# Examen de Aprendizaje Automático

ETSINF, Universitat Politècnica de València, 28 de enero de 2015

**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 $\boxed{C}$ Se ha evaluado un sistema de Aprendizaje Automático mediante la técnica de *validación cruzada en B bloques* ("B-fold Cross Validation") con $B = 100$ y utilizando un conjunto de datos de entrenamiento que contiene 1000 muestras. Se han obtenido un total de 22 errores. Indicar cuál de las afirmaciones siguientes es razonable:

A) La talla de entrenamiento efectiva es 990 muestras y el error estimado es $2.2\,\% \pm 0.1\,\%$
B) La talla de entrenamiento efectiva es de 900 muestras y el error estimado es $2.2\,\%$
C) La talla de test efectiva es de 1000 muestras y el error estimado es $2.2\,\% \pm 0.7\,\%$
D) El error estimado es $22\,\% \pm 7\,\%$.

2 $\boxed{A}$ Considerar la siguiente modificación de la función de Widrow y Hoff

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

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

3 $\boxed{A}$ Considerar el aprendizaje mediante máquinas de vectores soportes y márgenes blandos con una muestra de aprendizaje $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ no separable linealmente. Si un multiplicador de Lagrange óptimo $\alpha_j^*$, asociado a la restricción $c_j\ (\boldsymbol{\theta}^t \boldsymbol{x}_d jn + \theta_0)\ \geq\ 1 - \zeta_j$, $1 \leq j \leq N$, es cero, entonces:

A) La muestra $\boldsymbol{x}_j$ está clasificada correctamente.
B) La muestra $\boldsymbol{x}_j$ está mal clasificada.
C) La muestra $\boldsymbol{x}_j$ está clasificada correctamente pero $\boldsymbol{\theta}$ y $\theta_0$ no es canónico con respecto a la muestra.
D) La muestra $\boldsymbol{x}_j$ es un vector soporte.

4 $\boxed{A}$ La distribución de probabilidad conjunta en una red bayesiana de tres nodos $A, B$ y $C$ es $P(A, B, C) = P(C)\ P(A \mid C)\ P(B \mid C)$. Marcar cuál es la afirmación correcta:

A) $P(A, B \mid C)\ =\ P(A \mid C)\ P(B \mid C)$
B) En general, $P(A, B \mid C)\ \neq\ P(A \mid C)\ P(B \mid C)$
C) $P(A, B \mid C)\ =\ P(C \mid A)\ P(C \mid B)$
D) $P(A, B \mid C)\ =\ P(A)\ P(B)$

# 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 | 2 | 4 | 3 | 2 | 4 | 4 |
| $x_{i2}$ | 4 | 1 | 3 | 2 | 4 | 5 | 4 | 3 |
| Clase | +1 | +1 | +1 | +1 | −1 | −1 | +1 | −1 |
| $\alpha_i^\star$ | 3.38 | 0 | 0 | 5.75 | 9.13 | 0 | 10 | 10 |

**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 $(1,1)^t$.

**a)** Pesos de la función discriminante:

$$\boldsymbol{\theta}^\star = c_1\ \alpha_1^\star\ \mathbf{x_1} + c_4\ \alpha_4^\star\ \mathbf{x_4} + c_5\ \alpha_5^\star\ \mathbf{x_5} + c_7\ \alpha_7^\star\ \mathbf{x_7} + c_8\ \alpha_8^\star\ \mathbf{x_8}$$

$$\theta_1^* = (+1)\ (1)\ (3.38) + (+1)\ (4)\ (5.75) + (-1)\ (3)\ (9.13) + (+1)\ (4)\ (10) + (-1)\ (4)\ (10) \approx -1.0$$
$$\theta_2^* = (+1)\ (4)\ (3.38) + (+1)\ (2)\ (5.75) + (-1)\ (4)\ (9.13) + (+1)\ (4)\ (10) + (-1)\ (3)\ (10) = -1.5$$

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

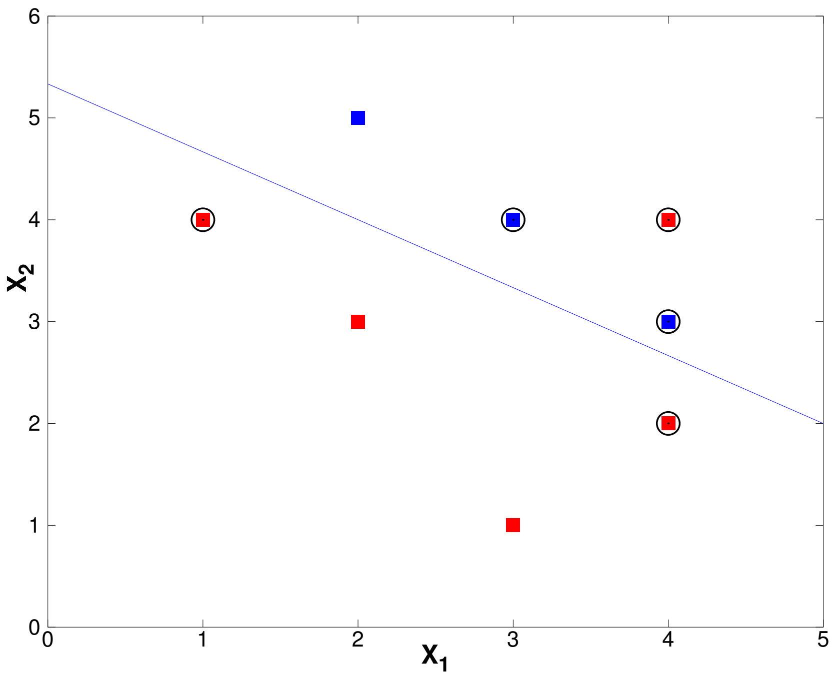
$$\theta_0^\star = c_4 - \boldsymbol{\theta}^{\star t}\mathbf{x_4} = 1 - ((-1.0)\ (4) + (-1.5)\ (2)) = 8.0$$

**b)** Frontera de separación y representación gráfica:

Ecuación de la frontera lineal de separación: $8.0 - 1.0\ x_1 - 1.5\ x_2 = 0 \quad \rightarrow \quad x_2 \approx -0.67\ x_1 + 5.3\,.$

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

Representación gráfica:

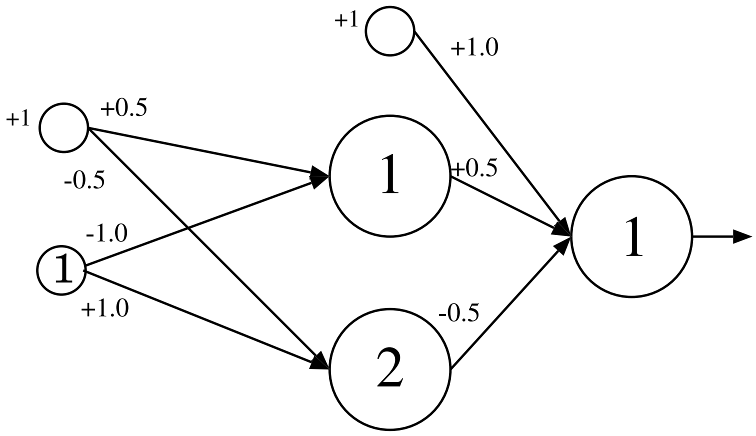



**c)** Clasificación de la muestra $(1,1)^t$:

El valor de la función discriminante para este vector es: $\theta_0^* + \theta_1^*\ 1 + \theta_2^*\ 1 = 5.5 > 0 \Rightarrow$ 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 sigmoid, y factor de aprendizaje $\rho = 1.0$.




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 $\theta^1_{21}$, que va del nodo 1 de entrada al nodo 2 de la capa oculta, y $\theta^2_{12}$, que va del nodo 2 de la capa oculta al nodo de la capa de salida.

a) Las salidas de todas las unidades

$$\begin{aligned}
\phi^1_1 = \theta^1_{11}\ x_1 + \theta^1_{10} &= -0.5\,; & s^1_1 = \tfrac{1}{1+\exp(-\phi^1_1)} &= 0.378 \\
\phi^1_2 = \theta^1_{21}\ x_1 + \theta^1_{10} &= 0.5\,; & s^1_2 = \tfrac{1}{1+\exp(-\phi^1_2)} &= 0.622 \\
\phi^2_1 = \theta^2_{11}\ s^1_1 + \theta^2_{12}\ s^1_2 + \theta^2_{10} &= 0.878\,; & s^2_1 = \tfrac{1}{1+\exp(-\phi^2_1)} &= 0.706
\end{aligned}$$

a) El error en la capa de salida es: $\delta^2_1 = (t_1 - s^2_1)\ s^2_1\ (1 - s^2_1) = -0.354$

Los errores en la capa de oculta son: $\delta^1_1 = \delta^2_1\ \theta^2_{11}\ s^1_1\ (1 - s^1_1) = -0.042$; $\quad \delta^1_2 = \delta^2_1\ \theta^2_{12}\ s^1_2\ (1 - s^1_2) = 0.042$

b) El nuevo peso $\theta^2_{12}$ es: $\theta^2_{12}\ =\ \theta^2_{12} + \rho\ \delta^2_1\ s^1_2\ =\ (-0.5) + (1)\ (-0.354)\ (0.622)\ =\ -0.720$

El nuevo peso $\theta^1_{21}$ es: $\theta^1_{21}\ =\ \theta^1_{21} + \rho\ \delta^1_2\ x_1\ =\ (+1.0) + (1)\ (0.042)\ (1.0)\ =\ 1.042$

# 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.

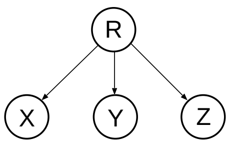
| T | "a" | "b" | "c" |
|---|---|---|---|
| 1 | 1/3 | 0 | 2/3 |
| 2 | 1/4 | 1/2 | 1/4 |
| 3 | 0 | 3/5 | 2/5 |

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)

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

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

$$= \frac{\frac{1}{3}\ \frac{3}{5}\ \frac{3}{5}\ \frac{3}{5}}{\frac{1}{3}\ 0\ 0\ 0 + \frac{1}{3}\ \frac{1}{2}\ \frac{1}{2}\ \frac{1}{2} + \frac{1}{3}\ \frac{3}{5}\ \frac{3}{5}\ \frac{3}{5}} = \frac{216}{341} \approx 0.633$$