1. $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^{6}$
$= \{((-3,1),0), ((-3,3),0), ((1,2),1), ((2,1),1), ((1,-2),2), ((2,-3)$

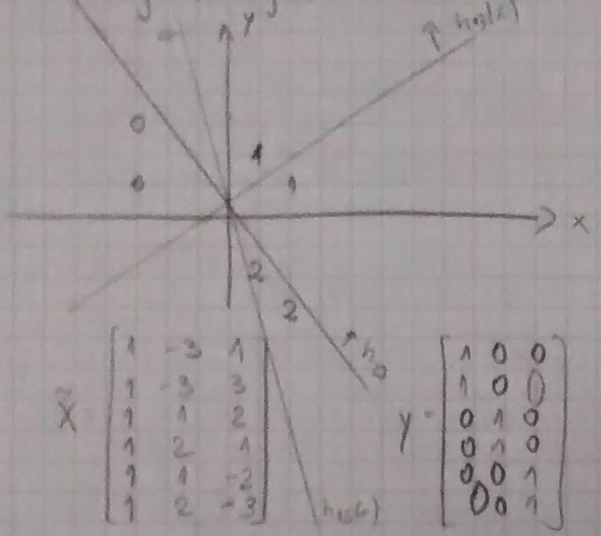
a) jedan-naspram ostalih $\Rightarrow$ $x \in \mathcal{G}_j$ ako $h_j(\tilde{x}) \geq 0$

$h(x) = \underset{j}{\mathrm{argmax}}\, h_j(x)$

$h_j(x) = ?$

$h_j(x) = \tilde{w}_j^T \tilde{x}$

*



$$\tilde{X} = \begin{bmatrix} 1 & -3 & 1 \\ 1 & -3 & 3 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \\ 1 & 1 & -2 \\ 1 & 2 & -3 \end{bmatrix} \qquad y = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\tilde{W} = \tilde{X}^{+} y = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T y = \begin{bmatrix} 0.335 & 0.259 & 0.406 \\ -0.217 & 0.234 & -0.017 \\ -0.005 & 0.222 & -0.217 \end{bmatrix}$$

$h_1(x) = -0.217x - 0.005y + 0.335$

$h_2(x) = 0.234x + 0.222y + 0.259$

$h_3(x) = -0.017x - 0.217y + 0.406$

b) $h_{12}(x) = h_1(x) - h_2(x) = -0.451x - 0.227y + 0.076$

$h_{13}(x) = h_1(x) - h_3(x) = -0.2x + 0.212y - 0.147$

$h_{23}(x) = h_2(x) - h_3(x) = 0.251x + 0.439y - 0.071$

$y_{12} = \dfrac{-451x_{12} - 76}{227}$ $\qquad y_{23} = -0.572x_{23} + 0.162$

$y_{13} = 0.943x + 0.694$

c) $x = (-1,3)$

$h_1(\tilde{x}) = 0.537$ $\qquad \boxed{h_2(\tilde{x}) = 0.691 \qquad h_3(\tilde{x}) = -0.235}$

$h(x) = \mathrm{argmax}\, h_j(x)$

d) Nema vjerojatnosti pripadanja klasi.

- reg. ne daje na izl. vrijed. između 0 i 1

(2.) a) Funkcija $f: \mathbb{R}^n \to \mathbb{R}$ je konveksna akko

1) njezina domena $dom(f)$ je konveksni skup
2) za svaki $x_1, x_2 \in dom(f)$ i svaki $\alpha \in [0, 1]$ vrijedi

$$f(x) = f(\alpha x_1 + (1-\alpha) x_2) \leq \alpha f(x_1) + (1-\alpha) f($$

b) F-ja f je kvazikonveksna (ili unimodalna) akko je njezina domena $dom\, f$ konveksna, te ako za svaki $x, y \in$
, $0 \leq \theta \leq 1$ vrijedi

$$f(\theta x + (1-\theta) y) \leq \max\{f(x), f(y)\}$$

Kvazikonveksnost je poopćenje konveksnosti, svaka je konveksna f-ja unimodalna, ali obrat ne vrijedi.

...

pokazati

ima

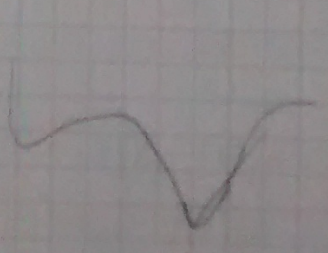
jedan

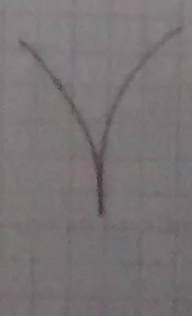
minimum

n

nije

konveks





) $E_p(\tilde{w}|D)$ = kriterij perceptrona

misclassification rate $E_m(\tilde{w}|D)$

a) $E_m(\tilde{w}|D)$ nije derivabilna pa ju ne možemo izravno minimizirati. // g.s ne može => po dijelovima konst!

b) $E_p(\tilde{w}|D) = \sum_{i=1}^{n} \max(0, -\tilde{w}^T \phi(x^{(i)}) y^{(i)})$

↓

to je suma linearnih f-ja (lin. f-je su konveksne)

=> max konv. f-ja je konveksan

=> suma konveksnih f-ja je konveksna => $E_p(\tilde{w}|D)$ je konv

Konveksnost f-je je bitna jer nam treba 1 minimum za grad. spust, ukoliko ne bi bila konveksna postojalo bi više lokalnih minimuma i grad. spust bi mogao naći lokalni umjesto globalni minimum

c) $L(h(x), y)$ za perceptron

za $x^{(1)} = (1, 0, 0)$ $L(h(x^{(1)}), y^{(1)}) = 0.0$

Klasifikacija za ovaj primjer nije ispravna jer ispravno klasificirani primjeri ne pridonose f-iji pogreške, tj. njihov gubitak je 0

$x^{(1)} = (1, 1, 1)$

ako je ispravno klasif. onda je $L = 0$

Postupak perceptrona ne konvergira ako primjeri nisu linearno odvojivi.

→ Postupak se ponavlja sve dok su primjeri ispravno klasificirani, ako primjeri nisu lin. odvojivi nije moguće ispravno klasif. sve primjere pa se postupak nikada neće zaustaviti.

) Rezultati kod postupka perceptrona ovise o početnim težinama i redoslijedu predočavanja primjera

Zašto? → postoji šire područje gdje je f-ja pogreške

→ algoritam će se zaustaviti na drugom mjestu?

→ pa ovisno otkuda se usele dobiva se drugačije rješenje

Grupno? pa opet je problem o poč. težinama

4) a) LR $n=100$ značajki

broj parametra LR: $n+1 = 101$ parametar

broj parametra BK: $\frac{n}{2}(n+1) + 2n + 1 = 5251$ parametar

b) $\Rightarrow k=5$

$kn + k - 1 = 500 + 499 = 999$

$\frac{kn}{2}(kn + k - 1) + 2kn + k - 1 = \frac{500}{2}(500 + 499) + 2 \cdot 500 + 499$

$= 251249$

c) $k = 10$

$n = ?$ da se isplati diskriminativni model

ima uvijek manje parametara?

5) samo za 1 (5)

rezultat

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

$\eta = 0{,}005$

$\phi(x) = \tilde{x}$

$\tilde{w}^{(0)} = (0,0,0)$

Do li konvergira?

→ D je linearno odvojiv pa će konvergirati

1) $f(\tilde{w}^T \phi(x^{(1)})) = f\left([0\ 0\ 0]\begin{bmatrix}-3\\1\\1\end{bmatrix}\right) = f(0) = 1$

$\eta\, \phi(x^{(1)})\, y^{(1)} = [-0{,}015,\ 0{,}005,\ 0{,}005]$

$\tilde{w}^{(1)} \leftarrow [0\ 0\ 0] - [-0{,}015\ \ 0{,}005\ \ 0{,}005]$

$\tilde{w}^{(1)} = [0{,}015\ \ -0{,}005\ \ -0{,}005]$

2) $f(\tilde{w}^{T(1)} \phi(x^{(2)})) = f\left([0{,}015,\ -0{,}005,\ -0{,}005]\begin{bmatrix}-1\\3\\1\end{bmatrix}\right) =$
$= f(-0{,}035) = -1$ ✓

$\tilde{w}^{(2)} = \tilde{w}^{(1)}$

3) $f(\tilde{w}^{T(2)} \phi(x^{(3)})) = f\left([0{,}015,\ -0{,}005,\ -0{,}005]\begin{bmatrix}1\\2\\1\end{bmatrix}\right) =$
$= f(0) = 1$ ✓

$\tilde{w}^{(3)} = \tilde{w}^{(2)}$

4) $f(\tilde{w}^{T(3)} \phi(x^{(4)})) = f\left([0{,}015\ \ -0{,}005\ \ -0{,}005]\begin{bmatrix}2\\1\\1\end{bmatrix}\right) =$
$= f(0{,}02) = 1$ ✓

(8) a) Logistička reg. izravno modelira aposteriornu $P(C_1|x)$

$$h(x) = P(C_1|x)$$

$$P(C_1|x) = \frac{p(x|C_1) \cdot P(C_1)}{P(x)}$$

U slučaju 2 klase C_1 i C_2:

$$P(C_2|x) = 1 - P(C_1|x)$$

$$P(C_1|x) = \frac{P(C_1) \cdot p(x|C_1)}{P(C_1) \cdot p(x|C_1) + P(C_2) \cdot p(x|C_2)}$$

$$P(C_1|x) = \frac{1}{1 + \frac{P(C_2) \cdot p(x|C_2)}{P(C_1) \cdot p(x|C_1)}}$$

$$\alpha = \ln \frac{P(C_1) \cdot p(x|C_1)}{P(C_2) \cdot p(x|C_2)}$$

$$P(C_1|x) = \frac{1}{1 + e^{-\alpha}} = \sigma(\alpha)$$

Pretpostavimo da se $p(x|C_1)$ i $p(x|C_2)$ pokoravaju G. r. s dijeljenom kov. mat., α možemo izraziti kao lin. komb. tj.

$$\alpha = \ln \frac{P(C_1) \cdot p(x|C_1)}{P(C_2) \cdot p(x|C_2)} = w^T x + w_0 = \tilde{w}^T \tilde{x}$$

$$\alpha = \ln P(C_1) + \ln p(x|C_1) - \ln P(C_2) - \ln p(x|C_2)$$

$$= x^T \Sigma^{-1} \mu_1 - \frac{1}{2} \mu_1^T \Sigma^{-1} \mu_1 - x^T \Sigma^{-1} \mu_2 + \frac{1}{2} \mu_2^T \Sigma^{-1} \mu_2 + \ln P(C_1) - \ln P(C_2)$$

$$= \underbrace{x^T \Sigma^{-1} (\mu_1 - \mu_2)}_{w} \underbrace{- \frac{1}{2} \mu_1^T \Sigma^{-1} \mu_1 + \frac{1}{2} \mu_2^T \Sigma^{-1} \mu_2 + \ln \frac{P(C_1)}{P(C_2)}}_{w_0}$$

Ispada da vrijedi:

$$h(x) = P(C_1|x) = \sigma(\alpha) = \sigma(\tilde{w}^T \tilde{x})$$

b) $E_x(\tilde{w}|D) = -\ln \mathcal{L}(\tilde{w}|D)$

$D = \{x^i, y^i\}_{i=1}^{N}$

za $y^i = 1$ x je u klasi C_1

za primjer x, y nam je slučajna var. Bernoullijeva razdioba

Po vjerojatnosti $P(y=1|x) = P(C_1|x) = h(x)$

$P(y|x) = h(x)^y (1-h(x))^{1-y}$

Izravno se modelira $P(y|x)$, izglednost se izražava u odnosu na y

$\mathcal{L}(\tilde{w}|D) = \ln P(D|\tilde{w}) = \ln \prod_{i=1}^{N} P(y^i|x^i) = \ln \prod_{i=1}^{N} h(x^i)^{y^i} (1-h(x^i))^{1-y^i}$

$E(\tilde{w}|D) = -\sum_{i=1}^{N} \ln h(x^i)^{y^i} (1-h(x^i))^{1-y^i}$

$E_x(\tilde{w}|D) = -\sum_{i=1}^{N} \{y^i \ln h(x^i) + (1-y^i) \ln(1-h(x^i))\}$

c) $\nabla E_x(\tilde{w}|D) = \sum_{i=1}^{N} (h(x^i) - y^i)\tilde{x}$

d) Regul. pogreška odgovara MAP procjeni na skup. D

e) $y \in \{-1, +1\}$ umjesto $y = \{0, 1\}$

pogrešne klasif. $-1 \Rightarrow 1$

$1 \Rightarrow -1$

$L(h(x), y) = -(-1-y)\ln(-1-h(x)) - (1-y)\ln(1-h(x))$

ne znam

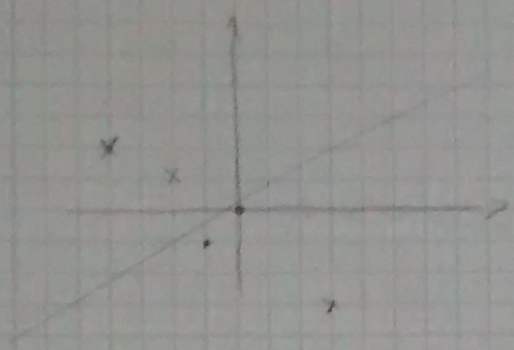
jel ok?

ima na slajdu?

① $\phi: \mathbb{R}^n \to \mathbb{R}^m$

$D = \{(x^i, y^i)\}_{i=1}^{5} = \{((-1,1),0), ((0,3),0), ((3,-3),1), ((-3,-1),1), ((-4,2), \ldots$

a)



$P(C_1) = \frac{2}{5}$

$P(C_2) = \frac{3}{5}$

$\mu_1 = -\frac{1}{5} = -0{,}2$

$\mu_2 = -0{,}2$

$h_{12}(x) = h_1(x) - h_2(x) = \ln \frac{P(x|C_1) \cdot P(C_1)}{P(x|C_2) \cdot P(C_2)}$

$w_1 = \Sigma^{-1}(\mu_1 - \mu_2) =$

$w_0 = -\frac{1}{2}\mu_1^T \Sigma^{-1} \mu_1 + \frac{1}{2}\mu_2^T \Sigma^{-1} \mu_2 + \ln \frac{P(C_1)}{P(C_2)}$

b) $\phi(x) = (1, \phi_1(x), \phi_2(x))$

$\phi_j = \exp\{-\frac{\|x - \mu_j\|^2}{2\sigma_j^2}\}$

$\mu_1 = (0,0)$

$\mu_2 = (-3,3)$

$\sigma_1 = \sigma_2 = 1$

$(-1,1) \Rightarrow (1, e^{-\frac{2}{2}}, e^{\ldots})$

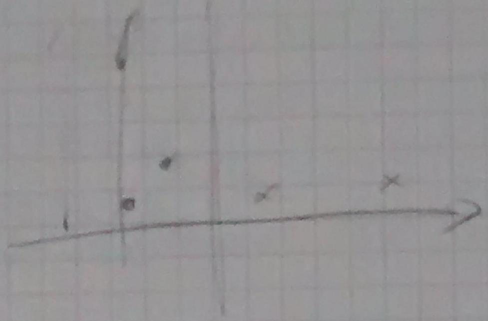
$\sqrt{(-1)^2 + (1)^2} = \sqrt{2}$, $\sqrt{(-1+3)^2 + (1-3)^2} = \sqrt{4} = 2$

$(0,0) \Rightarrow (1, 1, e^{-\sqrt{18}})$

$(3,-3) \Rightarrow (1, e^{-\sqrt{18}}, 1)$

$(-2, 1) \Rightarrow (1, e^{-13}, e^{-13})$
$(-4, 2) \Rightarrow (1, e^{-20}, e^{-13})$

c)



d) parametara 3 ← vektor w
hiperparametara ← baze {μ 6

u praksi vrijednost hiperparametara?
cross-validation

utječu li hiperpar na složenost modela? NE
↳ utjecalo bi broj parametara

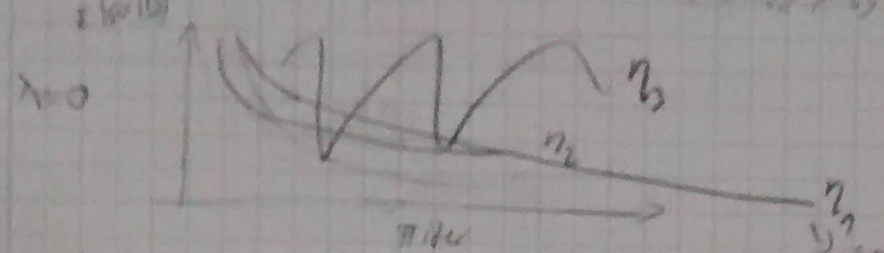
8

a) $E(w|D) = ?$

$\lambda = 0$

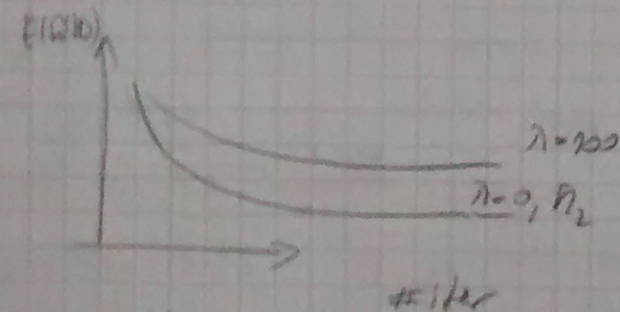
η_1, η_2, η_3

$\eta_1 \approx \eta_2 > \eta_3$



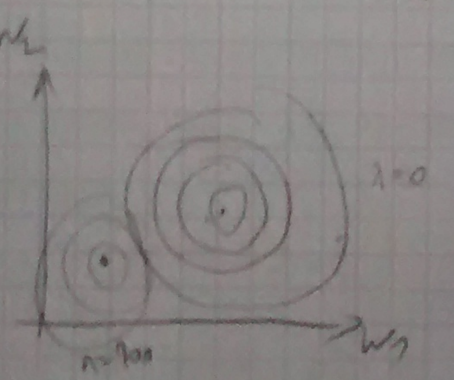
η1 sporije pada ali se bolje prilagodi

b)

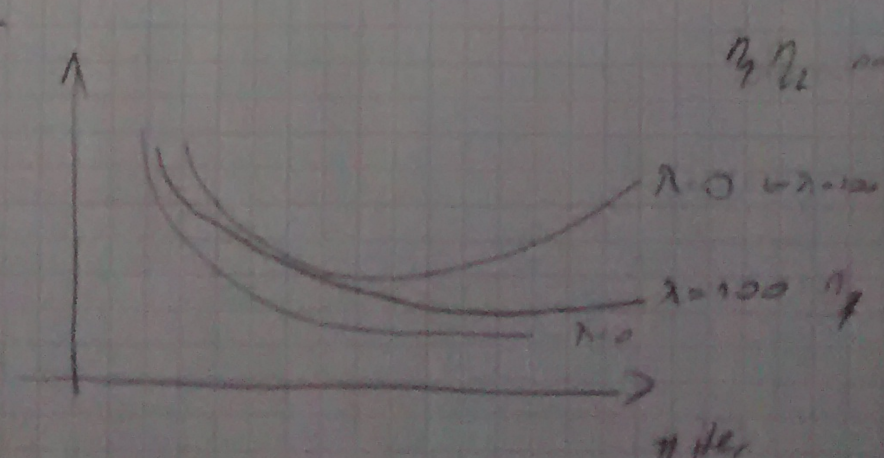


uzmemo η_2

c)

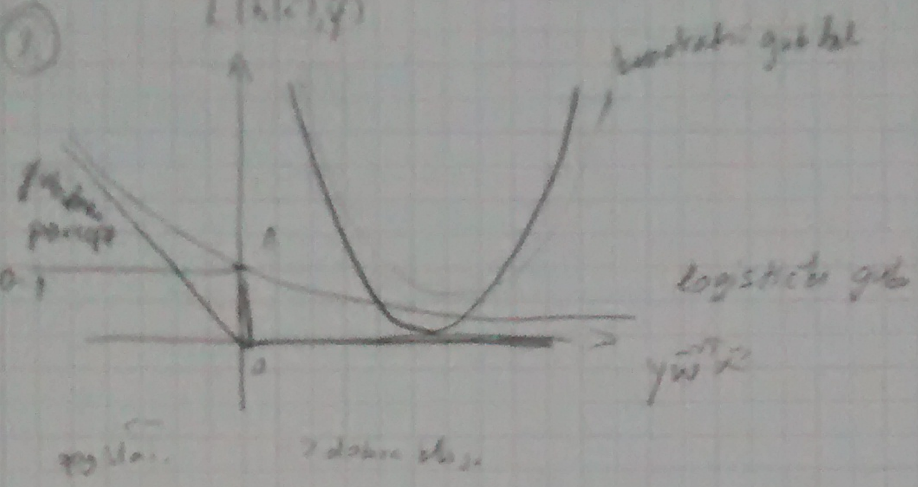


d)



$\eta_1 \eta_2$

2



a) kažnjava dobro klasf. primjere (tj. ispravne primjere, ali daleko od granice) → nije prikladna kada želimo min broj pog klasif

b) Iako log. reg. također kažnjava ispravne f-je, kvadratni gubitak raste, a log. gubitak se približava nuli

c) Iz slike se vidi kako za ispravno klas. primjere f-ja g. teži prema nuli, pa se stoga čini w prilagođavaju podacima.

d) Logističi i kvadratni gubitak će biti veći od pog (!) klasif. primjera jer se kažnjavaju i ispravni primjeri