Examen de recuperación de Aprendizaje Automático

ETSINF, Universitat Politècnica de València, 28 de enero de 2014

Apellidos: | | **Nombre:** | |

Cuestiones (2 puntos; tiempo estimado: 30 minutos)

Marca cada recuadro con una única opción de entre las dadas.

1 $\boxed{\text{B}}$ 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 de entrenamiento 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 2 %
C) La talla de entrenamiento efectiva es de 900 muestras y el error es del 20 %
D) La talla de entrenamiento efectiva es de 1000 muestras y y el error es del 20 %.

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

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

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

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

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

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

3 $\boxed{\text{A}}$ Un clasificador implementado mediante una red neuronal con L capas ocultas en la que todas las funciones de activación son lineales, es equivalente a:

A) un clasificador basado en funciones discriminantes lineales
B) un clasificador basado en funciones discriminantes lineales generalizadas cuyas fronteras de decisión son no-lineales
C) un clasificador implementado mediante una red neuronal con $L - 1$ capas ocultas
D) para que la red pueda usarse para clasificación, al menos las funciones de activación de la capa de salida han de ser no-lineales.

4 $\boxed{\text{A}}$ Sea $\mathcal{C}$ un conjunto de variables aleatorias (VA). Un concepto importante en el que se basan las técnicas de redes bayesianas es:

A) todas las probabilidades condicionales e incondicionales en las que participan VA's de $\mathcal{C}$ se pueden obtener mediante las reglas basicas de inferencia estadística, a partir de la probabilidad conjunta de todas las VA de $\mathcal{C}$.
B) la probabilidad incondicional de una VA $a \in \mathcal{C}$ solo depende de las VA's de los nodos con los que está conectado el nodo de a.
C) el grafo que representa las VA's de $\mathcal{C}$ ha de ser acíclico y conexo.
D) ha de existir indepenencia condicional entre al menos dos VA's de $\mathcal{C}$.

Problema 1 (4 puntos; tiempo estimado: 30 minutos)

En la siguiente tabla se muestra una muestra de entrenamiento linealmente separable en $\mathbb{R}^2$ y los correspondientes multiplicadores de Lagrange óptimos obtenidos al entrenar una máquina de vectores soporte con esta muestra:

i	1	2	3	4	5	6	7	8	9
x_{i1}	1	1	2	2	4	3	3	5	5
x_{i2}	2	3	2	1	1	6	3	1	2
Clase	+1	+1	+1	+1	-1	-1	-1	-1	-1
$\alpha_i^\star$	0	0	1.111	0	0.222	0	0.889	0	0

a) Obtener la función discriminante lineal correspondiente

b) Dibujar la función discriminante lineal

c) Clasificar la muestra $(5,5)^t$.

a) Pesos de la función discriminante:

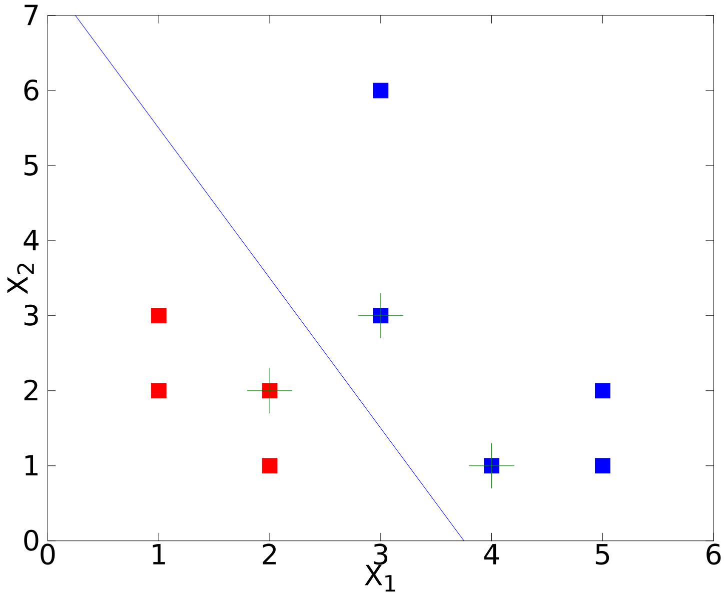
$\boldsymbol{\theta}^\star = c_3\alpha_3^\star \mathbf{x_3} + c_5\alpha_5^\star \mathbf{x_5} + c_7\alpha_7^\star \mathbf{x_7}$

$\theta_1^* = (+1)\ (2)\ (1.111) + (-1)\ (4)\ (0.222) + (-1)\ (3)\ (0.889) = -1.333$
$\theta_2^* = (+1)\ (2)\ (1.111) + (-1)\ (1)\ (0.222) + (-1)\ (3)\ (0.889) = -0.667$

Usando el vector soporte $\mathbf{x_3}$:

$\theta_0^\star = c_3 - \boldsymbol{\theta}^{\star t}\mathbf{x_3} = 1 - (-1.333 \cdot 2 - 0.667 \cdot 2) = 5.000$

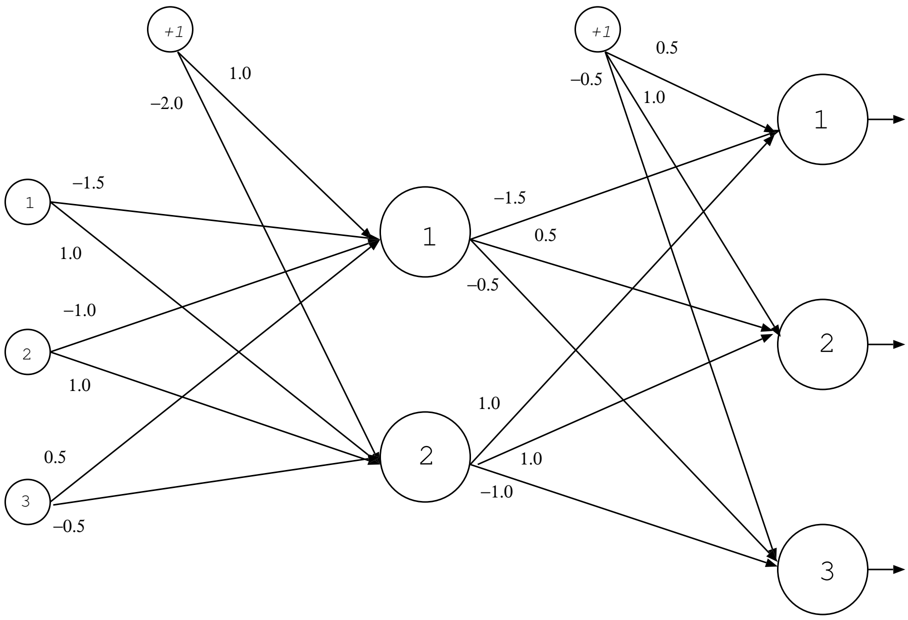
b) Dibujo de la función discriminante:



c) Clasificación de la muestra $(5,5)^t$:
$\theta_0^* + \theta_1^*\ 5 + \theta_2^*\ 5 = -5 < 0 \Rightarrow$ clase -1 (o la 2)

Problema 2 (4 puntos; tiempo estimado: 30 minutos)

Sea $\mathcal{P}$ el perceptrón multicapa de la figura, que se pretende utilizar para resolver un problema de regresión.



Se asume que la función de activación de los nodos de la capa de salida es lineal y la de los nodos de la capa oculta es de tipo escalón, definida como:

$$g_E(z) = \mathrm{sgn}(z) = \begin{cases} -1 & \text{if } z < 0 \\ +1 & \text{if } z \geq 0 \end{cases}$$

Sea $\mathbf{x} = (2.0,\ 2.0,\ 0.0)^t$ un vector de entrada y sean $-5.0,\ 1.0, -2.0,$ los valores deseados para dicha muestra en los nodos de salida $1, 2, 3$, respectivamente. Calcular:

a) los tres valores de salida de $\mathcal{P}$ cuando se observa $\mathbf{x}$ en la entrada,

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

a) Valores de la capa oculta:

$s_1^1 = -1; s_2^1 = 1$

Valores de la capa de salida:

$s_1^2 = 3; s_2^2 = 1.5; s_3^2 = -1.0$

b) Los errores en la capa de salida son:

$\delta_1^2 = t_1 - s_1^1 = -8.0; \qquad \delta_2^2 = t_2 - s_2^1 = -0.5; \qquad \delta_3^2 = t_3 - s_3^1 = -1.0$

La función en escalón no tiene derivada en 0, y vale 0 en cualquier otro punto, por lo tanto los errores en la capa oculta no se pueden calcular cuando la función discriminante vale 0. Como las funciones discriminates en la capa oculta para el vector son -4 y 2, el error en la capa oculta sería 0. No obstante, la propagación del error de la capa de salida a la oculta sin contar la derivada de la función de activación es:

nodo 1: $\delta_1^2\ \theta_{11}^2 + \delta_2^2\ \theta_{21}^2 + \delta_3^2\ \theta_{31}^2 = 12.25$; $\qquad$ nodo 2: $\delta_1^2\ \theta_{12}^2 + \delta_2^2\ \theta_{22}^2 + \delta_3^2\ \theta_{32}^2 = -7.5$