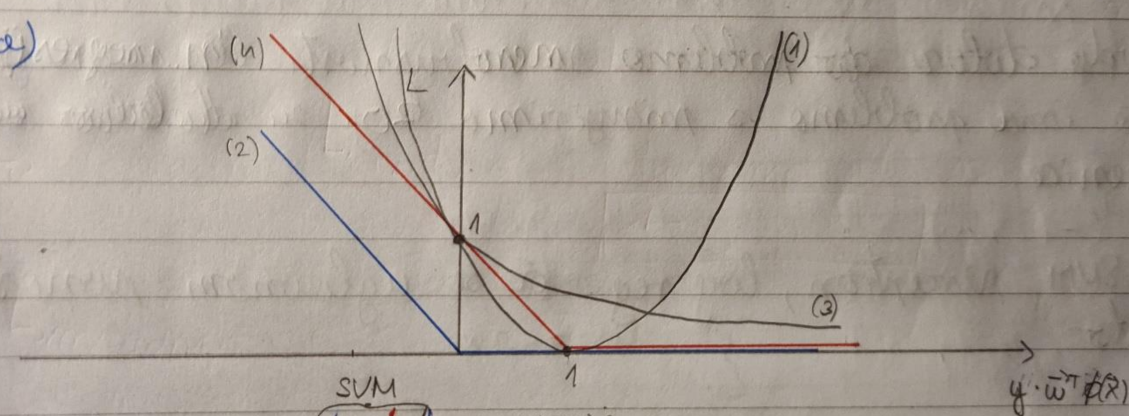
SVM II.

podnaučenost prenaučenost

$\boxed{3.}$ (1,1), 0 (2,3), 0 (4,3), 1 [redacted]

(0,2), 0 (3,1), 1

a)

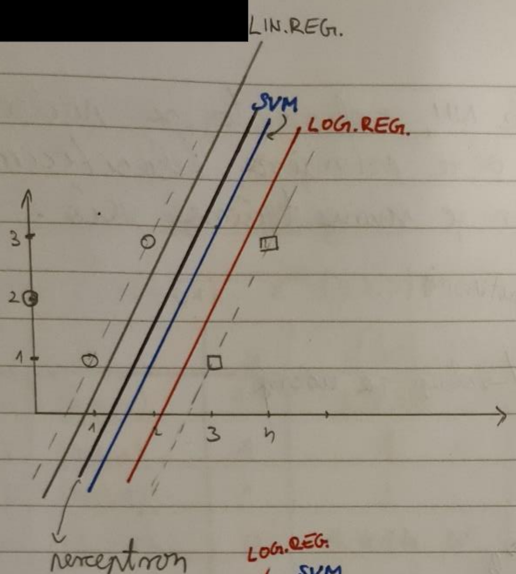


b)

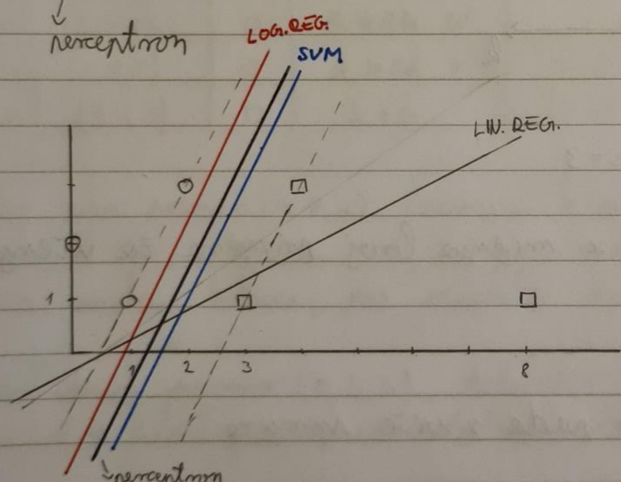
[redacted]

[redacted]

b)



c)



→ dolazi do problema nerobusnosti linearne regresije koja ima problem s primjerima koji su daleko od granice. Ostali algoritmi klasificiraju ispravno. Perceptron može ispast drugačije jer ovisi o početnim inicijalizacijama težina. SVM će biti isti jer su potporni vektori isti. Dok će se logistička regresija pomaknuti malo ulijevo. Prije je bilo desnije jer je bilo manje primjera.

d) Zato jer ne kažnjava primjere koji su točno klasificirani (ne nanose gubitak) pa je model jednostavniji, više je težina pritegnuto na 0.

JEZGRENE METODE

[redacted]

3

a) $$K(\vec{x},\vec{z}) = (\vec{x}^T z + 1)^2 = (\vec{x}^T\vec{z})^2 + 2\vec{x}^T\vec{z} + 1 =$$

$$= (x_1 z_1 + x_2 z_2)^2 + 2(x_1 z_1 + x_2 z_2) + 1 =$$

$$= x_1^2 z_1^2 + \underbrace{2x_1 z_1 x_2 z_2}_{\sqrt{2}x_1x_2\sqrt{2}z_1z_2} + x_2^2 z_2^2 + \underbrace{2x_1 z_1}_{\sqrt{2}x_1\sqrt{2}z_1} + \underbrace{2x_2 z_2}_{\sqrt{2}x_2\sqrt{2}z_2} + 1$$

$$[1, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2, x_1^2, x_2^2] = \phi(\vec{x})$$
$$[1, \sqrt{2}z_1, \sqrt{2}z_2, \sqrt{2}z_1z_2, z_1^2, z_2^2] - \phi(\vec{z})$$

$$K(\vec{x},\vec{z}) = (\vec{x}^T z + 1)^2 = \phi(\vec{x})\cdot\phi(\vec{z})$$

↳ Ovo K je Mercerova i to je bitno zato što tada znamo da odgovara skalarnom produktu u nekom prostoru značajki

b) $$\phi(\vec{x}) = [1, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2, x_1^2, x_2^2] \qquad x = (2,3)$$

$$= [\underbrace{1, 2.828, 4.243, 8.485, 4, 9}_{\text{5-dim prostor iz 2-dim prostora}}]$$

c)

→ s obzirom da smo preslikali u 5-D prostor, bit će veći ako uzmemo 6 primjera

d)

$x^{(1)}$ $(0,0), 0$ $\quad$ $(1,0)\ 1 = x^{(3)}$

$x^{(2)}$ $(1,1), 0$ $\quad$ $(0,1)\ 1 = x^{(4)}$

$K(\vec{x},\vec{z}) = (\vec{x}^T \vec{z} + 1)^2$

$\phi(\vec{x}^{(1)}) = (1, 0, 0, 0, 0, 0)\ , 0$

$\phi(\vec{x}^{(2)}) = (1, \sqrt{2}, \sqrt{2}, \sqrt{2}, 1, 1)\ , 0$

$\phi(\vec{x}^{(3)}) = (1, \sqrt{2}, 0, 0, 1, 0)\ , 1$

$\phi(\vec{x}^{(4)}) = (1, 0, \sqrt{2}, 0, 0, 1)\ , 1$

→ vidimo da će biti linearno odvojivo, postoji razlika po dimenzijama pa možemo zaključiti da ćemo ih odvojiti

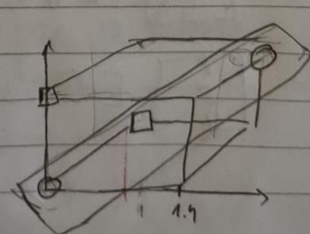
$K = (\vec{x}^T \vec{z})^2$ $\quad$ $\phi(x) = (\sqrt{2}\, x_1 x_2, x_1^2, x_2^2)$

$\phi(\vec{x}^{(1)}) = (0, 0, 0)\ , 0$

$\phi(\vec{x}^{(2)}) = (\sqrt{2}, 1, 1)\ , 0$

$\phi(\vec{x}^{(3)}) = (0, 1, 0)\ , 1$

$\phi(\vec{x}^{(4)}) = (0, 0, 1)\ , 1$



→ također postoji hiperravnina koja ovo može odvojiti

NEPARAMETARSKE METODE

NEPARAMETARSKE METODE [redacted]

$\boxed{2}$ Zadaci za učenje

a) $x^{(1)} = (4,2,1)$ $x^{(2)} = (0,3,3)$

	x	y	$d(x^{(i)}, x^{(1)})$	$d(x^{(i)}, x^{(2)})$
1	(4,4,0)	1	2.236 ✓	5.099 ✓
2	(4,3,1)	1	1 ✓	4.472 ✓
3	(6,0,2)	1	3	6.782
4	(5,2,2)	0	1.414 ✓	5.196 ✓
5	(5,1,1)	0	1.414 ✓	5.745 ✓
6	(7,2,0)	0	3.16	7.681

→ za prvi primjer (4,2,1) uzimam 4 najbliža primjera iz skupa D, a to su: 1., 2., 4., 5. primjer. Imamo po 2 glasa za svaku klasu, pa recimo da ćemo odabrat klasu 0.

→ za drugi primjer (0,3,3), također je izjednačen broj glasova za obje klase pa ćemo odabrat ponovno klasu 0.

b) $k = \frac{1}{1 + \|\vec{x} - \vec{x}'\|^2}$

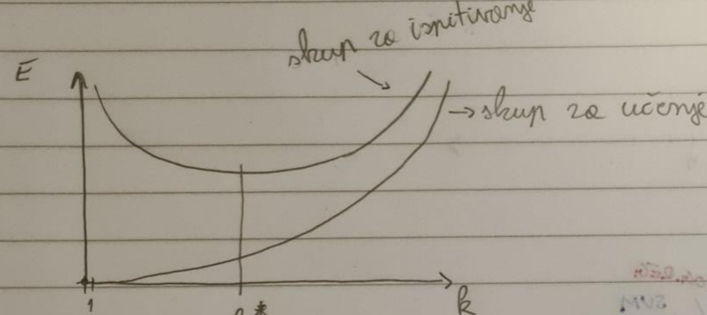
	x	y	$d(x^{(i)}, x^{(1)})$	$d(x^{(i)}, x^{(2)})$
1	(4,4,0)	1	0.7667 { 0.1667	0.037
2	(4,3,1)	1	0.5	0.1058 { 0.0476
3	(6,0,2)	1	0.1	0.02127
4	(5,2,2)	0	0.333	0.0357
5	(5,1,1)	0	0.751 { 0.333	0.08177 { 0.0294
6	(7,2,0)	0	0.091	0.01667

[redacted]

[redacted]

→ s obzirom na težinski k-NN, nakon što se zbrajaju težine, vidimo da bi se oba primjera klasificirala u klasu s oznakom 1, jer je suma težina veća.

c)



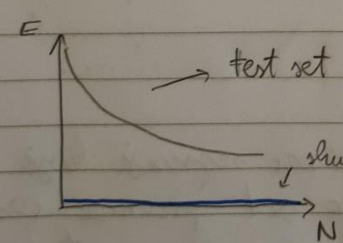
d)

→ za neki N, $k=1$, $k=3$

↓

na x-os kako se mijenja broj primjera za učenje

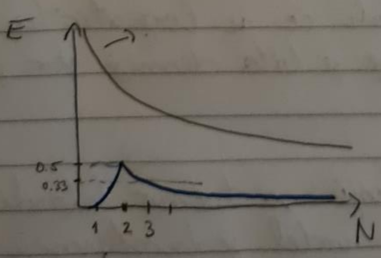
$k=1$



→ test set → pada s više primjera

skup za učenje → bit će uvijek 0, neovisno o broju primjera

$k=3$



→ ako je 2 primjera, pogreška je max 0.5, ako su u različitim klasama