Mia Krstičević 0036540586

VO8

1.1. a) Algoritam SVM se temelji na ideji maksimalne margine. Cilj je postaviti hiperravninu tako da bude najviše udaljena od najbližih primjera (margina) iz dviju klasa.

Znamo da je granica između klasa $h(x)=0$.

Model SVM-a: $h(x;w) = w^T x$

$\Rightarrow$ također, pretpostavljamo da su primjeri linearno odvojivi

$$\forall (x^{(i)}, y^{(i)}) \in D \quad y^{(i)} h(x^{(i)}) \geq 0$$

Ovo će uvijek vrijediti ako $h(x^{(i)}) = y^{(i)}$, tj. da su primjeri ispravno klasificirani.

1. potrebno je definirati najbliži primjer

$$d = \frac{h(x)}{\|w\|} \rightarrow \min_i \left\{ \frac{y^{(i)}(w^T x^{(i)} + w_0)}{\|w\|} \right\} = \frac{1}{\|w\|} \min_i \left\{ y^{(i)}(w^T x^{(i)} + w_0) \right\}$$

2. cilj je maksimizirati marginu

$$\underset{w, w_0}{\operatorname{argmax}} \left\{ \frac{1}{\|w\|} \min_i \underbrace{\left\{ y^{(i)}(w^T x^{(i)} + w_0) \right\}}_{1} \right\} = \underset{w, w_0}{\operatorname{argmax}} \frac{1}{\|w\|}$$

Poteškoća slijedi u tome što imamo min unutar argmax izraza. Pretpostavimo da je za primjer koji je najbliži margini, izlaz jednak nekoj konstanti, a s obzirom da vektor možemo skalirati vrijedi: $y^{(i)}(w^T x^{(i)} + w_0) = 1$ tj.

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

Za sve ispravno klasificirane primjere, ne samo one na margini!

Da izraz bude u standardnom obliku slijedi

$$\operatorname{argmin} \frac{1}{2} \|w\|^2$$ $\Rightarrow$ stavili smo kvadrat jer je L2-norma konveksna i neneg., a to nama treba za ovaj problem

uz uvjet: $y^{(i)}(w^T x^{(i)} + w_0) \geq 1 \quad i = 1, \dots, N$

Mia Krstičević 0036540586

b) Kvadratni program (QP) je poseban slučaj konveksnog optim. problema ($f(x)$ je konveksna) kod kojeg je ciljna f-ja kvadratna, a ograničenja su afine funkcije. Takav je problem maksimalne margine.

Jedna od metoda rješavanja QP-a je metoda Lagrangeove dualnosti (u kombinaciji s SMO). Cilj je da ograničenja eksplicitno ugradimo u ciljnu funkciju.

$$\mathcal{L}(x, \alpha, \beta) = f(x) + \sum_{i=1}^{m} \alpha_i g_i(x) + \sum_{i=1}^{p} \beta_i h_i(x) \qquad \alpha_i \geq 0$$

α_i, β_i - Lagr. multiplikator 15:15

Rješenje optim. problema s ograničenjem je stacionarna točka (gradijent Lagr. f-je) $\nabla \mathcal{L} = 0$, tu vrijede KKT uvjeti:

$$g_i(x) \leq 0 \quad i = 1, \ldots, m$$
$$h_i(x) = 0 \quad i = 1, \ldots, p$$
$$\alpha_i \geq 0 \quad i = 1, \ldots, m$$
$$\alpha_i g_i(x) = 0 \quad i = 1, \ldots, m$$

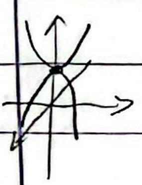
$h_i(x)$ je ograničenje jednakosti. $h(x)=0$ naše rješenje mora zadovoljiti. $h(x) = 0$ je n-dimen. površina, i svaka točka koja se nalazi na toj površ. je jedino moguća.

Pritom $g_i(x)$ i α_i su ograničenja nejednakosti. Tu postoje 2 slučaja: minimum f-je se nalazi unutar ostvarivog područja ili izvan.

$$\mathcal{L}(x, \alpha) = f(x) + \alpha g(x)$$ $\hookrightarrow g(x) = 0$, $\alpha = 0$

Za minimum mora vrijediti ili $g(x) = 0$ ili $\alpha = 0$, zato vrijedi $\alpha_i g_i(x) = 0$. → kompl. labavost

$$\tilde{\mathcal{L}}(\alpha, \beta) = \min_{x} \mathcal{L}(x, \alpha, \beta) \rightarrow$$ dualna Lagrangeova funkcija



Dualni problem je nalaženje donje ograde primarnog problema.

Ilia Krstičević 0036540586

c) Načelo dualnosti nam govori da optim. problemu možemo pristupiti na 2 načina: primarni i dualni problem.

$$\mathcal{L}(w, w_0, \alpha) = \frac{1}{2}\|w\|^2 - \sum \alpha_i \{ y^{(i)} (w^T x^{(i)} + w_0) - 1 \}$$

Kako u primarnom problemu minimiziramo $f(x)$, za doći u točku sedla u dualnom problemu moramo maksimizirati dualnu funkciju.

Prema tome,

minim. po primarnim varijablama

$$\tilde{\mathcal{L}}(\alpha) = \min_{w, w_0} \mathcal{L}(w, w_0, \alpha)$$

$$\frac{\partial \mathcal{L}(w, w_0, \alpha)}{\partial w} = 0 \Rightarrow w = \sum \alpha_i y^{(i)} x^{(i)}$$

$$\frac{\partial \mathcal{L}(w, w_0, \alpha)}{\partial w_0} = 0 \Rightarrow \underbrace{\sum \alpha_i y^{(i)}}_{=0} = 0$$

dualna Lagr. funkcija

$$\tilde{\mathcal{L}}(\alpha) = \min_{w, w_0} \left(\frac{1}{2}\|w\|^2 - \sum \alpha_i y^{(i)} w^T x^{(i)} - w_0 \sum \alpha_i y^{(i)} + \sum \alpha_i \right)$$

$$\rightarrow = \sum^N \alpha_i - \frac{1}{2} \sum^N \sum^N \alpha_i \alpha_j y^{(i)} y^{(j)} (x^{(i)})^T x^{(j)}$$

Želimo maksim.

$$\sum^N \alpha_i - \frac{1}{2} \sum^N \sum^N \alpha_i \alpha_j y^{(i)} y^{(j)} (x^{(i)})^T x^{(j)}$$

uz ogran.

$$\alpha_i \geq 0 \qquad \sum \alpha_i y^{(i)} = 0$$

KKT uvjeti:

uvjeti proizlaze iz uvjeta kod primarnog problema

$$y^{(i)} (w^T x^{(i)} + w_0) \geq 1 \qquad \alpha_i \geq 0$$

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

d) U dualnom obliku nemamo više težine, već dualne param. α. $w = \sum^N \alpha_i y^{(i)} x^{(i)}$. Također su rješenja jednostavnija (račun ski).

↳ pohranjujemo primjere i njihove oznake

e)

$$h(\vec{x}) = \underbrace{w^T x + w_0}_{\text{primarno}} = \underbrace{\sum^N \alpha_i y^i x^T (x^i) + w_0}_{\text{dualno}}$$

Mia Krstičević 0036540586

f) Potporni vektori su oni koji točno leže na margini.

To znamo iz uvjeta kompl labavosti:

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

Tu vrijedi da je $\alpha_i = 0$ ili $y^{(i)} h(x^{(i)}) = 1$. Ako je $\alpha_i = 0$ taj vektor ne pridonosi izračunu funkcije h, a ako vrijedi $y^{(i)} h(x^{(i)}) = 1$, primjer točno leži na margini.

g) Skaliranje je potrebno prvenstveno radi ravnoteže među značajkama (proporcionalnost) i preciznu udaljenost u prostoru značajki. Nužna je za ispravnu optimizaciju.

Mia Krstičević 0036540586

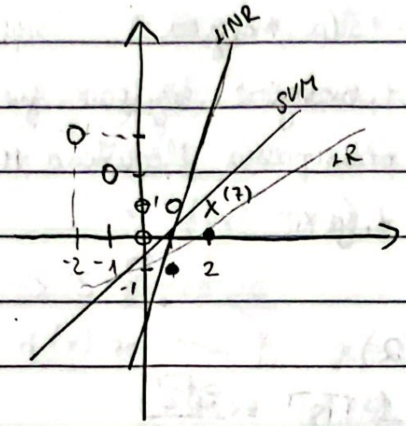
2.4. $D = \{((-2,3),0), ((-1,2),0), ((0,1),0), ((0,0),0), ((1,1),0), ((1,-1),1), ((2,0),1)\}$

log LR, SVM   LINR $\to h(x) = 1\{w^T x \geq 0\}$

za SVM i LINR $y=-1$ umjesto $y=0$

LR $E(w|D) \approx 0$

$x^{(7)} = (2,0)$ $d(m) = ?$ udaljenost $x^{(7)}$ od granice između klasa



LINR želi minimiz. kvadr. pogrešku - smještamo između $(1,1)$ i $(1,-1)$. + vidimo iz f-je $h(x)$

SVM stavljamo točno na sredinu između 2 klasa.

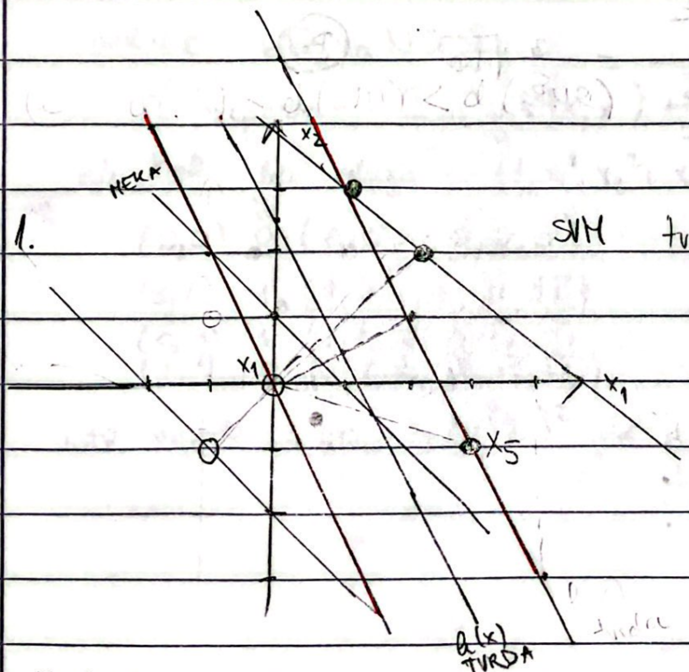
Ako je za LR $E(w|D) \approx 0$ znači da je najkraća udaljenost. (najmanja pogreška)

LINR - MSE $\frac{1}{2}\Sigma(y-h(x))^2$

(c) $d(LR) < d(SVM) < d(LINR)$

VOG

2. 1.



SVM tvrda i meka margina $C=1$

$\alpha_1 = 1$ $\alpha_2, \alpha_3, \alpha_4 > 0$

$\alpha_5 = 1$

$\Rightarrow \alpha_1$ i α_5 unutar margina

sl. 2 stranica $\to$

Za tvrdu marginu niti jedan primjer ne smije biti unutar margine!

$y^{(i)} h(x^{(i)}) \geq 1$

1. granica je za $w^T x + b = 0$

sustav $w_1 x_1 + w_2 x_2 + b = y$

nagib: $\frac{-1-3}{3-1} = -2$

$\begin{cases} y_{t1} = -2x + 5 \\ y_{t2} = -2x \end{cases}$

tvrda $y = -2x + 2.5$

Mia Krstičević 0036540586

tvrda: širina $\frac{1}{2}\sqrt{2^2+1^2} = \frac{\sqrt{5}}{2}$

meka: C je hiperparametar, određuje koliko će kažnjavati pogrešne klasifikacije. $0 \leq \alpha_i \leq C = 1$ $y_i(wx_i + b) \geq 1 - \xi_i$

što manji C, manja kazna - šira margina

Ako su $\alpha_i = 1$ znači da su unutar margine. sigurno: jer je to maksimalna dopuštena vr. C=1.
↳ točke potpore

$\alpha_i = C - \beta_i$ $w = \sum_i^n \alpha_i y_i x_i$

$\beta_i = 0$

širina: od (-1,-1) do (2,2)

$\frac{1}{2}\sqrt{(2-(-1))^2+(2-(-1))^2} = \frac{1}{2}\sqrt{18} = \frac{3\sqrt{2}}{2}$

$\frac{\text{mekana}}{\text{tvrda}} = \frac{\frac{3\sqrt{2}}{2}}{\frac{\sqrt{5}}{2}} = \frac{3}{5}\sqrt{10}$ B

Mia Krstičević 0036540586

VIO

1.3 a) $X = \mathbb{R}^2$ $K(x,z) = (x^T z + 1)^2$ $n=2$ Mercerova j.

Da bi jezgreni trik funkcionirao, jezgrena f-ja mora biti Mercerova, odnosno mora biti pozitivno semidefinitna.

$$K(x,z) = (x^T z + 1)^2 = ((x_1,x_2)^T(z_1,z_2)+1)^2 = (x_1 z_1 + x_2 z_2 + 1)^2$$
$$= (x_1 z_1 + x_2 z_2)^2 + 2x_1 z_1 + 2x_2 z_2 + 1$$
$$= x_1^2 z_1^2 + 2x_1 z_1 x_2 z_2 + 2x_1 z_1 + 2x_2 z_2 + 1 + x_2^2 z_2^2$$
$$= (1, x_1\sqrt{2}, x_2\sqrt{2}, x_1 x_2\sqrt{2}, x_1^2, x_2^2)^T (1, z_1\sqrt{2}, z_2\sqrt{2}, z_1 z_2\sqrt{2}, z_1^2, z_2^2)$$
$$= \phi(x)^T \phi(z)$$

-> 5-dimenzija

b) $\phi(x)$ za $n=2$ $x = (3,2)$

$$\phi(x) = (1, 4, 9, 6\sqrt{2}, 2\sqrt{2}, 3\sqrt{2})$$

c) Preslikali smo s $n=2$ u dim. 5. tako da će broj neparametarske inačice biti za $N > 5$.

neparam. $\sum \alpha_i y^i K(x, x^i) + w_0$

d) Je li XOR lin. odvojiv? $(1, x_1^2, x_2^2, x_1 x_2\sqrt{2}, x_1\sqrt{2}, x_2\sqrt{2})$

	y	
$(0,0)$	0	$(1,0,0,0,0,0)$
$(0,1)$	1	$(1,0,1,0,0,\sqrt{2})$
$(1,0)$	1	$(1,1,0,0,\sqrt{2},0)$
$(1,1)$	0	$(1,1,\sqrt{2},\sqrt{2},\sqrt{2},\sqrt{2})$

-> XOR je linearno odvojiv

isto vrijedi i za $K(x^T z)^2$ jer je ona Mercerova jezgra polinomijalna.

$(x^T x' + 1)^P$

Mia Krstičević 0036540586

VII

1.2. a) $\{((4,4,0),1), ((4,3,1),1), ((6,0,2),1), ((5,2,2),0), ((5,1,1),0), ((7,2,0),0)\}$

$x^{(1)} = (4,2,1) \quad x^{(2)} = (0,3,3)$ 4-NN

pr. $\sqrt{(4-4)^2+(4-2)^2+(0-1)^2}$

$\|x-x^1\|$	$\|x-x^2\|$
$\sqrt{5} \approx 2.24$	$\sqrt{26} \approx 5.1$
1	$2\sqrt{5} \approx 4.47$
3	$\sqrt{46} \approx 6.78$
$\sqrt{2}$	$3\sqrt{3} \approx 5.2$
$\sqrt{2}$	$\sqrt{33} \approx 5.74$
$\sqrt{10}$	$\sqrt{59} \approx 7.7$

Za $x^{(1)}$ najbliži je $\overset{1}{(4,3,1)}, \overset{1}{(4,4,0)}, \overset{0}{(5,2,2)}, \overset{0}{(5,1,1)}$

Za $x^{(2)}$ najbliži je $(4,3,1)(4,4,0)$, -//-

Nema odluke jer argmax $(2,2)$ za oba

b) 4-NN inverzna kv jezgra

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

$\|x-x^1\|$	$\|x-x^2\|$
$\frac{1}{6}$	$\frac{1}{27}$
$\frac{1}{2}$	$\frac{1}{21}$
$\frac{1}{10}$	$\frac{1}{47}$
$\frac{1}{3}$	$\frac{1}{28}$
$\frac{1}{3}$	$\frac{1}{34}$
$\frac{1}{11}$	$\frac{1}{60}$

težinski model:

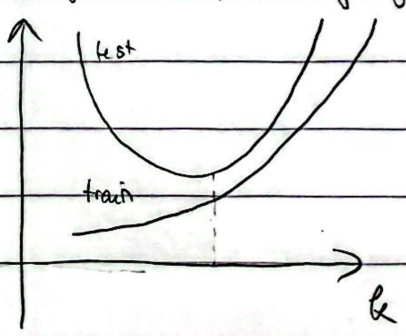
$$h(x) = \text{argmax} \sum k(x,x^i) \cdot 1\{y^i = j\}$$

$$h_1 = \text{argmax}\left(\frac{1}{6}+\frac{1}{2}, \frac{1}{3}+\frac{1}{3}\right) = \text{argmax}\left(\frac{2}{3}, \frac{2}{3}\right) \times$$

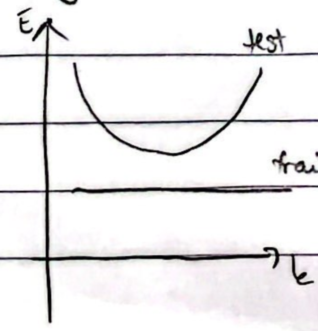
$$h_2 = \text{argmax}\left(\frac{1}{27}+\frac{1}{21}, \frac{1}{28}+\frac{1}{34}\right) = \text{argmax}(0.085, 0.065) = 1$$

↳ klasa 1

c) što je veći k, model je jednostavniji



d) $k=1$



$k=3$