Examen de Aprendizaje Automático

ETSINF, Universitat Politècnica de València, 18 de enero de 2021

Apellidos: ______ **Nombre:** ______ **Grupo:** ______

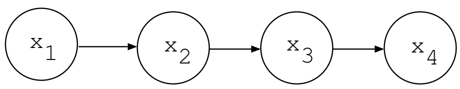
Cuestiones (2 puntos; tiempo estimado: 30 minutos)

Marca cada recuadro con una única opción de entre las dadas. Cada acierto suma 1/2 puntos y cada fallo resta 1/6 puntos.

1 $\boxed{C}$ Si la función de activación en la capa de salida de un perceptrón de dos capas es lineal, las fórmulas que permiten modificar los pesos de dicha capa de salida en el algoritmo BackProp verifican que (solo una respuesta es correcta):

A) $\Delta\theta^2_{ij} = \rho\ (t_i - s^2_i)\ g(\phi^2_i)\ (1 - g(\phi^2_i))\ s^1_j$
B) $\Delta\theta^2_{ij} = \rho\ (t_i - s^2_i)\ g(\phi^2_i)\ s^1_j$
C) $\Delta\theta^2_{ij} = \rho\ (t_i - s^2_i)\ s^1_j$
D) $\Delta\theta^2_{ij} = \rho\ (t_i - s^2_i)\ (1 - g(\phi^2_i))\ s^1_j$

2 $\boxed{A}$ En la red bayesiana lineal



¿cuál de las relaciones siguientes es correcta?

A) $P(x_1, x_4 \mid x_2) \ = \ P(x_1 \mid x_2)\ P(x_4 \mid x_2)$
B) $P(x_1, x_4 \mid x_2) \ = \ P(x_1 \mid x_2)\ P(x_3 \mid x_2)$
C) $P(x_1, x_4 \mid x_2) \ = \ P(x_3 \mid x_2)\ P(x_4 \mid x_2)$
D) $P(x_1, x_4 \mid x_2) \ = \ P(x_1)\ P(x_4)$

3 $\boxed{D}$ Se ha evaluado un sistema de Aprendizaje Automático mediante la técnica de *validación cruzada en B bloques* ("B-fold Cross Validation") con $B = 10$ y utilizando un conjunto de datos etiquetados que contiene 1000 muestras. Se han obtenido un total de 20 errores. Indicar cuál de las afirmaciones siguientes es correcta:

A) La talla de entrenamiento efectiva es de 1000 muestras y la talla de test efectiva es 1000 muestras.
B) La talla de entrenamiento efectiva es de 900 muestras y el error es del $20.0 \pm 0.2\,\%$
C) La talla de entrenamiento efectiva es de 1000 muestras y y el error es del $20.0 \pm 0.2\,\%$
D) La talla de entrenamiento efectiva es de 900 muestras y el error es del $2.0 \pm 0.9\,\%$

4 $\boxed{A}$ Considerar la siguiente modificación de la función de Widrow y Hoff

$$q_S(\boldsymbol{\theta}) = \sum_{n=1}^{N} \left(\boldsymbol{\theta}^t \boldsymbol{x}_n - y_n\right) + \frac{\lambda}{2}\ \boldsymbol{\theta},$$

Al aplicar la técnica de descenso por gradiente, en la iteración k el vector de pesos, $\boldsymbol{\theta}$, se modifica como: $\boldsymbol{\theta}(k+1) = \boldsymbol{\theta}(k) - \rho_k \boldsymbol{\nabla} q_S(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}(k)}$. En esta expresión, el gradiente, $\boldsymbol{\nabla} q_S(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}(k)}$, es:

A) $\displaystyle\sum_{n=1}^{N} \boldsymbol{x}_n + \frac{\lambda}{2}$

B) $\displaystyle\sum_{n=1}^{N} \boldsymbol{x}_n + 1$

C) $\displaystyle\sum_{n=1}^{N} \boldsymbol{x}_n + \lambda\ \boldsymbol{\theta}(k)$

D) $\displaystyle\sum_{n=1}^{N} \boldsymbol{\theta}(k)^t \boldsymbol{x}_n + 1$

Problema 1 (3 puntos; tiempo estimado: 30 minutos)

En la siguiente tabla se presenta una muestra de entrenamiento no linealmente separable en $\mathbb{R}^2$ y los correspondientes multiplicadores de Lagrange óptimos obtenidos al entrenar una máquina de vectores soporte con esta muestra (y C=10):

i	1	2	3	4	5	6	7	8
x_{i1}	1	3	4	4	3	1	3	2
x_{i2}	4	2	1	3	1	2	3	3
Clase	+1	+1	–1	–1	–1	+1	+1	–1
$\alpha_i^\star$	0.0	10.0	0.0	6.0	6.4	2.4	10.0	10.0

a) Obtener la función discriminante lineal correspondiente y el valor del margen.

b) Calcular las tolerancias de cada muestra de aprendizaje.

c) Representar gráficamente la frontera lineal de separación entre clases, los márgenes y las muestras de entrenamiento, indicando cuáles son vectores soporte.

d) Clasificar la muestra $(4,4)^t$.

a) Pesos de la función discriminante:

$\boldsymbol{\theta}^\star = c_2\ \alpha_2^\star\ \mathbf{x_2} + c_4\ \alpha_4^\star\ \mathbf{x_4} + c_5\ \alpha_5^\star\ \mathbf{x_5} + c_5\ \alpha_6^\star\ \mathbf{x_6} + c_5\ \alpha_7^\star\ \mathbf{x_7} + c_5\ \alpha_8^\star\ \mathbf{x_8}$

$$\theta_1^* \;=\; (+1)\ 10.0\ 3 + (-1)\ 6.0\ 4 + (-1)\ 6.4\ 3 + (+1)\ 2.4\ 1 + (+1)\ 10.0\ 3 + (-1)\ 10\ 2 \;=\; -0.8$$
$$\theta_2^* \;=\; (+1)\ 10.0\ 2 + (-1)\ 6.0\ 3 + (-1)\ 6.4\ 1 + (+1)\ 2.4\ 2 + (+1)\ 10.0\ 3 + (-1)\ 10\ 3 \;=\; +0.4$$

Usando el vector soporte $\mathbf{x_4}$ (que verifica la condición : $0 < \alpha_4^* < C = 10$)

$\theta_0^\star = c_4 - \boldsymbol{\theta}^{\star t}\mathbf{x_4} = -1 - ((-0.8)\ (4) + (0.4)\ (3)) \;=\; 1.0$

Margen: $\dfrac{2}{\|\boldsymbol{\theta}\|} \approx 2.23$

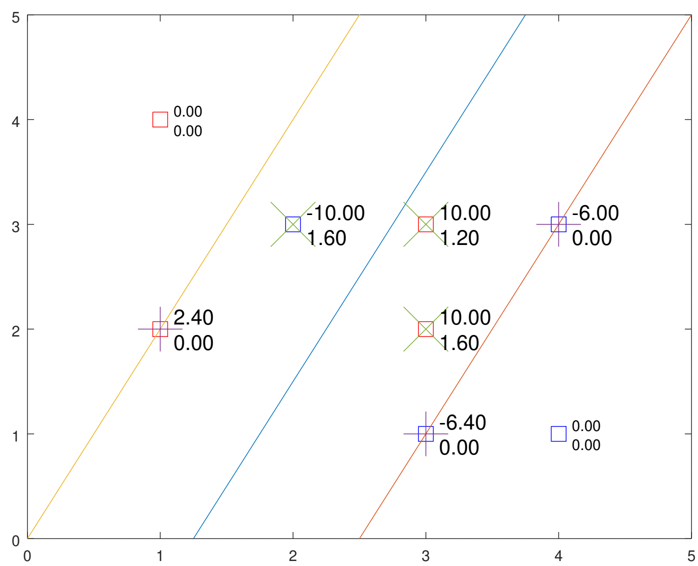
b) Calcular las tolerancias de cada muestra de aprendizaje:

$\zeta_1 = 0.0$; $\zeta_2 = 1 \;-\; c_2\ (\boldsymbol{\theta}^{\star t}\mathbf{x_2} + \theta_0^\star) = 1.6$; $\zeta_3 = 0.0$; $\zeta_4 = 0.0$;
$\zeta_5 = 0.0$; $\zeta_6 = 0.0$; $\zeta_7 = 1 \;-\; c_7\ (\boldsymbol{\theta}^{\star t}\mathbf{x_7} + \theta_0^\star) = 1.2$; $\zeta_8 = 1 \;-\; c_8\ (\boldsymbol{\theta}^{\star t}\mathbf{x_8} + \theta_0^\star) = 1.6$

c) Frontera de separación y representación gráfica:

Ecuación de la frontera lineal de separación: $-0.8\ x_1 + 0.4\ x_2 + 1.0 \;=\; 0$

Los vectores de entrenamiento son todos los de la tabla. De ellos, los vectores soporte son: $(2,3)^t, (4,2)^t, (3,2)^t, (2,2)^t\ (1,2)^t$.
El margen lo definen las dos rectas paralelas a la frontera de separación, cada una de ellas situada a una distancia de $2.23/2 \approx 1.12$ y cuyas ecuaciones son: $-0.8\ x_1 + 0.4\ x_2 + 1.0 \;=\; +1$ y $-0.8\ x_1 + 0.4\ x_2 + 1.0 \;=\; -1$
Representación gráfica:



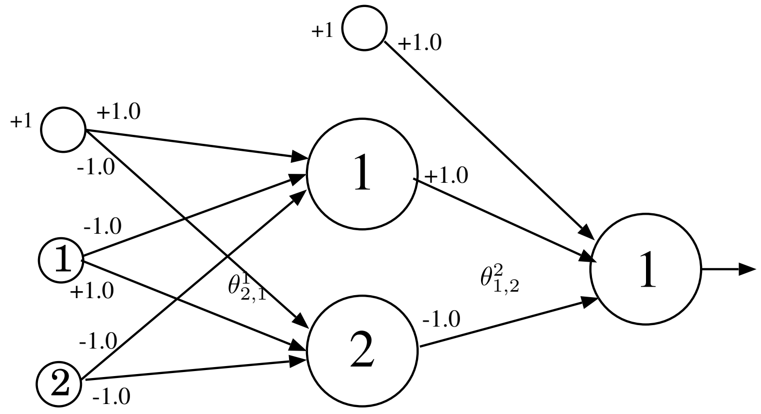
c) Clasificación de la muestra $(4,4)^t$:

El valor de la función discriminante para este vector es:

$\theta_1^*\ x_1 + \theta_2^*\ x_2\ +\ \theta_0^* = (-0.8) * 4 + (0.467) * 4 + 1.0 = -0.6 < 0\ \Rightarrow$ clase -1.

Problema 2 (3 puntos; tiempo estimado: 30 minutos)

El perceptrón multicapa de la figura se utiliza para resolver un problema de regresión, con función de activación del nodo de la capa de salida de tipo lineal y de los nodos de la capa oculta de tipo *sigmoide*, y factor de aprendizaje $\rho = 1.0$.



Dado un par de entrenamiento $(\boldsymbol{x}^t, t) = ((+2, +2), -1)$, calcular:

a) Las salidas de todos los nodos.

b) Los correspondientes errores en el nodo de la capa de salida y en los nodos de la capa oculta.

c) Los nuevos valores de los pesos de las conexiones $\theta^2_{1,2}$ y $\theta^1_{2,1}$.

a) Las salidas de la capa oculta son:

$\phi^1_1 = \theta^1_{1,0} + \theta^1_{1,1}\ x_1 + \theta^1_{1,2}\ x_2 = -3 \qquad s^1_1 = f_s(\phi^1_1) = +0.047426$

$\phi^1_2 = \theta^1_{2,0} + \theta^1_{2,1}\ x_1 + \theta^1_{2,2}\ x_2 = -1 \qquad s^1_2 = f_s(\phi^1_1) = +0.268941$

La salida de la capa de salida es:

$\phi^2_1 = \theta^2_{1,0} + \theta^2_{1,1}\ s^1_1 + \theta^2_{1,2}\ s^1_2 = +0.77848 \qquad s^2_1 = f_l(\phi^2_1) = +0.77848$

b) El error en la capa de salida es:

$\delta^2_1 = (t_1 - s^2_1) = -1.7785$

Los errores en la capa oculta son:

$\delta^1_1 = (\delta^2_1\ \theta^2_{1,1})\ f'_S(\phi^1_1) = (\delta^2_1\ \theta^2_{1,1})\ s^1_1\ (1 - s^1_1) = -0.08035$

$\delta^1_2 = (\delta^2_1\ \theta^2_{1,2})\ f'_S(\phi^1_2) = (\delta^2_1\ \theta^2_{1,2})\ s^1_2\ (1 - s^1_2) = +0.34967$

c) El nuevo peso $\theta^2_{1,2}$ es:

$\theta^2_{1,2} = \theta^2_{1,2} + \Delta\theta^2_{1,2} = \theta^2_{1,2} + \rho\ \delta^2_1\ s^1_2 = -1.0 + 1.0\ (-1.7785)\ 0.268941 = -1.47831$

El nuevo peso $\theta^1_{2,1}$ es:

$\theta^1_{2,1} = \theta^1_{2,1} + \Delta\theta^1_{2,1} = \theta^1_{2,1} + \rho\ \delta^1_2\ x_1 = +1.0 + 1.0\ (+0.34967)\ 2.0 = +1.69934$

Problema 3 (2 puntos; tiempo estimado: 20 minutos)

Las variables aleatorias A, B, C, D, E toman valores en el conjunto $\{0, 1\}$ y la variable F en el conjunto $\{x, y, z\}$. La distribución de probabilidad conjunta de estas variables viene dada por

$$P(A, B, C, D, E, F) \;=\; P(A)\ P(B \mid A)\ P(C \mid B)\ P(D \mid A)\ P(E \mid D)\ P(F \mid D, E),$$

con las correspondientes distribuciones de probabilidad:

$$\begin{array}{rclrcl}
P(A=1) &=& 0.3 & P(A=0) &=& 0.7 \\
P(B=1 \mid A=1) &=& 0.4 & P(B=0 \mid A=1) &=& 0.6 \\
P(B=1 \mid A=0) &=& 0.6 & P(B=0 \mid A=0) &=& 0.4 \\
P(C=1 \mid B=1) &=& 0.5 & P(C=0 \mid B=1) &=& 0.5 \\
P(C=1 \mid B=0) &=& 0.3 & P(C=0 \mid B=0) &=& 0.7 \\
P(D=1 \mid A=1) &=& 0.9 & P(D=0 \mid A=1) &=& 0.1 \\
P(D=1 \mid A=0) &=& 0.1 & P(D=0 \mid A=0) &=& 0.9 \\
P(E=1 \mid D=1) &=& 0.2 & P(E=0 \mid D=1) &=& 0.8 \\
P(E=1 \mid D=0) &=& 0.1 & P(E=0 \mid D=0) &=& 0.9
\end{array}$$

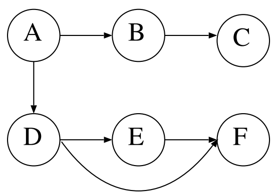
$$\begin{array}{rclrclrcl}
P(F=x \mid D=0, E=0) &=& 0.1 & P(F=y \mid D=0, E=0) &=& 0.2 & P(F=z \mid D=0, E=0) &=& 0.7 \\
P(F=x \mid D=0, E=1) &=& 0.0 & P(F=y \mid D=0, E=1) &=& 0.2 & P(F=z \mid D=0, E=1) &=& 0.8 \\
P(F=x \mid D=1, E=0) &=& 0.3 & P(F=y \mid D=1, E=0) &=& 0.3 & P(F=z \mid D=1, E=0) &=& 0.4 \\
P(F=x \mid D=1, E=1) &=& 0.4 & P(F=y \mid D=1, E=1) &=& 0.3 & P(F=z \mid D=1, E=1) &=& 0.3
\end{array}$$

a) Representar gráficamente la red bayesiana correspondiente

b) Obtener una expresión simplificada de $P(F \mid A, B, C, D)$.

c) Dados $A = B = C = D = 1$, ¿Cuál es la mejor predicción para el valor de F?.

a) Representar gráficamente la red bayesiana correspondiente



b) Obtener una expresión simplificada de $P(F \mid A, B, C, D)$

$$\begin{aligned}
P(F \mid A, B, C, D) &= \frac{P(A, B, C, D, F)}{P(A, B, C, D)} \\
&= \frac{\sum_e P(A)\ P(B \mid A)\ P(C \mid B)\ P(D \mid A)\ P(E=e \mid D)\ P(F \mid D, E=e)}{\sum_{e,f} P(A)\ P(B \mid A)\ P(C \mid B)\ P(D \mid A)\ P(E=e \mid D)\ P(F=f \mid D, E=e)} \\
&= \frac{\cancel{P(A)}\ \cancel{P(B \mid A)}\ \cancel{P(C \mid B)}\ \cancel{P(D \mid A)}\ \sum_e P(E=e \mid D)\ P(F \mid D, E=e)}{\cancel{P(A)}\ \cancel{P(B \mid A)}\ \cancel{P(C \mid B)}\ \cancel{P(D \mid A)}\ \sum_e P(E=e \mid D)\ \sum_f P(F=f \mid D, E)} \\
&= \sum_e P(E=e \mid D)\ P(F \mid D, E=e)
\end{aligned}$$

c) Dados $A = B = C = D = 1$, ¿Cuál es el mejor valor de F que se puede predecir?

$$f^\star \;=\; \underset{f \in \{x,y,z\}}{\arg\max}\ P(F=f \mid A=1, B=1, C=1, D=1) \;=\; \underset{f \in \{x,y,z\}}{\arg\max} \sum_e P(E=e \mid D=1)\ P(F=f \mid D=1, E=e)$$

$$P(F=x \mid A=1, B=1, C=1, D=1) \;=\; \sum_e P(E=e \mid D=1)\ P(F=x \mid D=1, E=e) \;=\; 0.2{\cdot}0.4 + 0.8{\cdot}0.3 \;=\; 0.32$$

$$P(F=y \mid A=1, B=1, C=1, D=1) \;=\; \sum_e P(E=e \mid D=1)\ P(F=y \mid D=1, E=e) \;=\; 0.2{\cdot}0.3 + 0.8{\cdot}0.3 \;=\; 0.30$$

$$P(F=z \mid A=1, B=1, C=1, D=1) \;=\; 1 \;-\; 0.32 \;-\; 0.30 \;=\; 0.38$$

El valor óptimo de F es $f^\star = z$