UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Cuaderno de trabajo:

Regresión logística

Albert Sanchis

DSIC

Departamento de Sistemas
Informáticos y Computación

Objetivos formativos

- Calcular vectores de logits
- Aplicar la función softmax a los vectores de logits
- Aplicar la regla de decisión de un clasificador basado en la función softmax
- Calcular el gradiente de la función NLL
- Aplicar descenso por gradiente para actualizar la matriz de pesos

- ***Cuestión 1***: La siguiente tabla presenta un conjunto de 2 muestras de entrenamiento $(n = \{1, 2\})$ de 2 dimensiones (x_{n1}, x_{n2}) procedentes de 2 clases $(c_n = \{1, 2\})$:

n	x_{n1}	x_{n2}	c_n
1	1	1	2
2	0	1	1

Adicionalmente, la siguiente tabla representa una matriz de pesos iniciales $\mathbf{W}$ con los pesos de cada una de las clases por columnas (en notación homogénea):

$\boldsymbol{w}_1$	$\boldsymbol{w}_2$
0	0
$-0{,}25$	$0{,}25$
0	0

Se pide:

1. Calcula el vector de logits asociado a cada muestra de entrenamiento.

$$\boldsymbol{a} = f(\boldsymbol{x}; \mathbf{W}) = \mathbf{W}^t \boldsymbol{x}$$

$$\boldsymbol{a}_1 = \mathbf{W}^t \boldsymbol{x}_1 = \begin{pmatrix} 0 & -0{,}25 & 0 \\ 0 & 0{,}25 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -0{,}25 \\ 0{,}25 \end{pmatrix}$$

$$\boldsymbol{a}_2 = \mathbf{W}^t \boldsymbol{x}_2 = \begin{pmatrix} 0 & -0{,}25 & 0 \\ 0 & 0{,}25 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

2. Aplica la función softmax al vector de logits de cada muestra de entrenamiento.

$$\boldsymbol{\mu_1} = \mathcal{S}(\boldsymbol{a}_1) = \begin{pmatrix} \frac{e^{-0,25}}{e^{-0,25}+e^{0,25}} \\ \frac{e^{0,25}}{e^{-0,25}+e^{0,25}} \end{pmatrix} = \begin{pmatrix} 0,3775 \\ 0,6225 \end{pmatrix}$$

$$\boldsymbol{\mu_2} = \mathcal{S}(\boldsymbol{a}_2) = \begin{pmatrix} \frac{e^{0}}{e^{0}+e^{0}} \\ \frac{e^{0}}{e^{0}+e^{0}} \end{pmatrix} = \begin{pmatrix} 0,5 \\ 0,5 \end{pmatrix}$$

3. Clasifica cada una de las muestras de entrenamiento. En caso de empate, elige cualquier clase.

$$c(\boldsymbol{x}) = \arg\max_c \ \mathcal{S}(\boldsymbol{a})_c \quad \textbf{\textit{donde}} \quad \boldsymbol{a} = \mathbf{W}^t \boldsymbol{x}$$

$$c(\boldsymbol{x_1}) = \arg\max_c \ \mathcal{S}(\boldsymbol{a}_1)_c = 2$$

$$c(\boldsymbol{x_2}) = \arg\max_c \ \mathcal{S}(\boldsymbol{a}_2)_c = 1 \quad (o \ \ 2)$$

4. Calcula el gradiente de la función NLL en el punto de la matriz de pesos iniciales.

$$\begin{pmatrix} \frac{\partial\,\mathrm{NLL}}{\partial W_{11}} & \cdots & \frac{\partial\,\mathrm{NLL}}{\partial W_{1C}} \\ \vdots & \ddots & \vdots \\ \frac{\partial\,\mathrm{NLL}}{\partial W_{D1}} & \cdots & \frac{\partial\,\mathrm{NLL}}{\partial W_{DC}} \end{pmatrix} = \frac{\partial\,\mathrm{NLL}}{\partial \mathbf{W}^t} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n(\boldsymbol{\mu}_n - \boldsymbol{y}_n)^t$$

$$\boldsymbol{x}_1(\boldsymbol{\mu}_1 - \boldsymbol{y}_1)^t = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \left(\begin{pmatrix} 0{,}3775 & 0{,}6225 \end{pmatrix} - \begin{pmatrix} 0 & 1 \end{pmatrix} \right)$$

$$= \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 0{,}3775 & -0{,}3775 \end{pmatrix}$$

$$= \begin{pmatrix} 0{,}3775 & -0{,}3775 \\ 0{,}3775 & -0{,}3775 \\ 0{,}3775 & -0{,}3775 \end{pmatrix}$$

$$
\begin{aligned}
\boldsymbol{x}_2(\boldsymbol{\mu}_2 - \boldsymbol{y}_2)^t &= \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \left(\begin{pmatrix} 0{,}5 & 0{,}5 \end{pmatrix} - \begin{pmatrix} 1 & 0 \end{pmatrix} \right) \\
&= \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} -0{,}5 & 0{,}5 \end{pmatrix} \\
&= \begin{pmatrix} -0{,}5 & 0{,}5 \\ 0 & 0 \\ -0{,}5 & 0{,}5 \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n(\boldsymbol{\mu}_n - \boldsymbol{y}_n)^t &= \frac{1}{2}\left(\begin{pmatrix} 0{,}3775 & -0{,}3775 \\ 0{,}3775 & -0{,}3775 \\ 0{,}3775 & -0{,}3775 \end{pmatrix} + \begin{pmatrix} -0{,}5 & 0{,}5 \\ 0 & 0 \\ -0{,}5 & 0{,}5 \end{pmatrix}\right) \\
&= \begin{pmatrix} -0{,}06 & 0{,}06 \\ 0{,}19 & -0{,}19 \\ -0{,}06 & 0{,}06 \end{pmatrix}
\end{aligned}
$$

5. Actualiza la matriz de pesos iniciales aplicando descenso por gradiente con factor de aprendizaje $\eta = 1{,}0$

$$\mathbf{W}_{i+1} = \mathbf{W}_i - \eta_i \, \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{x}_n (\boldsymbol{\mu}_n - \boldsymbol{y}_n)^t$$

$$\begin{aligned}
\mathbf{W}_1 &= \mathbf{W}_0 - \eta \, \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{x}_n (\boldsymbol{\mu}_n - \boldsymbol{y}_n)^t \\
&= \begin{pmatrix} 0 & 0 \\ -0{,}25 & 0{,}25 \\ 0 & 0 \end{pmatrix} - 1{,}0 \begin{pmatrix} -0{,}06 & 0{,}06 \\ 0{,}19 & -0{,}19 \\ -0{,}06 & 0{,}06 \end{pmatrix} \\
&= \begin{pmatrix} 0{,}06 & -0{,}06 \\ -0{,}44 & 0{,}44 \\ 0{,}06 & -0{,}06 \end{pmatrix}
\end{aligned}$$