Examen de Aprendizaje Automático

ETSINF, Universitat Politècnica de València, 11 de enero de 2016

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 $\boxed{D}$ Se ha evaluado un sistema de Aprendizaje Automático mediante la técnica de *Exclusión individual* ("Leaving One Out") y utilizando un conjunto de datos que contiene 200 muestras. Se han obtenido un total de 10 errores. Indicar cuál de las afirmaciones siguientes es razonable:

A) La talla de entrenamiento efectiva es 190 muestras, la del test es de 10 muestras y el error estimado es $5.0\,\% \pm 0.3\,\%$
B) La talla de entrenamiento efectiva es de 199 muestras, la del test es de 1 muestra y el error estimado es $5.0 \pm 3.0\,\%$
C) La talla de entrenamiento efectiva es de 200 muestras, la del test es de 10 muestras y el error estimado es $5.0 \pm 0.3\,\%$
D) La talla de entrenamiento efectiva es de 199 muestras, la del test es de 200 muestras y el error estimado es $5.0 \pm 3.0\,\%$

2 $\boxed{D}$ En el problema de optimización con restricciones

$$\begin{array}{ll} \text{minimizar} & q(\mathbf{\Theta}), \quad \mathbf{\Theta} \in \mathbb{R}^D \\ \text{sujecto a} & v_i(\mathbf{\Theta}) \leq 0, \quad 1 \leq i \leq k \end{array}$$

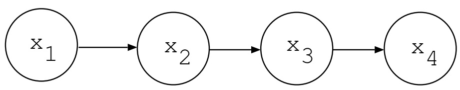
se cumplen las condiciones complementarias de Karush-Kuhn-Tucker $\alpha_i^\star v_i(\mathbf{\Theta}^\star) = 0$ para $1 \leq i \leq k$. Indicar cuál de las siguientes afirmaciones se deduce de ellas:

A) Si para un i, $\alpha_i^\star = 0$, entonces $v_i(\mathbf{\Theta}^\star) \ > \ 0$
B) Si para un i, $\alpha_i^\star = 0$, entonces $v_i(\mathbf{\Theta}^\star) = 0$,
C) Si para un i, $v_i(\mathbf{\Theta}^\star) = 0$, entonces $\alpha_i^\star = 0$
D) Si para un i, $\alpha_i^\star > 0$, entonces $v_i(\mathbf{\Theta}^\star) \ = \ 0$

3 $\boxed{C}$ En la estimación por máxima verosimilitud de los parámetros de una mezcla de K gaussianas de matriz de covarianza común y conocida a partir de N vectores de entrenamiento, los parámetros a estimar son: el vector-media $\boldsymbol{\mu}_k$ y el peso π_k de cada gaussiana, $k, 1 \leq k \leq K$. Identificar cuál de las siguientes afirnaciones es *correcta*:

A) Se puede usar *descenso por gradiente*, ya que los valores de $\boldsymbol{\mu}_k$ no están sujetos a ninguna restricción, lo que hace innecesario recurrir a la técnica de los *multiplicadores de Lagrange*.
B) La solución se obtiene en un paso, utilizando directamente la *optimización lagrangiana* de la verosimilitud de los N vectores de entrenamiento. En este caso, hay un único multiplicador de Lagrange, β, asociado a la restricción de igualdad: $\sum_{k=1}^K \pi_k = 1$.
C) El método más adecuado es el de *esperanza-maximización* (EM), el cual garantiza que que se cumple la restricción $\sum_{k=1}^K \pi_k = 1$. Esto es así gracias a que, en cada iteración de EM, los valores de $\pi_k, 1 \leq k \leq K$, se obtienen como medias de valores de variables latentes, usando una expresión que se deriva analíticamente mediante la técnica de los *multiplicadores de Lagrange* con la restricción indicada.
D) El método más adecuado sería el de *esperanza-maximización* (EM), pero no es posible utilizarlo ya que EM es un método iterativo que no garantiza el cumplimiento de la restricción de igualdad: $\sum_{k=1}^K \pi_k = 1$.

4 $\boxed{B}$ En la red bayesiana lineal



¿cuál de las relaciones siguientes es falsa en general?

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

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

En la siguiente tabla se presenta una muestra de entrenamiento no linealmente separable 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
x_{i1}	1	1	1	1	1
x_{i2}	1	2	3	4	5
Clase	+1	+1	–1	+1	–1
$\alpha_i^\star$	0	3.56	10	10	3.56

a) Obtener la función discriminante lineal correspondiente

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

c) Clasificar la muestra $(1, 4.5)^t$.

a) Pesos de la función discriminante:

$\boldsymbol{\theta}^\star = c_2\ \alpha_1^\star\ \mathbf{x_2} + c_3\ \alpha_4^\star\ \mathbf{x_3} + c_4\ \alpha_5^\star\ \mathbf{x_4} + c_5\ \alpha_7^\star\ \mathbf{x_5}$

$\theta_1^* \ = \ 0.0$

$\theta_2^* \approx -0.67$

Usando el vector soporte $\mathbf{x_5}$ (que verifica la condición : $0 < \alpha_5^* < C$)

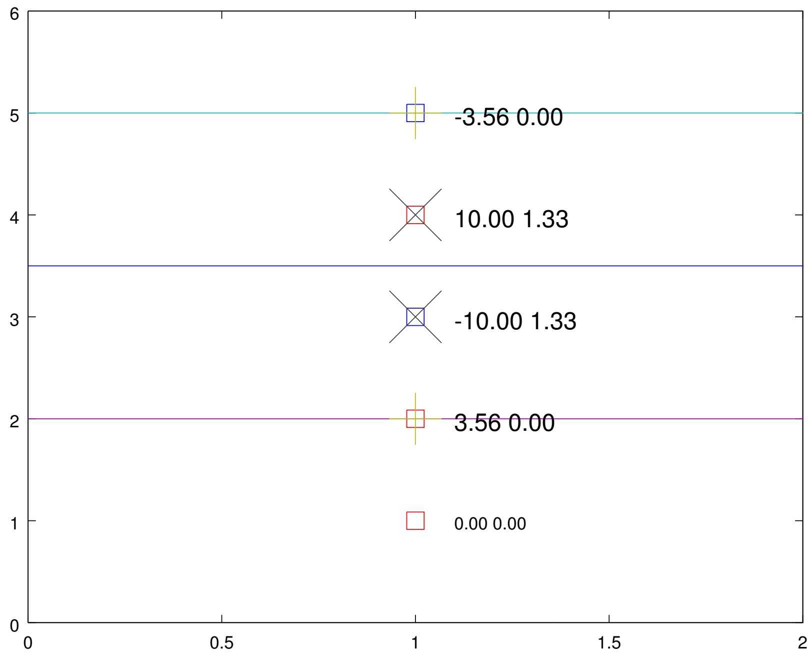
$\theta_0^\star = c_5 - {\boldsymbol{\theta}^\star}^t \mathbf{x_5} \approx 2.33$

b) Frontera de separación y representación gráfica:

Ecuación de la frontera lineal de separación: $2.33 - 0.67\ x_2 \ = \ 0$

Los vectores de entrenamiento son todos los de la tabla. De ellos, los vectores soporte son: $(1, 2)^t, (1, 3)^t, (1, 4)^t, (1, 5)^t$.

Representación gráfica:

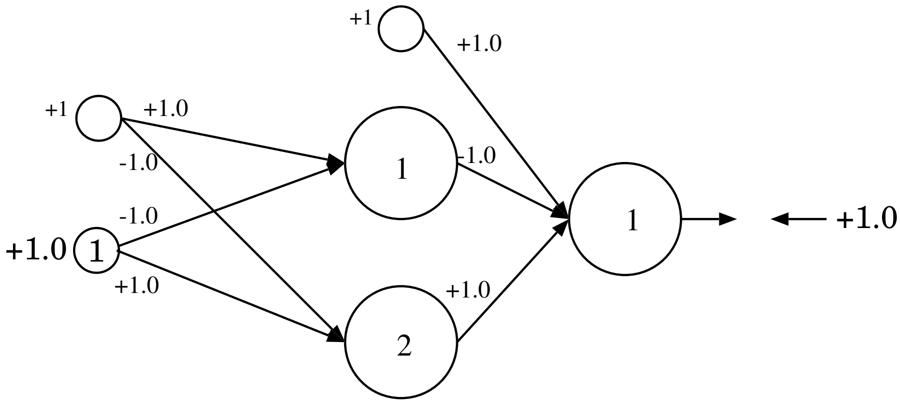


c) Clasificación de la muestra $(1, 4.5)^t$:

El valor de la función discriminante para este vector es: $2.33 - 0.67\ x_2 \ \approx \ -0.67 < 0 \ \Rightarrow$ 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 de los nodos de la capa de salida y de la capa oculta de tipo *tangente hiperbólica*, y factor de aprendizaje $\rho = 0.5$.



Dado un vector de entrada $x = 1$ y su valor deseado de salida $t = +1$, Calcular:

a) las salidas de todas las unidades

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

c) Los nuevos valores de los pesos de las conexiones θ^1_{21}, que va del nodo 1 de entrada al nodo 2 de la capa oculta, y θ^2_{12}, que va del nodo 2 de la capa oculta al nodo de la capa de salida.

a) Las salidas de todas las unidades

$$
\begin{aligned}
\phi^1_1 = \theta^1_{11}\ x_1 + \theta^1_{10} &= 0.0\,; & s^1_1 &= \tfrac{\exp(\phi^1_1)-\exp(-\phi^1_1)}{\exp(\phi^1_1)+\exp(-\phi^1_1)} = 0.0 \\
\phi^1_2 = \theta^1_{21}\ x_1 + \theta^1_{10} &= 0.0\,; & s^1_2 &= \tfrac{\exp(\phi^1_2)-\exp(-\phi^1_2)}{\exp(\phi^1_2)+\exp(-\phi^1_2)} = 0.0 \\
\phi^2_1 = \theta^2_{11}\ s^1_1 + \theta^2_{12}\ s^1_2 + \theta^2_{10} &= 1.0\,; & s^2_1 &= \tfrac{\exp(\phi^2_1)-\exp(-\phi^2_1)}{\exp(\phi^2_1)+\exp(-\phi^2_1)} = 0.76159
\end{aligned}
$$

b) El error en la capa de salida es:
$\delta^2_1 = (t_1 - s^2_1)\ (1 - (s^2_1)^2) = +0.10012$
Los errores en la capa de oculta son:
$\delta^1_1 = \delta^2_1\ \theta^2_{11}\ (1 - (s^1_1)^2) = -0.10012;$ $\quad \delta^1_2 = \delta^2_1\ \theta^2_{12}\ (1 - (s^1_2)^2) = +0.10012$

c) El nuevo peso θ^2_{12} es: $\quad \theta^2_{12} = \theta^2_{12} + \rho\ \delta^2_1\ s^1_2 = (+1.0) + (0.5)\ (+0.10012)\ (0.0) = 1.0$
El nuevo peso θ^1_{21} es: $\quad \theta^1_{21} = \theta^1_{21} + \rho\ \delta^1_2\ x_1 = (+1.0) + (0.5)\ (+0.10012)\ (1.0) = 1.0501$

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

Considerar la red bayesiana $\mathcal{R}$ definida como $P(R, X, Y, Z) = P(R)\ P(X \mid R)\ P(Y \mid R)\ P(Z \mid R)$, cuya variable R toma valores en $\{1, 2, 3\}$ y las variables X, Y, Z, en el conjunto $\{\texttt{"a"},\texttt{"b"},\texttt{"c"}\}$. Las distribuciones de probabilidad asociadas son como sigue:

- $P(R)$ es uniforme: $P(R = 1) = P(R = 2) = P(R = 3)$
- $P(X \mid R)$, $P(Y \mid R)$ y $P(Z \mid R)$ son idénticas y vienen dadas en la tabla T.

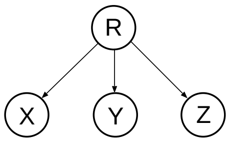
T	"a"	"b"	"c"
1	1/3	0	2/3
2	1/4	1/2	1/4
3	0	3/5	2/5

a) Representar gráficamente la red

b) Obtener una expresión simplificada de $P(X, Y, Z \mid R)$ en función de las distribuciones que definen $\mathcal{R}$ y calcular $P(X = \texttt{"a"}, Y = \texttt{"a"}, Z = \texttt{"a"} \mid R = 1)$

c) Calcular $P(R = 3 \mid X = \texttt{"b"}, Y = \texttt{"b"}, Z = \texttt{"b"})$

a) Representación gráfica de la red:



b) Expresión simplificada de $P(X, Y, Z \mid R)$:

$$P(X, Y, Z \mid R) = \frac{P(R, X, Y, Z)}{P(R)} = P(X \mid R)\ P(Y \mid R)\ P(Z \mid R)$$

$$P(X = \texttt{"a"}, Y = \texttt{"a"}, Z = \texttt{"a"} \mid R = 1) = \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot = \frac{1}{27}$$

c)

$$P(R = 3 \mid X = \texttt{"b"}, Y = \texttt{"b"}, Z = \texttt{"b"}) = \frac{P(R = 3, X = "b", Y = "b", Z = "b")}{P(X = "b", Y = "b", Z = "b")}$$

$$= \frac{P(R = 3)\ P(X = "b" \mid R = 3)\ P(Y = "b" \mid R = 3)\ P(Z = "b" \mid R = 3)}{\sum_{r \in \{1,2,3\}} P(R = r)\ P(X = "b" \mid R = r)\ P(Y = "b" \mid R = r)\ P(Z = "b" \mid R = r)}$$

$$= \frac{\frac{1}{3}\ \frac{3}{5}\ \frac{3}{5}\ \frac{3}{5}}{\frac{1}{3}\ 0\ 0\ 0\ +\ \frac{1}{3}\ \frac{1}{2}\ \frac{1}{2}\ \frac{1}{2}\ +\ \frac{1}{3}\ \frac{3}{5}\ \frac{3}{5}\ \frac{3}{5}} \approx 0.633$$