Examen de Aprendizaje Automático

ETSINF, Universitat Politècnica de València, 27 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 $\boxed{\text{A}}$ Identifica cuál de las siguientes afirmaciones es *errónea o impropia*:

A) La *teoría de la decisión estadística* es idónea para problemas de *clasificación*, pero es inadecuda para problemas de *regresión* en aprendizaje de funciones.
B) *Clasificación* puede considerarse como un caso particular de *regresión* y los problemas de *sobreajuste* y *sobregeneralización* le afectan de forma similar
C) En el *aprendizaje activo* se asume que un agente externo al sistema se encarga de etiquetar datos de entrenamiento seleccionados por el sistema
D) El *aprendizaje adaptativo* es esencialmente un modo *supervisado* de aprendizaje

2 $\boxed{\text{A}}$ Se ha evaluado un sistema de Aprendizaje Automático mediante un proceso de *validación cruzada en B bloques* ("B-fold Cross Validation") con $B = 8$ y 1000 muestras etiquetadas. En este proceso se han producido 15 errores en total. Indicar cuál de las afirmaciones siguientes es razonable:

A) La talla de entrenamiento efectiva es 875 muestras y el error estimado es $1.5\,\% \pm 0.75\,\%$
B) La talla de entrenamiento efectiva es de 992 muestras y el error estimado es inferior al $2\,\%$
C) La talla de test efectiva es de 125 muestras y el error estimado es $12\,\% \pm 5.7\,\%$.
D) Como en cada bloque de test solo hay 125 muestras, el error es muy variable, pudiéndose estimar como $1.5\,\% \pm 5.7\,\%$.

3 $\boxed{\text{A}}$ Se desea ajustar por mínimos cuadrados la función $f : \mathbb{R}^2 \to R$, definida como: $y = f(\boldsymbol{x}) \stackrel{\text{def}}{=} ax_1^2 + bx_2^2 + cx_1x_2$ a una secuencia de N pares entrada-salida: $S = (\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2) \dots, (\boldsymbol{x}_N, y_N))$. La técnica empleada es minimizar por descenso por gradiente la función de error cuadrático:

$$q(a, b, c) = \sum_{n=1}^{N} (f(\boldsymbol{x}_n) - y_n)^2$$

Identifica la afirmación acertada de entre las siguientes:

A) El vector gradiente es: $2 \sum_{n=1}^{N} (f(\boldsymbol{x}_n) - y_n) \cdot \left(\; x_{n1}^2, \; x_{n2}^2, \; x_{n1}x_{n2} \; \right)^t$
B) El gradiente es $2ax_1 + 2bx_2 + cx_1x_2$
C) El descenso por gradiente solo es aplicable a funciones convexas, pero $q(\cdot)$ no lo es.
D) La técnica de descenso por gradiente no es aplicable en este caso ya que la función a ajustar, $f(\cdot)$, no es lineal.

4 $\boxed{\text{A}}$ Considerar el aprendizaje mediante máquinas de vectores soportes y márgenes blandos con una muestra de aprendizaje $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N$ no separable linealmente. Si un multiplicador de Lagrange óptimo α_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 θ_0 no es canónico con respecto a la muestra.
D) La muestra $\boldsymbol{x}_j$ es un vector soporte.

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
x_{i1}	4	1	2	3	4	2
x_{i2}	1	2	2	2	2	3
Clase	-1	$+1$	-1	$+1$	-1	$+1$
$\alpha_i^\star$	0	3.11	10.00	10.00	3.78	0.67

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:

$$\boldsymbol{\theta}^\star = c_2\ \alpha_2^\star\ \mathbf{x_2} + c_3\ \alpha_3^\star\ \mathbf{x_3} + c_4\ \alpha_4^\star\ \mathbf{x_4} + c_5\ \alpha_5^\star\ \mathbf{x_5} + c_6\ \alpha_6^\star\ \mathbf{x_6}$$

$$\theta_1^* = (+1)\ (3.11)\ (1) + (-1)\ (10)\ (2) + (+1)\ (10)\ (3) + (-1)\ (3.78)\ (4) + (+1)\ (0.67)\ (2) \approx -0.67$$
$$\theta_2^* = (+1)\ (3.11)\ (2) + (-1)\ (10)\ (2) + (+1)\ (10)\ (2) + (-1)\ (3.78)\ (2) + (+1)\ (0.67)\ (3) \approx +0.67$$

Usando el vector soporte $\mathbf{x_5}$ (que verifica la condición : $0 < \alpha_5^* < C = 10$)

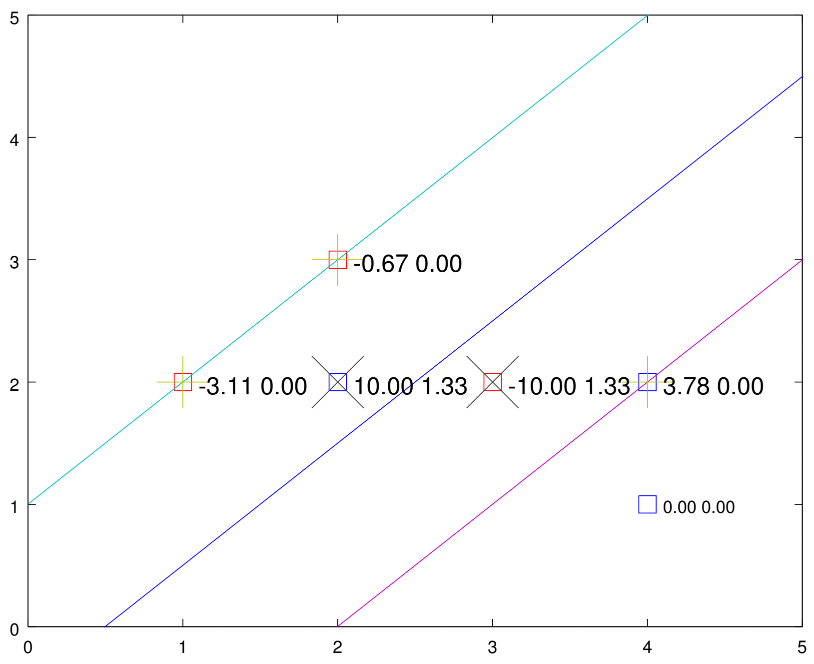
$$\theta_0^\star = c_5 - \boldsymbol{\theta}^{\star t}\mathbf{x_5} = -1 - ((-0.67)\ (4) + (0.67)\ (2)) = 0.34$$

b) Frontera de separación y representación gráfica:

Ecuación de la frontera lineal de separación: $-0.67\ x_1 + 0.67\ x_2 + 0.34 = 0$

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

Representación gráfica:

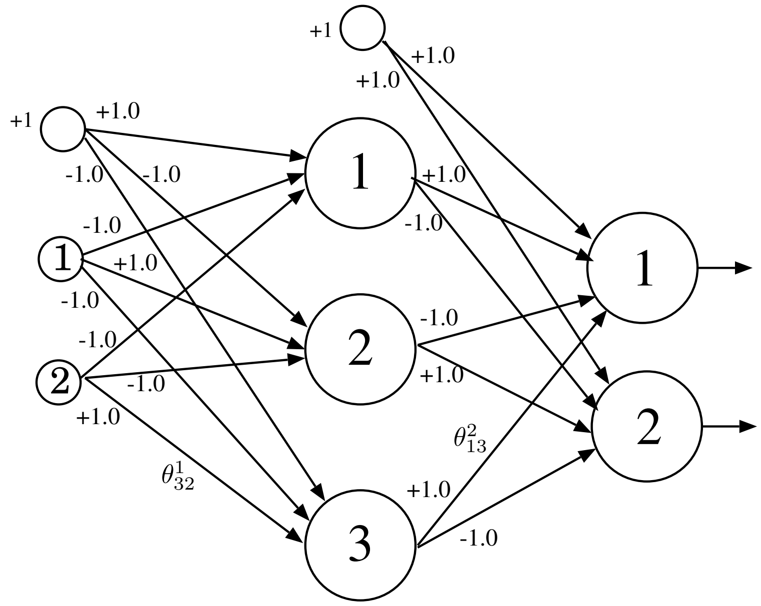


c) Clasificación de la muestra $(1,1)^t$:

El valor de la función discriminante para este vector es: $\theta_0^* + \theta_1^*\ 1 + \theta_2^*\ 1 = +0.34 > 0 \Rightarrow$ 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 *sigmoide*, y factor de aprendizaje $\rho = 1.0$.



Dado un vector de entrada $\boldsymbol{x}^t = (1, -1)$, las salidas de las unidades de la capa de salida son $s^2_1 = 0.740$ y $s^2_2 = 0.722$ y las de las unidades ocultas son $s^1_1 = 0.731$, $s^1_2 = 0.731$ y $s^1_3 = 0.047$. Si el valor deseado de salida es $\boldsymbol{t}^t = (1, 0)$, calcular:

a) Los correspondientes errores en los nodos de la capa de salida y en los nodos de la capa oculta.

b) Los nuevos valores de los pesos de las conexiones θ^1_{32} y θ^2_{13}.

a) Los errores en la capa de salida son:

$$\delta^2_1 = (t_1 - s^2_1)\ s^2_1\ (1 - s^2_1) = +0.050;$$
$$\delta^2_2 = (t_2 - s^2_2)\ s^2_2\ (1 - s^2_2) = -0.145$$

Los errores en la capa de oculta son:

$$\delta^1_1 = (\delta^2_1\ \theta^2_{11} + \delta^2_2\ \theta^2_{21})\ s^1_1\ (1 - s^1_1) = +0.038;$$
$$\delta^1_2 = (\delta^2_1\ \theta^2_{12} + \delta^2_2\ \theta^2_{22})\ s^1_2\ (1 - s^1_2) = -0.038;$$
$$\delta^1_3 = (\delta^2_1\ \theta^2_{13} + \delta^2_2\ \theta^2_{23})\ s^1_3\ (1 - s^1_3) = +0.009$$

b) El nuevo peso θ^2_{13} es: $\theta^2_{13} = \theta^2_{13} + \rho\ \delta^2_1\ s^1_3 = (+1.0) + (1)\ (0.050)\ (0.047) = 1.002;$

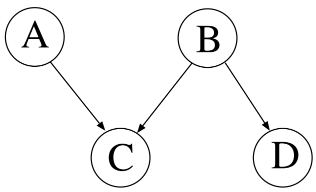
El nuevo peso θ^1_{32} es: $\theta^1_{32} = \theta^1_{32} + \rho\ \delta^1_3\ x_2 = (+1.0) + (1)\ (0.009)\ (-1.0) = 0.991$

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

Las variables aleatorias A, B, C, D toman valores en el conjunto $\{0, 1\}$. La distribución de probabilidad conjunta de estas variables viene dada por $P(A, B, C, D) = P(A)\ P(B)\ P(C \mid A, B)\ P(D \mid B)$, donde las distribuciones de probabilidad asociadas son:

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

Y cuya representación gráfica es



a) Calcular la probabilidad conjunta para $A = 1,\ B = 1, C = 1$ y $D = 1$.

b) Obtener una expresión simplificada de $P(A \mid B, C, D)$ y calcular su valor para $A = 1$ cuando $B = 1, C = 1$ y $D = 1$.

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

a) Calcular la probabilidad conjunta para $A = 1,\ B = 1, C = 1$ y $D = 1$.

$$P(A=1, B=1, C=1, D=1) = P(A=1)\ P(B=1)\ P(C \mid A=1, B=1)\ P(D \mid B=1) = 0.3\ 0.4\ 0.4\ 0.7 = 0.0336$$

b) Obtener una expresión simplificada de $P(A \mid B, C, D)$ y calcular su valor para $A = 1$ cuando $B = 1, C = 1$ y $D = 1$.

$$
\begin{aligned}
P(A \mid B, C, D) &= \frac{P(A, B, C, D)}{P(B, C, D)} = \frac{P(A)\ P(B)\ P(C \mid A, B)\ P(D \mid B)}{P(B)\ P(D \mid B)\ \sum_a P(A=a)\ P(C \mid A=a, B)}\\
&= \frac{P(A)\ P(C \mid A, B)}{P(A=0)\ P(C \mid A=0, B) + P(A=1)\ P(C \mid A=1, B)}
\end{aligned}
$$

$$P(A=1 \mid B=1, C=1, D=1) = \frac{0.3 \cdot 0.4}{0.7 \cdot 0.2\ +\ 0.3 \cdot 0.4} = 0.4615$$

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

$a^\star = \arg\max_{a \in \{0,1\}} P(A = a \mid B = 1, C = 1, D = 1)$

$P(A = 0 \mid B = 1, C = 1, D = 1) = 1 - 0.4615 = 0.5385 \geq 0.4615 \Rightarrow$ valor óptimo es $A = 0$