Examen de Aprendizaje Automático

ETSINF, Universitat Politècnica de València, 10 de enero de 2018

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 $\boxed{C}$ Sea S un conjunto de 1000 datos supervisados o etiquetados. Para el diseño de un sistema de reconocimiento de formas, se utilizan datos de S tanto para aprender los parámetros del modelo de reconocimiento, $\mathcal{M}$, como para estimar el error de reconocimiento dicho modelo. Para ello se ha utilizado el método de *validación cruzada en* 10 *bloques*, obteniéndose 4, 0, 4, 3, 4, 0, 3, 5, 4, 3 errores. Indicar cual de las siguientes afirmaciones es *correcta*.

A) El test efectivo es de 100 muestras, el conjunto de entrenamiento efectivo es de 900 muestras, el error empírico $\hat{p}$ es del 3 % y el intervalo de confianza es de 3.0 ± 0.1 %

B) El test efectivo es de 1000 muestras, el conjunto de entrenamiento efectivo es de 1000 muestras, el error empírico $\hat{p}$ es del 3 % y el intervalo de confianza es de 3 ± 1 %

C) El test efectivo es de 1000 muestras, el conjunto de entrenamiento efectivo es de 900 muestras, el error empírico $\hat{p}$ es del 3 % y el intervalo de confianza es de 3 ± 1 %

D) El test efectivo es de 1000 muestras, el conjunto de entrenamiento efectivo es de 900 muestras, el error empírico $\hat{p}$ es del 6 % y el intervalo de confianza es de 6 ± 1 %

2 $\boxed{A}$ Sea $S = \{(\boldsymbol{x}_1, c_1), \ldots, (\boldsymbol{x}_N, c_N)\}$, $\boldsymbol{x}_n \in \mathbb{R}^D$, $c_n \in \{+1, -1\}$ una muestra de entrenamiento. El algoritmo perceptrón muestra a muestra ("online") trata de encontrar una solución $\hat{\boldsymbol{\theta}}$ que satisfaga el sistema de N inecuaciones. Si queremos una solución con *margen*, esto es encontrar una solución $\hat{\boldsymbol{\theta}}$ que satisfaga el sistema de N inecuaciones: $c_n\ \boldsymbol{\theta}^t\boldsymbol{x}_n \geq b,\ \ 1 \leq n \leq N$, para $b \in \mathbb{R}^{\geq 0}$. Indicar cuál de los siguientes algoritmos implementa la solución con margen, partiendo de que $\boldsymbol{\theta}(1)\ =\ \text{arbitrario}$:

A) $\boldsymbol{\theta}(k+1) = \begin{cases} \boldsymbol{\theta}(k) & c(k)\ \boldsymbol{\theta}^t\boldsymbol{x}(k) \geq b \\ \boldsymbol{\theta}(k) + \rho_k\ c(k)\ \boldsymbol{x}(k) & c(k)\ \boldsymbol{\theta}^t\boldsymbol{x}(k) < b \end{cases}$

B) $\boldsymbol{\theta}(k+1) = \begin{cases} \boldsymbol{\theta}(k) & c(k)\ \boldsymbol{\theta}^t\boldsymbol{x}(k) < b \\ \boldsymbol{\theta}(k) + \rho_k\ c(k)\ \boldsymbol{x}(k) & c(k)\ \boldsymbol{\theta}^t\boldsymbol{x}(k) \geq b \end{cases}$

C) $\boldsymbol{\theta}(k+1) = \begin{cases} b\ \boldsymbol{\theta}(k) & c(k)\ \boldsymbol{\theta}^t\boldsymbol{x}(k) \geq 0 \\ b\ \boldsymbol{\theta}(k) + \rho_k\ c(k)\ \boldsymbol{x}(k) & c(k)\ \boldsymbol{\theta}^t\boldsymbol{x}(k) < 0 \end{cases}$

D) $\boldsymbol{\theta}(k+1) = \begin{cases} \boldsymbol{\theta}(k) & c(k)\ \boldsymbol{\theta}^t\boldsymbol{x}(k) \geq b \\ \boldsymbol{\theta}(k) + \rho_k\ b\ c(k)\ \boldsymbol{x}(k) & c(k)\ \boldsymbol{\theta}^t\boldsymbol{x}(k) < b \end{cases}$

3 $\boxed{A}$ Las siguientes afirmaciones se refieren al método Esperanza Maximización (EM) aplicado a una muestra de entrenamiento S. Identificar cuál de ellas es *correcta*:

A) EM es útil para estimar valores máximo-verosímiles de los parámetros de modelos estadísticos a partir de S cuando hay variables latentes o ocultos.
B) EM no se puede aplicar en la estimación de valores máximo-verosímiles de los parámetros de modelos estadísticos a partir de S cuando no hay variables latentes o ocultas.
C) La rapidez de convergencia de EM puede mejorarse eligiendo un factor de aprendizaje adecuado para S.
D) La rapidez de convergencia de EM siempre puede mejorarse inicializando los parámetros a cero.

4 $\boxed{D}$ En la red bayesiana cuya distribución conjunta es $P(x_1, x_2, x_3, x_4) = P(x_1)\ P(x_2 \mid x_1)\ P(x_3 \mid x_1)\ P(x_4 \mid x_2, x_3)$ ¿cuál de las relaciones siguientes es *falsa* en general?

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

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

Para el aprendizaje de una máquina de vectores soporte se dispone de la siguiente muestra de entrenamiento linealmente separable:

$$S = \{((6,6), -1), ((6,2), -1), ((1,6), +1), ((2,2), +1), ((2,3), +1), ((4,2), -1), ((3,5), -1), ((3,4), -1), ((1,2), +1), ((2,1), +1)\}$$

Los multiplicadores de Lagrange óptimos son: $\boldsymbol{\alpha}^* = (0.0,\ 0.0,\ 0.25,\ 0.0,\ 1.0,\ 0.0,\ 0.0,\ 1.25,\ 0.0,\ 0.0)^t$.

a) Obtener la función discriminante lineal correspondiente

b) Obtener la ecuación de la frontera de decisión entre clases y representar gráficamente los puntos de entrenamiento y dicha frontera.

c) Calcular el margen óptimo

d) Clasificar la muestra $(3, 1)$.

a) Función discriminante lineal (FDL):

- Vector de pesos:

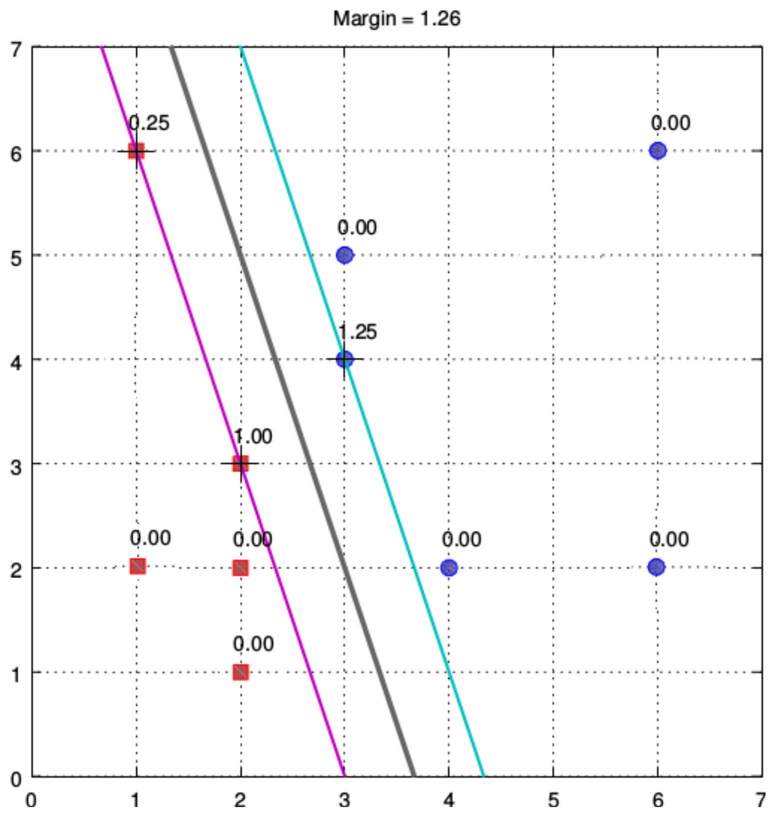
$$\begin{aligned} \theta_1^\star &= +1 \cdot 0.25 \cdot 1\ +1 \cdot 1.0 \cdot 2\ -1 \cdot 1.25 \cdot 3 = -1.5 \\ \theta_2^\star &= +1 \cdot 0.25 \cdot 6\ +1 \cdot 1.0 \cdot 3\ -1 \cdot 1.25 \cdot 4 = -0.5 \end{aligned}$$

- Peso umbral (con el tercer vector de entrenamiento): $\theta_0^* = (+1) - (-1.5 \cdot 1 - 0.5 \cdot 6) = 5.5$
- FDL: $\phi(\boldsymbol{x}) = -1.5\ x_1 - 0.5\ x_2 + 5.5$

b) Ecuación de la frontera de decisión:

$$-1.5\ x_1 - 0.5\ x_2 + 5.5 = 0 \quad \Rightarrow \quad x_2 = -3x_1 + 11$$

Representación gráfica:



c) Margen óptimo:

$$\frac{2}{\|\boldsymbol{\theta}^\star\|} = \frac{2}{\sqrt{(-1.5)^2 + (-0.5)^2}} = 1.265$$

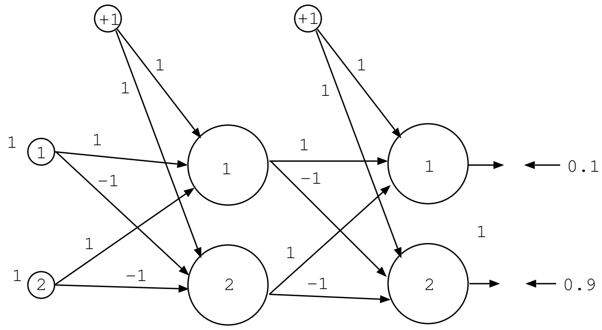
Alternativamente:

$$2\left(\sum_{n \in \mathcal{V}} \alpha_n^\star\right)^{-1/2} = \frac{2}{\sqrt{0.25 + 1.0 + 1.25}} = 1.265$$

d) Clasificación de la muestra $(3, 1)$: $\phi(3, 1)\ = -1.5 \cdot 3 - 0.5 \cdot 1 + 5.5\ = 0.5\ >\ 0 \quad \Rightarrow \quad \text{clase} = +1$

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

La red hacia adelante ("feedforward") 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 $\rho = 1.0$.



Dados los pesos iniciales indicados en la figura, un vector de entrada $\boldsymbol{x}^t = (1,1)$ y su valor deseado de salida $t = (0.1, 0.9)$, Calcular:

a) las salidas de todas las unidades

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

c) Los nuevos valores de los pesos de las conexiones al nodo 2 de la capa oculta.

a) Las salidas de todas las unidades

Capa oculta

$$\phi_1^1 = \theta_{10}^1 + \theta_{11}^1\ x_1 + \theta_{12}^1\ x_2 = 3 \qquad s_1^1 = \frac{1}{1+\exp(-\phi_1^1)} = 0.953$$

$$\phi_2^1 = \theta_{20}^1 + \theta_{21}^1\ x_1 + \theta_{22}^1\ x_2 = -1 \qquad s_2^1 = \frac{1}{1+\exp(-\phi_2^1)} = 0.269$$

Capa de salida

$$\phi_1^2 = \theta_{10}^2 + \theta_{11}^2\ s_1^1 + \theta_{12}^2\ s_2^1 = 2.221 \qquad s_1^2 = \frac{1}{1+\exp(-\phi_1^2)} = 0.902$$

$$\phi_2^2 = \theta_{20}^2 + \theta_{21}^2\ s_1^1 + \theta_{22}^2\ s_2^1 = -0.222 \qquad s_2^2 = \frac{1}{1+\exp(-\phi_2^2)} = 0.445$$

b) Los errores en la capa de salida son:

$$\delta_1^2 = (t_1 - s_1^2)\ s_1^2\ (1 - s_1^2) = -0.0708 \qquad \delta_2^2 = (t_2 - s_2^2)\ s_2^2\ (1 - s_2^2) = +0.1124$$

Los errores en la capa de oculta son:

$$\delta_1^1 = (\delta_1^2\ \theta_{11}^2 + \delta_2^2\ \theta_{21}^2)\ s_1^1\ (1 - s_1^1) = -0.0082 \qquad \delta_2^1 = (\delta_1^2\ \theta_{12}^2 + \delta_2^2\ \theta_{22}^2)\ s_2^1\ (1 - s_2^1) = -0.0360$$

c) Los nuevos pesos del nodo 2 son:

$$\theta_{20}^1 = \theta_{20}^1 + \rho\ \delta_2^1\ (+1) = 0.964$$

$$\theta_{21}^1 = \theta_{21}^1 + \rho\ \delta_2^1\ x_1 = -1.036$$

$$\theta_{22}^1 = \theta_{22}^1 + \rho\ \delta_2^1\ x_2 = -1.036$$

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

Considerar la red bayesiana $\mathcal{R}$ definida como $P(W,X,Y,Z) = P(W)\ P(X \mid W)\ P(Y \mid W)\ P(Z \mid X)$, cuyas variables aleatorias, W, X, Y, Z, toman valores en en el conjunto {a,b,c}. Las distribuciones de probabilidad asociadas son como sigue:

- $P(W)$ es uniforme: $P(W = \mathrm{a}) = P(W = \mathrm{b}) = P(W = \mathrm{c})$,
- $P(X \mid W)$, $P(Y \mid W)$ y $P(Z \mid X)$ vienen dadas en las siguientes tablas:

$P(x \mid w)$	x: a	b	c
w: a	1/2	0	1/2
b	1/4	1/2	1/4
c	1/5	3/5	1/5

$P(y \mid w)$	y: a	b	c
w: a	1/3	0	2/3
b	1/4	1/2	1/4
c	0	3/5	2/5

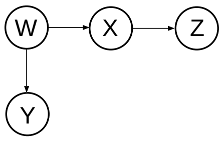
$P(z \mid x)$	z: a	b	c
x: a	1/3	0	2/3
b	1/4	1/2	1/4
c	0	3/5	2/5

a) Representar gráficamente la red

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

c) Obtener una expresión simplificada de $P(W \mid X,Y,Z)$, y calcular $P(W = \mathrm{c} \mid X = \mathrm{b},\ Y = \mathrm{b},\ Z = \mathrm{b})$

a) Representación gráfica de la red:



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

$$\begin{aligned} P(X,Y,Z \mid W) &= \frac{P(W,X,Y,Z)}{P(W)} = \frac{\cancel{P(W)}\ P(X \mid W)\ P(Y \mid W)\ P(Z \mid X)}{\cancel{P(W)}} \\ &= P(X \mid W)\ P(Y \mid W)\ P(Z \mid X) \end{aligned}$$

$$P(X = \mathrm{b},\ Y = \mathrm{b},\ Z = \mathrm{b} \mid W = \mathrm{c}) = \frac{3}{5} \cdot \frac{3}{5} \cdot \frac{1}{2} = \frac{9}{50} = 0.18$$

c) Expresión simplificada de $P(W \mid X,Y,Z)$:

$$\begin{aligned} P(W \mid X,Y,Z) &= \frac{P(W,X,Y,Z)}{P(X,Y,Z)} = \frac{P(W)\ P(X \mid W)\ P(Y \mid W)\ \cancel{P(Z \mid X)}}{\sum_{w \in \{\mathrm{a,b,c}\}} P(W = w) P(X \mid W = w)\ P(Y \mid W = w)\ \cancel{P(Z \mid X)}} \\ &= \frac{\cancel{(1/3)}\, P(X \mid W)\ P(Y \mid W)}{\cancel{(1/3)} \sum_{w \in \{\mathrm{a,b,c}\}} P(X \mid W = w)\ P(Y \mid W = w)} = \frac{P(X \mid W)\ P(Y \mid W)}{\sum_{w \in \{\mathrm{a,b,c}\}} P(X \mid W = w)\ P(Y \mid W = w)} \end{aligned}$$

$$\begin{aligned} P(W = \mathrm{c} \mid X = \mathrm{b},\ Y = \mathrm{b},\ Z = \mathrm{b}) &= \frac{P(X = \mathrm{b} \mid W = \mathrm{c})\ P(Y = \mathrm{b} \mid W = \mathrm{c})}{\sum_{w \in \{\mathrm{a,b,c}\}} P(X = \mathrm{b} \mid W = w)\ P(Y = \mathrm{b} \mid W = w)} \\ &= \frac{(3/5)^2}{0 + (1/2)^2 + (3/5)^2} = \frac{36}{61} \approx 0.59 \end{aligned}$$