VJ-5

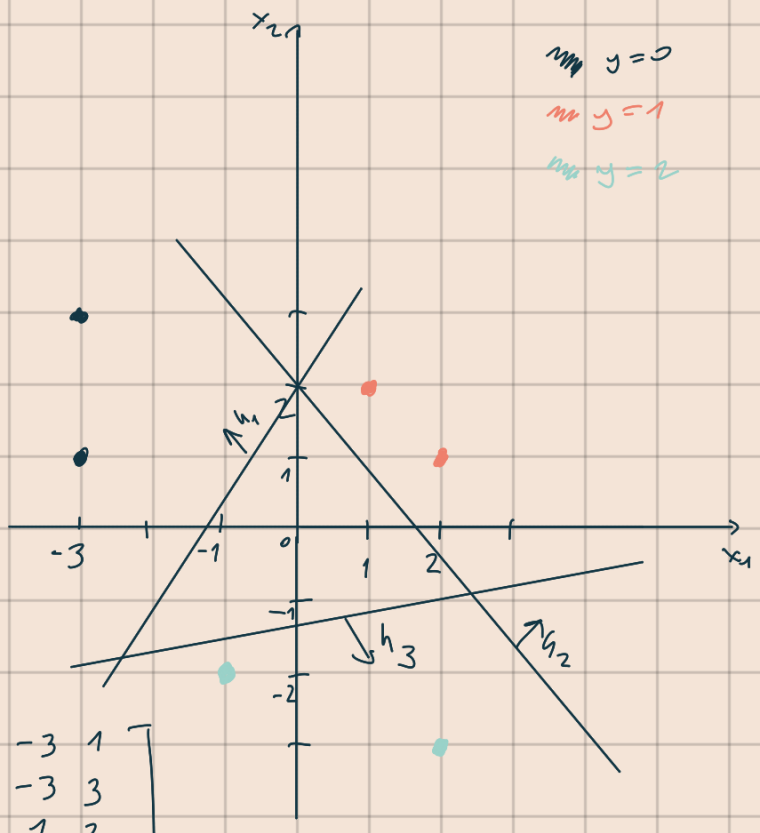
Ante Čamar, 0036540817

matrixcalc.org za inverz

1.2 $h(x) = \underset{j}{\operatorname{argmax}}\ h_j(x)$

a)

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



---

h_1, $y = 0$

$\vec{y}_1 = [1\ 1\ 0\ 0\ 0\ 0]^T$

$$\Phi = \begin{bmatrix} 1 & -3 & 1 \\ 1 & -3 & 3 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \\ 1 & 1 & -2 \\ 1 & 2 & -3 \end{bmatrix}$$

$\vec{w}_1^* = \Phi^+ \cdot \vec{y}_1$

$$\vec{w}_1 = \begin{bmatrix} 0,3351 \\ -0,2173 \\ -0.0052 \end{bmatrix}$$

$h_1(x) = 0,3351 - 0,2173 x_1 - 0,0052 x_2$

---

h_2, $y = 1$

$\vec{y}_2 = [0\ 0\ 1\ 1\ 0\ 0]$

$$\vec{w}_2 = \begin{bmatrix} 0.2592 \\ 0.2343 \\ 0.2225 \end{bmatrix}$$

$h_2(x) = 0,2592 + 0.2343 x_1 + 0.2225 x_2$

---

h_3, $y = 2$

$\vec{y}_3 = [0\ 0\ 0\ 0\ 1\ 1]$

$$\vec{w}_3 = \begin{bmatrix} 0.4058 \\ -0.017 \\ -0.2173 \end{bmatrix}$$

$h_3(x) = 0,4058 - 0.017 x_1 - 0,2173 x_2$

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

Ante Čamar, 0036540817

↳ OVO (One vs. One)

h_{01}, simetrala između h_1 i h_2

↳ $\vec{w}_{01} = \vec{w}_0 - \vec{w}_1$

$$\begin{bmatrix} 0,3351 \\ -0,2173 \\ -0.0052 \end{bmatrix} - \begin{bmatrix} 0.2592 \\ 0.2343 \\ 0.2225 \end{bmatrix} =$$

$$= w_{01} = \begin{bmatrix} 0.0759 \\ -0.451 \\ -0.227 \end{bmatrix}$$

$h_{1,2}$

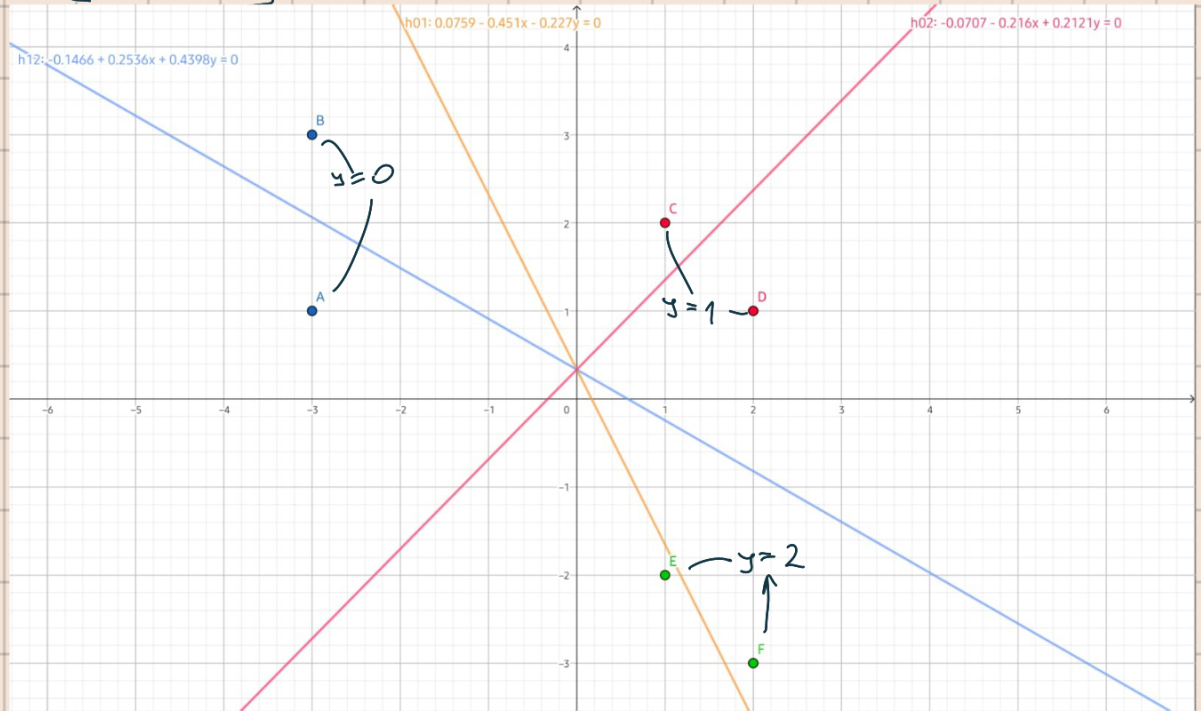
$$\begin{bmatrix} 0.2592 \\ 0.2343 \\ 0.2225 \end{bmatrix} - \begin{bmatrix} 0.4058 \\ -0.017 \\ -0.2173 \end{bmatrix}$$

$$w_{1,2} = \begin{bmatrix} -0.1466 \\ 0.236 \\ 0.4398 \end{bmatrix}$$

$h_{0,2}$

$$\begin{bmatrix} 0,3351 \\ -0,2173 \\ -0.0052 \end{bmatrix} - \begin{bmatrix} 0.4058 \\ -0.017 \\ -0.2173 \end{bmatrix}$$

$$= w_{0,2} = \begin{bmatrix} -0.0707 \\ -0.215 \\ 0.2121 \end{bmatrix}$$



Ante Čamer, 0036540817

c)

$X = (-1, 3)$ $y = ?$

Gledajući sa grafa: $y = 1$

Računajući sa OVR-ovi:

$h_1 = 0.5368$

$h_2 = 0.6924$

$h_3 = -0.2291$

$h(x) = \underset{y}{\text{argmax}}\ h_y(x)$

d) Ne možemo reći koja je vjerojatnost da neki primjer pripada nekoj klasi jer linearni modeli bijk smo uvijek koristili nemaju probabilističku interpretaciju.

e) OVR vs OVO

PREDNOST: Treniramo manje modela u odnosu na OVO (K vs $\binom{K}{2}$ gdje je K broj klasa).

NEDOSTATAK: Neuravnoteženost klasa - ako pretpostavimo da za svaku klasu imamo isti broj primjera onda za bilo koji klasifikator imamo 1:9 omjer primjera.

f) Tako što preriše križnjava primjere koji izvorno pripadaju nekoj klasi. Npr. $\{(50, 3), 1\}$

2.3

Ante Čamar, 0036540817

$N = 1000 \qquad \lambda = 0$

$n = 555$

$k = 4$

$|k_i| = \{ \overset{i=1}{400}, \overset{2}{300}, \overset{3}{200}, \overset{4}{100} \}$

OVO

Broj modela

$\binom{4}{2} = 6$

P	N	Dim ϕ
400	300	700 x
400	200	600 x
400	100	500 x
300	200	500 x
300	100	400 x
200	100	300 x

555 - nestabilni $N < n$

4

OVR

Broj modela = 4

P	N	Dim ϕ
400	600	1000 x 555
300	700	$N \gg n$
200	800	
100	900	

c)

VO-6

Ante Čamar, 0036540817

2.7

SGD

$\lambda = 1000$

$\eta = 0.01$

$\mathbb{R}^2 \to \mathbb{R}^3$

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

UČENJE:

$$w \leftarrow w\underbrace{(1-\eta\lambda)}_{-9} - \eta \sum_{i=1}^{\tilde{N}} \underbrace{(h(x^{(i)}) - y^{(i)})}_{\Delta_w L\ [1]} x^{(i)}$$

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

Novi primjer: $(-1,2), 1$

$$h = \frac{1}{1 + \exp(-w^T \phi(x))} \approx 0 \qquad 0 - 1 = -1 \qquad -1 \cdot x_i \to$$

$$w \leftarrow w(1-\eta\lambda) + \eta x^{(i)}$$

$-0,2 + 0.5 + 2.2 + 5.4$

$= 7.9$

$w \leftarrow -9w + \eta x$

$$w \leftarrow \begin{bmatrix} -1.8 \\ -4.5 \\ 9.9 \\ -24.3 \end{bmatrix} + \begin{bmatrix} 0 \\ -0,01 \\ 0,02 \\ -0,02 \end{bmatrix}$$

$$w \leftarrow \begin{bmatrix} -1,8 \\ -4.51 \\ 9.92 \\ -24,32 \end{bmatrix}$$

Razlika u w_1 je -5. **B**

$\Delta w_1 = -4.51 - 0,5 = -5.01 \sim -5$

VJ - 7

Ante Čamax, 0036540817

1.1.

a) Za ulazni primjer X, softmax uzima vrijednosti $w_k^T x$ za svaku od K klasa ($K \in \mathbb{N}$), te ih preslikava u K-dimenzijski vektor čije se komponente zbrajaju u 1. ($softmax: \mathbb{R}^K \to \mathbb{R}^K$, gdje je Σ zbroj 1).

To nam nosi 2 posljedice: 1. NORMALIZIRA sve vrijednosti tako da u zbroju budu 1

2. POVEČAVA veće vrijednosti i SMANJUJE male vrijednosti (zato se i zove SOFTmax).

$$Softmax = \frac{\exp(x_k)}{\sum_j \exp(x_j)}, \quad a_i = (2, 8, 1, 5)$$

$\sum_j \exp(x_j) \rightarrow e^2 + e^8 + e^1 + e^5 \approx 3139,48$

$Softmax(2) \approx 0,002$ $Softmax(8) \approx .9495$ $Softmax(1) \approx .001$ $Softmax(5) \approx .047$

b)

Koristimo zasebni vektor težina za svaku od K klasa, ali onda skalarni umnožak $w_k^T x$ propuštamo kroz softmax koja će se pobrinuti da se vjerojatnosti svih klasa zbrajaju na 1.

C) Ante Čavar, 0036540817

y_k → klasa

$$P(y|\mu) = \mu^y (1-\mu)^{1-y}$$ → Bernoullijeva razdioba

μ ... vjerojatnost ishoda

↳ ovo možemo poopćiti za K klasa

$$P(\mathbf{y}|\mu) = \prod_{k=1}^{K} \mu_k^{y_k}$$

Model linearne logističke regresije

$$h_k(x; w) = \frac{\exp(w_k^T \phi(x))}{\sum_j \exp(w_j^T \phi(x))} = \mu_k$$

→ vjerojatnost pripadnosti nekoj klasi

Logaritam vjerojatnosti označke → $\ln P(y|X)$

$$\ln[P(y|X)] = \ln \prod_{i=1}^{N} P(y^i|X)$$

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

("uzimamo" ln)

$$= \sum_{i=1}^{N} \sum_{k=1}^{K} y_k^i \ln[h_k(x_i, w)]$$

Pogreška modela → NEGATIVAN log. vjerој. označke

$$E(W|D) = -\sum_{i=1}^{N} \sum_{k=1}^{K} y_k^i \ln[h_k(x_i, w)]$$

$h_k < 1$, $\ln < 0$

↳ Z log ovaj razlog stavljamo minus ispred sume, jer želimo pozitivnu pogrešku (negativna pogreška bi zapravo bila "nagrada")