$\boxed{VO5}$

Emmanuel Njegovec
0036543588

Zad. 1.2)

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

a) postupak OVR

$$\Phi = \begin{bmatrix} 1 & -3 & 1 \\ 1 & -3 & 3 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \\ 1 & 1 & -2 \\ 1 & 2 & -3 \end{bmatrix} \qquad y_0 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad y_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad y_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

$$\vec{w_j} = \Phi^+ y_j$$

$$\Phi^+ = \begin{bmatrix} 0.186 & 0.159 & 0.126 & 0.134 & 0.199 & 0.207 \\ -0.14 & -0.077 & 0.106 & 0.128 & -0.019 & 0.003 \\ -0.058 & 0.052 & 0.123 & 0.099 & -0.097 & -0.12 \end{bmatrix}$$

$$\vec{w_0} = \Phi^+ y_0 = \begin{bmatrix} 0.335 \\ -0.217 \\ -0.006 \end{bmatrix} \qquad \vec{w_1} = \Phi^+ y_1 = \begin{bmatrix} 0.259 \\ 0.235 \\ 0.222 \end{bmatrix} \qquad \vec{w_2} = \Phi^+ y_2 = \begin{bmatrix} 0.407 \\ -0.017 \\ -0.218 \end{bmatrix}$$

$$h_0 = 0.335 - 0.217x_1 - 0.006x_2$$

$$h_1 = 0.259 + 0.235x_1 + 0.222x_2$$

$$h_2 = 0.407 - 0.017x_1 - 0.218x_2$$

b)

$$h_{ij} = \vec{w_0}_{ij} + \vec{w_1}_{ij} x_1 + \vec{w_2}_{ij} x_2$$

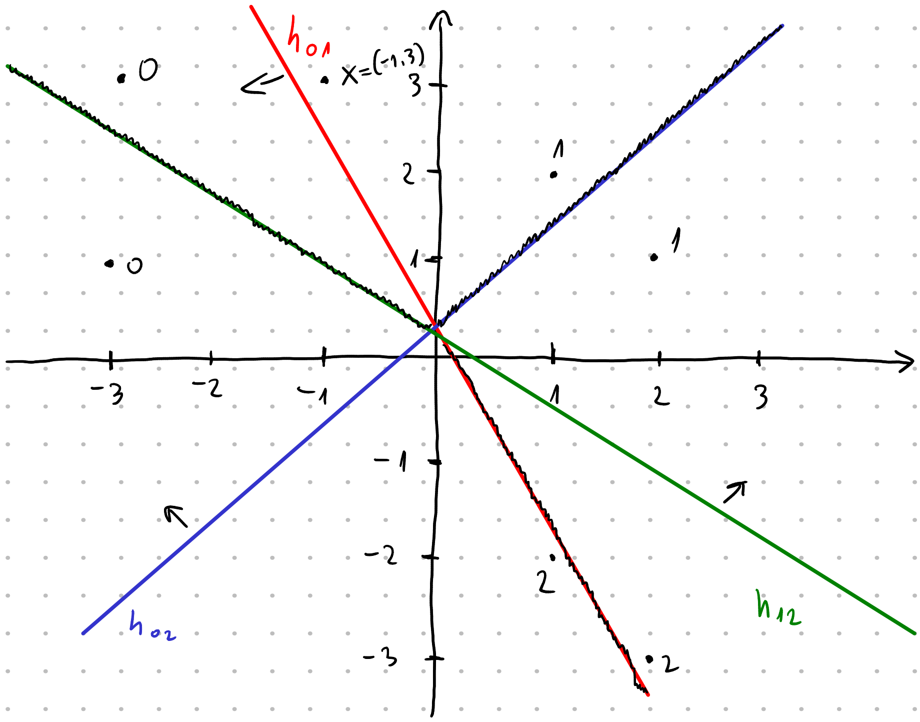
$$\vec{w_{01}} = \vec{w_0} - \vec{w_1} = \begin{bmatrix} 0.335 \\ -0.217 \\ -0.006 \end{bmatrix} - \begin{bmatrix} 0.259 \\ 0.235 \\ 0.222 \end{bmatrix} = \begin{bmatrix} 0.076 \\ -0.452 \\ -0.228 \end{bmatrix}$$

$$\vec{w_{02}} = \vec{w_0} - \vec{w_2} = \begin{bmatrix} -0.072 \\ -0.2 \\ 0.212 \end{bmatrix} \qquad \vec{w_{12}} = \vec{w_1} - \vec{w_2} = \begin{bmatrix} -0.148 \\ 0.252 \\ 0.44 \end{bmatrix}$$

$$h_{01} = 0.076 - 0.452x_1 - 0.228x_2$$

$$h_{02} = -0.072 - 0.2x_1 + 0.212x_2$$

$$h_{12} = -0.148 + 0.252x_1 + 0.44x_2$$



Emmanuel Njegovec
0036543588

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

- nalazi se između h_{01} i h_{12}
- ispod h_{01} $\Rightarrow$ klasa 1
- iznad h_{12}

d) Uporaba linearne regresije u svrhu klasifikacije nam ne daje informaciju o tome s kojom vjerojatnosti primjer pripada nekoj klasi jer linearna regresija ne daje probabilističku interpretaciju.

e) Prednost OVR-a nad OVO-m je manji broj modela koje treba trenirati, a nedostatak je taj da OVR lako dovodi do neuravnoteženih brojeva primjera između parova klasa za koje treniramo model.

f) Dva glavna problema korištenja linearne regresije za klasifikaciju su:

- nedostatak vjerojatnosne interpretacije
- nerobusan model koji je osjetljiv na vrijednosti odskačn, odnosno algoritam kažnjava točno klasificirane primjere koji su "duboko" u području neke klase.

Dodavanje primjera $((0,10),1)$ bi dovelo do pomicanja h_{12} na način da bi $((2,-3),2)$ bio klasificiran kao 1.

Zad. 2.3.)

Emanuel Njegovec

0036543588

$N=1000$

$n=555$

$K=4 \quad \{400, 300, 200, 100\}$

$\lambda=0$

$\Phi \Rightarrow$ matrica dizajna $\Rightarrow$ dimenzije $N \times (n+1) \Rightarrow$ uvjet stabilnosti: $N \geq n+1$

- binarna logistička regresija (OVO, OVR)

OVO

$h_{01}: K_0 + K_1 = 400+300 = \overset{\nearrow N}{700} \Rightarrow \dim \Phi = 700 \times 556$ ✓

$h_{02}: K_0 + K_2 = 400+200 = 600 \Rightarrow \dim \Phi = 600 \times 556$ ✓

$h_{03}: K_0 + K_3 = 500 \Rightarrow \dim \Phi = 500 \times 556$ ✗

$h_{12}: K_1 + K_2 = 500$ ✗

$h_{13}: K_1 + K_3 = 400$ ✗

$h_{23}: K_2 + K_3 = 300$ ✗

$\} \; 4 \quad$ C

OVR

$K_0 + K_1 + K_2 + K_3 = 1000 \Rightarrow \dim \Phi = 1000 \times 556$ ✓

$\boxed{\text{VOG}}$

Zad. 2.7.)

$\lambda = 1000$

$\eta = 0.01$

$\phi(x) = (1, x_1, x_2, x_1 x_2)$

$\vec{w} = (0.2, 0.5, -1.1, 2.7)$

$(\vec{x}, y) = ((-1, 2), 1)$

SGD

$\nabla_w L(y, h(\vec{x})) = (h(\vec{x}) - y)\, \Phi(\vec{x})$

$\Phi(\vec{x}) = (1, -1, 2, -2)$

$h(\vec{x}) = \sigma(\vec{w}^T \Phi(\vec{x})) = \sigma(-7.9) = \dfrac{1}{1+\exp(-(-7.9))} = 3.71 \cdot 10^{-4}$

$\nabla_w L(1, 3.71 \cdot 10^{-4}) = (3.71 \cdot 10^{-4} - 1) \cdot [1, -1, 2, -2]$

$= [-1, 1, -2, 2]$

$\vec{w} \leftarrow \vec{w}(1 - \eta\lambda) - \eta \nabla_w L$

$w_1 = w_1 (1 - 0.01 \cdot 1000) - 0.01 \cdot 1$

$w_1 = 0.5(1-10) - 0.01 = -4.51$

$\Delta w_1 = -4.51 - 0.5 = \underline{\underline{-5.01}}$ B

V07

Emanuel Njegovec
0036543588

Zad. 1.1.)

a) funkcija softmax : $\mathbb{R}^n \to \mathbb{R}^n$

$$\text{softmax}_k(x_1, x_2, \dots, x_n) = \frac{\exp(x_k)}{\sum_j \exp(x_j)} \to \exp(2)+\exp(8)+\exp(1)+\exp(5) = 3139.48$$

$\alpha = (2, 8, 1, 5)$

$$\text{softmax}_1(\alpha) = \frac{\exp(2)}{3139.48} = 2.3 \cdot 10^{-3}$$

$$\text{softmax}_2(\alpha) = \frac{\exp(8)}{3139.48} = 0.94$$

$$\text{softmax}_3(\alpha) = 8.6 \cdot 10^{-4}$$

$$\text{softmax}_4(\alpha) = 4.7 \cdot 10^{-2}$$

$$\Rightarrow \text{softmax}(\alpha) = [2.3 \cdot 10^{-3}, 0.94, 8.6 \cdot 10^{-4}, 4.7 \cdot 10^{-2}]$$

- dva efekta f-je:
  - normalizira sve vrijednosti t.d. u zbroju budu 1
  - pojačava veće vrijednosti i smanjuje manje

b) model multinomijalne logističke regresije:

- skup modela $\{h_k\}_k$, svaki model zadužen za k-tu klasu od K klasa

$$h_k(\vec{x}, \vec{W}) = \frac{\exp(\vec{w}_k^T \Phi(\vec{x}))}{\sum_j \exp(\vec{w}_j^T \Phi(\vec{x}))} = P(y = k \mid \vec{x}, \vec{W})$$

$\hookrightarrow \vec{W} = (\vec{w}_1, \vec{w}_2, \dots, \vec{w}_k)$

c)

oznake klasa : $y = (y_1, \dots, y_k)^T, \sum_k y_k = 1$

vektor parametara : $\mu = (\mu_1, \dots, \mu_k), \sum_k \mu_k = 1, \mu_k \geq 0$

distribucija kategoričke varijable : $P(y \mid \mu) = \prod_{k=1}^{k} \mu_k^{y_k}$

$$\ln P(y \mid x) = \ln \prod_{i=1}^{N} P(y^i \mid x) = \ln \prod_{i=1}^{N} \prod_{k=1}^{k} \mu_k^{y_k^i} = \ln \prod_{i=1}^{N} \prod_{k=1}^{k} h_k(x^i; W)^{y_k^i}$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{k} y_i^k \ln h_k(x^i; w)$$

$$E(w \mid D) = -\sum_{i=1}^{N} \sum_{k=1}^{k} y_k^i \ln h_k(x^i; W)$$