SU1: 2. DOMAĆA ZADAĆA

V05 - Zadaci za učenje

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

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

$\vec{w}_0 = \Phi^+ y_0 = \begin{bmatrix} 0.335 \\ -0.217 \\ -0.005 \end{bmatrix} \qquad h_0 = 0.335 - 0.217x_1 - 0.005x_2$

$\vec{w}_1 = \Phi^+ y_1 = \begin{bmatrix} 0.259 \\ 0.234 \\ 0.222 \end{bmatrix} \qquad h_1 = 0.259 + 0.234x_1 + 0.222x_2$

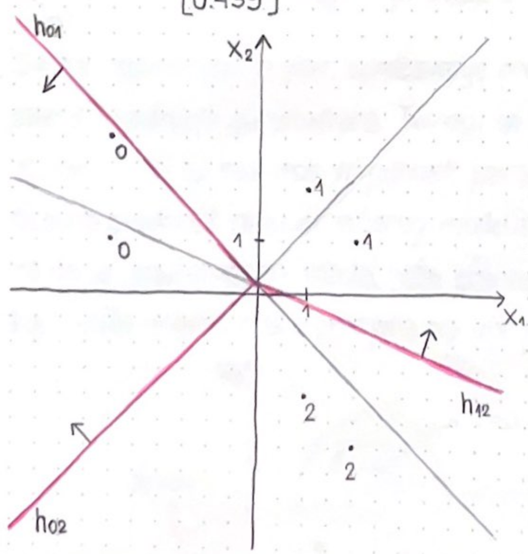
$\vec{w}_2 = \Phi^+ y_2 = \begin{bmatrix} 0.406 \\ -0.017 \\ -0.217 \end{bmatrix} \qquad h_2 = 0.406 - 0.017x_1 - 0.217x_2$

b) $h_{ij} = w_{0,ij} + w_{1,ij}x_1 + w_{2,ij}x_2$

$\vec{w}_{01} = \vec{w}_0 - \vec{w}_1 = \begin{bmatrix} 0.076 \\ -0.451 \\ -0.227 \end{bmatrix} \qquad h_{01} = 0.076 - 0.451x_1 - 0.227x_2 = 0 \Rightarrow x_2 = -1.987x_1 + 0.335$

$\vec{w}_{02} = \vec{w}_0 - \vec{w}_2 = \begin{bmatrix} -0.071 \\ -0.200 \\ 0.212 \end{bmatrix} \qquad h_{02} = -0.071 - 0.200x_1 + 0.212x_2 = 0 \Rightarrow x_2 = 0.943x_1 + 0.335$

$\vec{w}_{12} = \vec{w}_1 - \vec{w}_2 = \begin{bmatrix} -0.147 \\ 0.251 \\ 0.439 \end{bmatrix} \qquad h_{12} = -0.147 + 0.251x_1 + 0.439x_2 = 0 \Rightarrow x_2 = -0.572x_1 + 0.335$



d) Ne možemo znati koja je vjerojatnost da $x=(-1,3)$ pripada klasi 1 jer nemamo model s probabilističkom interpretacijom.

c) Primjer $x=(-1,3)$ bio bi klasificiran u klasi 1 jer se nalazi između hiperravnina h_{12} i h_{23}, odnosno na 'negativnoj' strani h_{01} i 'pozitivnoj' strani h_{12}.

e) Prednost OVR sheme nad ovo shemom jest da imamo manji broj modela koje moramo trenirati (linearna ovisnost o broju klasa umjesto kvadratne). S druge strane, nedostatak je taj da OVR shema lako rezultira neuravnoteženim brojem primjera između parova klasa za koje treniramo model.

f) Problem koji se javlja kod pokušaja klasifikacije linearnom regresijom jest to što izlazi takvog modela nemaju nikakvu vjerojatnosnu interpretaciju. Još veći nedostatak je nerobusnost takvog modela. Budući da algoritam regresije minimizira kvadratno odstupanje, „pretočno" klasificirani primjeri koji su duboko u području neke klase pomaknut će regresijsku granicu k sebi. To može dovesti do pogrešne klasifikacije nekih drugih primjera. Npr. u prethodnom primjeru bi dodavanje primjera ((7,3),1) pomaklo pravac h_{01} tako da bi primjer ((-3,3),0) postao pogrešno klasificiran u klasu 1.

V05 - zadaci s ispita

4. $\mathcal{D} = \{(x(i), y(i))\} = \{((1,0),+1), ((2,-3),-1), ((2,5),-1)\}$

$$X = \begin{bmatrix} 1 & 0 \\ 2 & -3 \\ 2 & 5 \end{bmatrix} \quad y = \begin{bmatrix} +1 \\ -1 \\ -1 \end{bmatrix} \quad \Phi = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 2 & -3 & -6 & 4 & 9 \\ 1 & 2 & 5 & 10 & 4 & 25 \end{bmatrix} \quad \vec{w} = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 2 \\ -2 \\ 0 \end{bmatrix}$$

$$E(\vec{w} \mid \mathcal{D}) = \sum_{i=1}^{3} \max(0, -w^T \Phi(x(i))\, y(i)) = \max(0,1) + \max(0,-16) + \max(0,8) = 9$$

V06 - Zadaci s ispita

5. $\Phi(x) = (1, x_1, x_2, x_1 x_2)$

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

$(\vec{x}, y) = ((-0.5, 2), 1) \Rightarrow \Phi(x) = \begin{bmatrix} 1 & -0.5 & 2 & -1 \end{bmatrix}$

$$h(\vec{x}; w) = \frac{1}{1 + \exp(-w^T \Phi(x))} = \frac{1}{1 + \exp(-4.95)} = 0.00703$$

$$\nabla L(y, h(x)) = (h(x) - y)\Phi(x) = \begin{bmatrix} -0.99297 & 0.49685 & -1.98594 & 0.99297 \end{bmatrix}$$

$\|\nabla L\|_2 = 2.48$

V07 - Zadaci za učenje

1\. a) Funkcija softmax za neki ulazni primjer x uzima vrijednosti $w_k^T x$ za svaku od K klasa te ih preslikava u K-dimenzijski vektor čije se komponente zbrajaju u 1.

$$\text{softmax}: \mathbb{R}^n \to \mathbb{R}^n \qquad \text{softmax}_k(x_1, \dots, x_n) = \frac{\exp(x_k)}{\sum_j \exp(x_j)}$$

Funkcija softmax ima dva efekta: normalizira vrijednosti tako da u zbroju budu 1 te „pojačava" veće vrijednosti i smanjuje manje vrijednosti.

$$\alpha = [2, 8, 1, 5]$$

$$\left.\begin{aligned} &\text{softmax}_2(\alpha) = \frac{\exp(2)}{\sum_j \exp(x_j)} = 0.002 \\ &\text{softmax}_8(\alpha) = 0.949 \\ &\text{softmax}_1(\alpha) = 0.001 \\ &\text{softmax}_5(\alpha) = 0.047 \end{aligned}\right\} \text{softmax}(\alpha) = [0.002\ 0.949\ 0.001\ 0.047]$$

b) Model multinomijalne logističke regresije definiramo kao skup modela $\{h_k\}_k$, gdje je svaki model zadužen za k-tu od ukupno K klasa:

$$h_k(x; w) = \frac{\exp(w_k^T \Phi(x))}{\sum_j \exp(w_j^T \Phi(x))} = P(y = k | x, w)$$

c) Oznake klasa definiramo kao kategoričke varijable i prikazujemo kao vektor indikatorskih binarnih varijabli: $y = (y_1, \dots, y_k)^T$, pri čemu je $\sum_k y_k = 1$.

Definiramo i vektor parametara $\mu = (\mu_1, \dots, \mu_k)$, pri čemu za parametre μ_k vrijedi $\sum_k \mu_k = 1$ i $\mu_k \geq 0$.

Distribuciju za kategoričku varijablu definiramo kao $P(y|\mu) = \prod_{k=1}^{K} \mu_k^{y_k}$.

$$\begin{aligned} \ln P(y|X) &= \ln \prod_{i=1}^{N} P(y^i | x) = \ln \prod_{i=1}^{N} \prod_{k=1}^{K} \mu_k^{y_k^i} = \ln \prod_{i=1}^{N} \prod_{k=1}^{K} h_k(x^i; w)^{y_k^i} \\ &= \sum_{i=1}^{N} \sum_{k=1}^{K} y_i^k \ln h_k(x^i; w) \end{aligned}$$

Funkcija pogreške koju minimiziramo negativan je logaritam vjerojatnosti oznaka: $E(w|\mathcal{D}) = -\sum_{i=1}^{N} \sum_{k=1}^{K} y_k^i \ln h_k(x^i; w)$.

V07 - Zadaci s ispita

7. $K = 3$

$$h_k(x) = \text{softmax}_k\left(\sum_{j=0}^{3} w_{j,k}\,\Phi_j(x)\right)$$

$d = 10$

→ ulazni sloj ima 3 dimenzije
→ drugi sloj ima 3 dimenzije
→ zadnji sloj ima 3 dimenzije

$\underbrace{3}_{\text{bias}} + 10 \cdot 3 = 33$

$\underbrace{3}_{\text{bias}} + 3 \cdot 3 = 12$

45 parametara