DZ-02 Mateo Jakšić, 0036535152

VO5 1.2.
2.3.
VO6 2.7.
VO7 1.1.

VO5 1.2. $K=3$

x_1	x_2	y
-3	1	0
-3	3	0
1	2	1
2	1	1
1	-2	2
2	-3	2

a) OVR

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

$$w_k = \Phi^+ \cdot y_k \Rightarrow w_0 = \left(\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -3 & -3 & 1 & 2 & 1 & 2 \\ 1 & 3 & 2 & 1 & -2 & -3 \end{bmatrix} \cdot \begin{bmatrix} 1 & -3 & 1 \\ 1 & -3 & 3 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \\ 1 & 1 & -2 \\ 1 & 2 & -3 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ -3 & -3 & 1 & 2 & 1 & 2 \\ 1 & 3 & 2 & 1 & -2 & -3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$= \dots$ (pomoću pythona) $=$

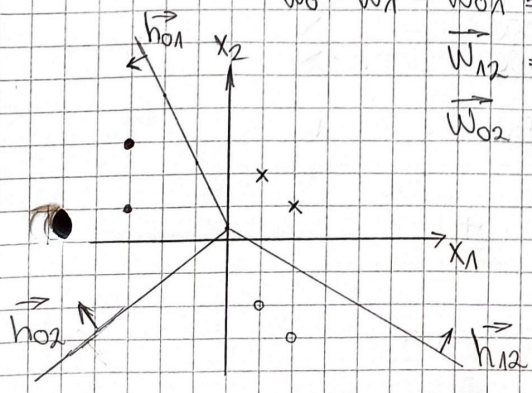
$$w_0 = \begin{bmatrix} 0.335 & -0.217 & -0.005 \end{bmatrix}^T$$
$$w_1 = \begin{bmatrix} 0.259 & 0.234 & 0.223 \end{bmatrix}^T$$
$$w_2 = \begin{bmatrix} 0.406 & -0.017 & -0.217 \end{bmatrix}^T$$

$\Downarrow$

$$h_0(x) = 0.335 - 0.217 x_1 - 0.005 x_2$$
$$h_1(x) = 0.259 + 0.234 x_1 + 0.223 x_2$$
$$h_2(x) = 0.406 - 0.017 x_1 - 0.217 x_2$$

b) $h_{01}(x) = w_{01,0} + w_{01,1} \cdot x_1 + w_{01,2} \cdot x_2$

$$\vec{w_0} - \vec{w_1} = \vec{w_{01}} = \begin{bmatrix} 0.076 & -0.451 & -0.218 \end{bmatrix}$$
$$\vec{w_{12}} = \begin{bmatrix} -0.147 & 0.251 & 0.440 \end{bmatrix}$$
$$\vec{w_{02}} = \begin{bmatrix} -0.071 & -0.200 & 0.212 \end{bmatrix}$$



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

$h_0(x) = 0.537$

$h_1(x) = 0.691$ ← max ⇒ kao klasu 1 jer za nju dobivamo najveću vrijednost hipoteze

$h_2(x) = 0.228$

d) Ne, jer OVR nema probabilističku interpretaciju

e) Prednost je da kod OVR treniramo manje modela, tj K modela, a kod OVO $\binom{K}{2}$

Nedostatak je ako ima manje primjera neke klase nasprem drugih zbog čega u želji smanjenja empirijske pogreške dođe do klasificiranja primjera iz manje klase u većinsku klasu

f) Jer nema probabilističku interpretaciju te ne možemo opisati probleme s više klasa, a to je poanta klasifikacije (imamo 3 klase u ovom primjeru

VOS 2.3 $N = 1000$, $n = 555$, $K = 4$, 400, 300, 200, 100 primjera

$\Phi \Rightarrow$ dim $N \times n+1$

stabilno ako $N \geq n+1$

OVO	T	⊥	dim	STABILNO
h_{01}	400	300	700×556	✓
h_{02}	400	200	600×556	✓
h_{03}	400	100	500×556	✗
h_{12}	300	200	500×556	✗
h_{13}	300	100	400×556	✗
h_{23}	200	100	300×556	✗
OVR			1000×556	✓

VOG 2.7 $\lambda = 1000$, $\eta = 0.01$, $\phi(x) = (1, x_1, x_2, x_1 x_2)$

$w = (0.2, 0.5, -1.1, 2.7)$

$(x, y) = ((-1, 2), 1)$, SGD

$\nabla_w L = (h(x) - y)\,\phi(x)$

$h(x) = \sigma(w^T \phi(x))$

$= \sigma\left(\begin{bmatrix} 0.2 & 0.5 & -1.1 & 2.7 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 2 \\ -2 \end{bmatrix}\right)$

$= \sigma(-7.9) = \frac{1}{1 + e^{+7.9}}$

$= 3.70606 \cdot 10^{-4} \approx 0$

$\nabla_w L = (0 - 1)\,\phi(x) = \begin{bmatrix} -1 \\ 1 \\ -2 \\ 2 \end{bmatrix}$

Mateo Jakšić 0036535152

$w_1 = w_0 + w_1 \eta \lambda - \eta \nabla_w L$

$= 0.5 - 0.5 \cdot 0.01 \cdot 1000 - 0.01 \cdot 1$

$= -4.51$

$\nabla w = 0.5 - (-4.51) \approx 5$ (B)

VO7. 1.1. a) ~~SOFTMAX~~, $\mathbb{R}^n \to \mathbb{R}^n$

$$\text{softmax}_k(x_1, \dots, x_n) = \frac{\exp(x_k)}{\sum_j \exp(x_k)}$$

$\text{softmax}_2(d) = 2.353 \cdot 10^{-3}$

$\text{softmax}_8(d) = 0.9495$

$\text{softmax}_1(d) = 8.658 \cdot 10^{-4}$

$\text{softmax}_5(d) = 0.0473$

1. normalizira sve vrijednosti da su u zbroju 1
2. povećava veće vrijednosti, smanjuje manje vrijednosti

b)

$$h_k(x, W) = \frac{\exp(w_k^T \phi(x))}{\sum_j \exp(w_j^T \phi(x))} = P(y = k \mid x, W)$$

$$W = [w_1 \; w_2 \; \dots \; w_K]$$

$$h(x) = \underset{k}{\operatorname{argmax}} \left(h_k(x, w) \right)$$

c)

$$E(w \mid D) = -\ln P(y \mid x, w)$$

$$= -\ln \prod_{i=1}^{N} P(y^i \mid x^i, w)$$

$$= -\ln \prod_{i=1}^{N} \prod_{k=1}^{K} h_k(x^i, w)^{y_k^i}$$

$$= -\sum_{i=1}^{N} \sum_{k=1}^{K} y_k^i \ln(h_k(x^i, w))$$

Mateo
Jakšić
0036535152