# Ejercicios B2T2

Sistemas Inteligentes (SIN)
ETSInf - Grado en Informática

Curso 2023/2024

1. En un problema de clasificación en 2 clases, para objetos representados mediante vectores de características bidimensionales, se tienen dos muestras de entrenamiento: $\vec{y}_1 = (0,0)^t$, $\vec{y}_2 = (1,1)^t$ de clases $c_1 = 1$, $c_2 = 2$, respectivamente.

   Mostrar una traza de ejecución del algoritmo Perceptrón, con *vectores de pesos iniciales* nulos, *factor de aprendizaje* $\alpha = 1$ y *margen* $b = 0{,}1$. La traza debe incluir las sucesivas actualizaciones de los vectores de pesos de las clases.

2. Se tiene un problema de clasificación en dos clases, 0 y 1, para objetos representados en $\{0,1\}^2$, esto es, vectores de bits de la forma $\vec{x} = (x_1, x_2)^t$ con $x_1, x_2 \in \{0,1\}$. Asimismo, disponemos de cuatro muestras de entrenamiento:

| $x_n$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|
| $x_{n1}$ | 0 | 0 | 1 | 1 |
| $x_{n2}$ | 0 | 1 | 0 | 1 |
| $c_n$ | 0 | 1 | 1 | 0 |

   Se pide:

   *a*) Aplicar una iteración del algoritmo Perceptrón con pesos iniciales nulos, constante de aprendizaje $\alpha = 1$ y margen $b = 0{,}1$, mostrando los pesos resultantes

   *b*) ¿Convergerá el algoritmo Perceptrón a una solución sin datos de entrenamiento mal clasificados? Razona la respuesta

3. Dado el conjunto de muestras bidimensionales $\mathcal{X} = \{(-2,-2), (0,0), (2,2)\}$, cada una respectivamente de las clases 1, 2 y 3, y los vectores de pesos $\vec{w}_1 = (0,-2,-2)$, $\vec{w}_2 = (-1,0,0)$ y $\vec{w}_3 = (-1,4,4)$, realizar una iteración del algoritmo Perceptrón con factor de aprendizaje $\alpha = 1$ y margen $\gamma = 0{,}1$ usando los pesos iniciales propuestos, mostrando los pesos resultantes y las regiones de decisión correspondientes a cada clase

4. Dado un problema de clasificación en dos clases con dos datos bidimensionales $\mathcal{D} = \{((1,0,0)^t, (1,0)^t), ((1,1,1)^t, (0,1)^t)\}$:

   *a*) Realiza tres iteraciones del algoritmo de aprendizaje del modelo de regresión logística que minimiza la neg-log-verosimilitud con descenso por gradiente ($\eta = 1{,}0$) a partir de la matriz de pesos iniciales nulos

   *b*) Calcula la probabilidad a posteriori de los datos a partir de la matriz de pesos final obtenida en el anterior apartado.

   *c*) Clasifica los datos por máxima probabilidad a posteriori.

5. 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 $D = 2$.

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

   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 $\eta = 0{,}1$.

**Soluciones**

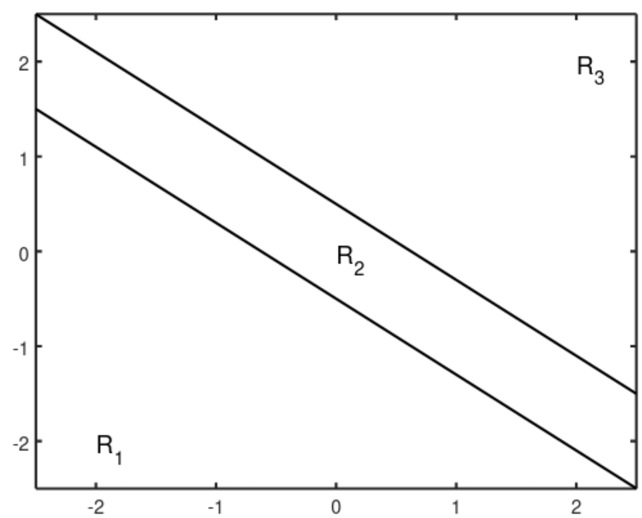
1. En total 3 iteraciones.

   - Iteración 1: $w_1 = (0,0,0)^t \to (1,0,0)^t \to (0,-1,-1)^t$, $w_2 = (0,0,0)^t \to (-1,0,0) \to (0,1,1)^t$
   - Iteración 2: $w_1 = (0,-1,-1)^t \to (1,-1,-1)^r$, $w_2 = (0,1,1)^t \to (-1,1,1)^t$
   - Iteración 3: no hay cambios

2. *a*) $\vec{w}_1 = \vec{w}_2 = (0,0,0)^t$

   *b*) No es posible porque las muestras no son linealmente separables ni siquiera usando el margen propuesto

3. Los pesos obtenidos son $\vec{w}_1 = (-1,-2,-2)$, $\vec{w}_2 = (0,0,0)$ y $\vec{w}_3 = (-2,4,4)$. Las regiones de decisión son:




4. *a*)

   Iteración 1: $W_1^t = \begin{pmatrix} 0{,}0 & -0{,}25 & -0{,}25 \\ 0{,}0 & 0{,}25 & 0{,}25 \end{pmatrix}$

   Iteración 2: $W_2^t = \begin{pmatrix} 0{,}115 & -0{,}385 & -0{,}385 \\ -0{,}115 & 0{,}385 & 0{,}385 \end{pmatrix}$

   Iteración 3: $W_3^t = \begin{pmatrix} 0{,}231 & -0{,}491 & -0{,}491 \\ -0{,}231 & 0{,}491 & 0{,}491 \end{pmatrix}$

   *b*) $P(c_1|x_1) = 0{,}613$, $P(c_2|x_1) = 0{,}387$, $P(c_1|x_2) = 0{,}182$, $P(c_2|x_2) = 0{,}818$

   *c*) Por máxima probabilidad a posteriori, $x_1$ se clasifica en $c_1$ y $x_2$ en $c_2$

5. 

$$\mathbf{W} = \begin{pmatrix} 1{,}0845 & -0{,}0422 & 0{,}9578 \\ -0{,}9156 & 0{,}9578 & -1{,}0422 \\ 0{,}0844 & -0{,}0422 & 0{,}9578 \end{pmatrix}$$