Ejercicio diapositivas

Dado el clasificador que se muestra abajo y la muestra $x = (1,1)^t$, aplica la función softmax a x y calcula el vector de probabilidades que indica la probabilidad de que la muestra x pertenezca a cada clase.

$g_A(y) = -1 - 2y_1 + y_2$ $\qquad$ $g_B(y) = -y_1 + 2y_2$

$$\boldsymbol{a} = (a_1, a_2)^t$$

$$\begin{matrix} g_A(y) = -1 - 2y_1 + y_2 \\ g_B(y) = -y_1 + 2y_2 \end{matrix} \xrightarrow{\begin{pmatrix} 1 \\ 1 \end{pmatrix}} \begin{pmatrix} -2 \\ 1 \end{pmatrix} \xrightarrow{\textit{softmax}} \begin{pmatrix} 0.047 \\ 0.953 \end{pmatrix}$$

$$softmax(a_1) = \frac{e^{-2}}{e^{-2} + e^{1}} = \frac{0.135}{0.135+2.718} = 0.047$$

$$softmax(a_2) = \frac{e^{-1}}{e^{-2} + e^{1}} = \frac{2.718}{0.135+2.718} = 0.953$$

1 ☐ Cuestiones

Sea un modelo de regresión logística en notación compacta (homogénea) para un problema de clasificaciónn en $C = 3$ clases y datos representados por vectores de dimensión $D = 2$.

$$p(\boldsymbol{y} \mid \boldsymbol{x}, \mathbf{W}) = \mathrm{Cat}(\boldsymbol{y} \mid \mathcal{S}(\mathbf{W}^t \boldsymbol{x})) \text{ con } \mathbf{W}^t = \begin{pmatrix} 0 & 1 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{C \times D} .$$

Dado $\boldsymbol{x} = (0.5, 0.5)^t$, la probabilidad P de que $\boldsymbol{x}$ pertenezca a la clase 1 es:

A) $P < 0.25$
B) $0.25 \leq P < 0.5$
C) $0.5 \leq P < 0.75$
D) $0.75 \leq P$

C=3 clases

D=2 dimensiones

$$\mathbf{W}^t = \begin{pmatrix} 0 & 1 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{array}{l} \text{w1} \\ \text{w2} \\ \text{w3} \end{array}$$

$$\mathbf{W}^t = \begin{pmatrix} 0 & 1 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

vector logits

Respuesta C

$$(0 \quad 1 \quad 1)\begin{pmatrix} 1 \\ 0.5 \\ 0.5 \end{pmatrix} = 1$$

$$softmax(a_1) = \frac{e^{1}}{e^{1}+e^{0}+e^{0}} = \frac{e}{2+e} = 0.576$$

$$(0 \quad -1 \quad 1)\begin{pmatrix} 1 \\ 0.5 \\ 0.5 \end{pmatrix} = 0$$

$$softmax(a_2) = \frac{e^{0}}{e^{0}+e^{1}+e^{0}} = \frac{1}{2+e} = 0.212$$

$$(0 \quad 0 \quad 0)\begin{pmatrix} 1 \\ 0.5 \\ 0.5 \end{pmatrix} = 0$$

$$softmax(a_3) = \frac{e^{0}}{e^{0}+e^{1}+e^{0}} = \frac{1}{2+e} = 0.212$$

Problemas

1. Sea un modelo de regresión logística en notación compacta (homogénea) para un problema de clasificación en $C = 3$ clases y datos representados por vectores de dimensión $D = 2$.

$$p(\boldsymbol{y} \mid \boldsymbol{x}, \mathbf{W}) = \mathrm{Cat}(\boldsymbol{y} \mid \mathcal{S}(\mathbf{W}^t \boldsymbol{x})) \text{ con } \mathbf{W}^t = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix} \in \mathbb{R}^{C \times D}$$

Actualiza el valor de $\mathbf{W}$ con una iteración de descenso por gradiente con el conjunto de entrenamiento $\mathcal{D} = \{(\boldsymbol{x} = (1,1,1)^t, y = 1)\}$ y factor de aprendizaje $\eta = 0.1$.

C=3 clases

D=2 dimensiones

$$\mathbf{W}^t = \begin{array}{c} \begin{pmatrix} \boxed{1 \quad -1 \quad 0} \\ \boxed{0 \quad 1 \quad 0} \\ \boxed{1 \quad -1 \quad 1} \end{pmatrix} \end{array} \begin{array}{l} \text{w1} \\ \text{w2} \\ \text{w3} \end{array}$$

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

x

clase 1

$y = 1$

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

one-hot

descenso de gradiente

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

n	$\boldsymbol{x_n}$	$\boldsymbol{y_n}$	$\boldsymbol{a_n} = (a_{n1}, a_{n2})$	$\boldsymbol{\mu_n} = (S(a_{n1}), S(a_{n2}))$	$\boldsymbol{\mu_n} - \boldsymbol{y_n}$	$\boldsymbol{x}_n(\boldsymbol{\mu}_n - \boldsymbol{y}_n)^t$
1	$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 1 \end{pmatrix}^t$	$\frac{e^{0}}{e^0+e^1+e^1}$ $\frac{e^{1}}{e^1+e^0+e^1}$ $\frac{e^{1}}{e^1+e^0+e^1}$ $\begin{pmatrix} 0.155 & 0.422 & 0.422 \end{pmatrix}^t$	$\begin{pmatrix} 0.155 \\ 0.422 \\ 0.422 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ $\begin{pmatrix} -0.845 & 0.422 & 0.422 \end{pmatrix}^t$	$\begin{pmatrix} -0.845 & 0.422 & 0.422 \\ -0.845 & 0.422 & 0.422 \\ -0.845 & 0.422 & 0.422 \end{pmatrix}$

	$\mathbf{W_0}^t$	$\mathbf{W_1}^t$
w1 w2 w3	$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1.0845 & -0.915 & 0.0845 \\ -0.0422 & 0.958 & -0.0422 \\ 0.958 & -1.0422 & 0.958 \end{pmatrix}$

$$\boldsymbol{x}_n(\boldsymbol{\mu}_n - \boldsymbol{y}_n)^t$$

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \text{X} \begin{pmatrix} -0.845 & 0.422 & 0.422 \end{pmatrix}^t = \begin{pmatrix} -0.845 & 0.422 & 0.422 \\ -0.845 & 0.422 & 0.422 \\ -0.845 & 0.422 & 0.422 \end{pmatrix}$$

$$\frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n(\boldsymbol{\mu}_n - \boldsymbol{y}_n)^t$$ Gradiente de la NLL

$$\underbrace{\begin{pmatrix} -0.845 & 0.422 & 0.422 \\ -0.845 & 0.422 & 0.422 \\ -0.845 & 0.422 & 0.422 \end{pmatrix}}_{\text{Gradiente de la NLL}} \cdot 0.1 = \begin{pmatrix} -0.0845 & 0.0422 & 0.0422 \\ -0.0845 & 0.0422 & 0.0422 \\ -0.0845 & 0.0422 & 0.0422 \end{pmatrix}$$

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

$$\begin{pmatrix} 1 & 0 & 1 \\ -1 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} -0.0845 & 0.0422 & 0.0422 \\ -0.0845 & 0.0422 & 0.0422 \\ -0.0845 & 0.0422 & 0.0422 \end{pmatrix} = \begin{pmatrix} 1.0845 & -0.0422 & 0.958 \\ -0.915 & 0.958 & -1.042 \\ 0.0845 & -0.0422 & 0.958 \end{pmatrix}$$