Examen de Aprendizaje Automático

ETSINF, Universitat Politècnica de València, 25 de enero de 2017

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 $\boxed{\text{C}}$ Sea S un conjunto de 1000 datos supervisados o etiquetados. En el diseño de un sistema de reconocimiento de formas se utiliza el método de *validación cruzada en 10 bloques* y se obtienen los errores siguientes: (2,4,2,3,0,7,4,4,4,1). Indicar cual de las siguientes afirmaciones es incorrecta.

A) El error estimado es $p_e = 3.1\,\%$

B) El intervalo de confianza al 95 % es ± 0.9

C) El test efectivo es de 100 muestras

D) El tamaño de entrenamiento efectivo es de 900 muestras

2 $\boxed{\text{D}}$ En el problema de aprendizaje de modelos probabilísticos con variables observables $\boldsymbol{x}_n$ y latentes $\boldsymbol{z}_n$

$$L_S(\boldsymbol{\Theta}) \;=\; \sum_{n=1}^{N} \log \left(\sum_{\boldsymbol{z}_n} P(\boldsymbol{x}_n, \boldsymbol{z}_n \mid \boldsymbol{\Theta}) \right)$$

se utiliza la técnica esperanza-maximización (EM). Indicar cuál de las siguientes afirmaciones es cierta:

A) En cada iteración, el paso E consiste en obtener una estimación de todas las variables $\boldsymbol{x}_n$ y $\boldsymbol{z}_n$, y el paso M, obtener los parámetros óptimos de $\boldsymbol{\Theta}$ utilizando la estimación de las variables $\boldsymbol{x}_n$ y $\boldsymbol{z}_n$ obtenidas en el paso E.

B) En cada iteración, el paso E consiste en obtener los valores de las variables latentes $\boldsymbol{z}_n$ que maximizan la función objetivo $L_S(\boldsymbol{\Theta})$, y el paso M, obtener los parámetros óptimos de $\boldsymbol{\Theta}$ utilizando la estimación de las variables $\boldsymbol{z}_n$ obtenidas en el paso E.

C) En cada iteración, el paso E consiste en obtener los valores de todas las variables $\boldsymbol{x}_n$ y $\boldsymbol{z}_n$ que maximizan la función $L_S(\boldsymbol{\Theta})$,, y el paso M, obtener los parámetros óptimos de $\boldsymbol{\Theta}$ utilizando la estimación de las variables $\boldsymbol{x}_n$ y $\boldsymbol{z}_n$ obtenidas en el paso E.

D) En cada iteración, el paso E consiste en obtener una estimación de las variables latentes $\boldsymbol{z}_n$, y el paso M, obtener los parámetros óptimos de $\boldsymbol{\Theta}$ utilizando la estimación de las variables $\boldsymbol{z}_n$ obtenidas en el paso E.

3 $\boxed{\text{B}}$ 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 α_j^*, asociado a la restricción $c_j\ (\boldsymbol{\theta}^t \boldsymbol{x}_d + \theta_0) \;\geq\; 1 - \zeta_j$, $1 \leq j \leq N$, es cero, entonces la muestra $\boldsymbol{x}_j$ está bien clasificada pero ¿cuál de las siguientes afirmaciones es falsa?:

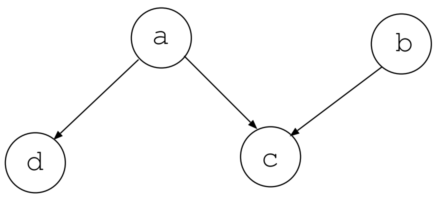
A) $\zeta_j = 0$

B) Se produce un error de margen

C) No hay error de margen

D) La muestra $\boldsymbol{x}_j$ no es un vector soporte

4 $\boxed{\text{A}}$ En la red bayesiana



¿cuál de las relaciones siguientes es verdadera?

A) $P(a, b) \;=\; P(a)\ P(b)$

B) $P(a, d) \;=\; P(a)\ P(d)$

C) $P(a, b \mid d) \;=\; P(a \mid d)\ P(b \mid d)$

D) $P(a, c \mid b) \;=\; P(a \mid b)\ P(c \mid b)$

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

La siguiente tabla contine 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
x_{i1}	1	1	2	1	4	3	6
x_{i2}	6	3	3	1	5	4	2
Clase	+1	+1	+1	+1	-1	-1	-1
$\alpha_i^\star$	0.25	0	1.0	0	0	1.25	0

a) Obtener la función discriminante lineal correspondiente

b) Calcular el margen óptimo

b) Obtener la ecuación de la frontera de separación entre clases y representarla gráficamente, junto con los datos de entrenamiento.

c) Clasificar la muestra $(2,6)^t$.

a) Función discriminante lineal (FDL):

- Vector de pesos:

$$\theta_1^\star = +1 \cdot 0.25 \cdot 1 \ +1 \cdot 1.0 \cdot 2 \ -1 \cdot 1.25 \cdot 3 = -1.5$$
$$\theta_2^\star = +1 \cdot 0.25 \cdot 6 \ +1 \cdot 1.0 \cdot 3 \ -1 \cdot 1.25 \cdot 4 = -0.5$$

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

c) Margen óptimo:

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

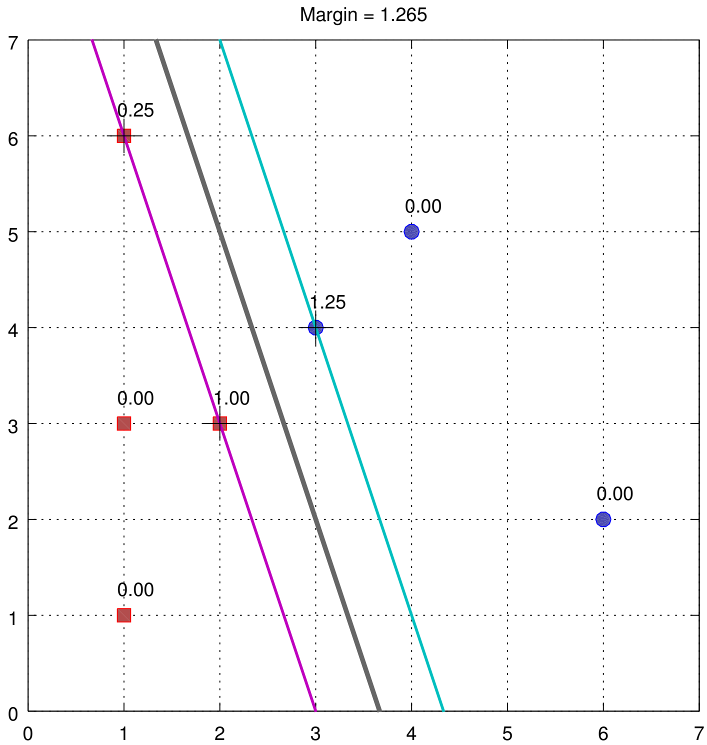
Alternativamente:

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

b) Ecuación de la frontera de decisión:

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

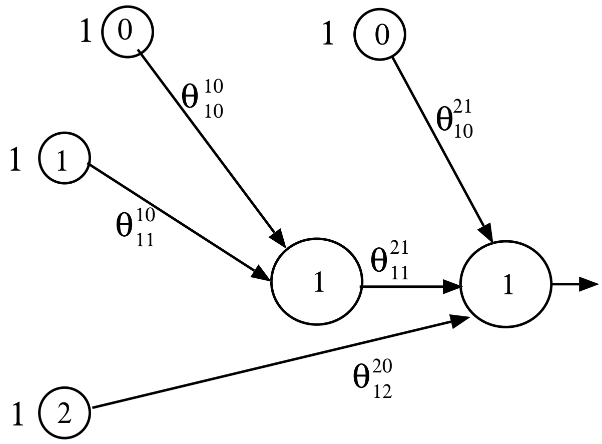
Representación gráfica:



d) Clasificación de la muestra $(2,6)$: $\phi(2,6) = -1.5 \cdot 2 - 0.5 \cdot 6 + 5.5 = -0.5 < 0 \Rightarrow$ 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 *tangente hiperbólica*, y factor de aprendizaje $\rho = 1.0$.



Dados unos pesos iniciales $\theta^{10}_{10} = \theta^{10}_{11} = \theta^{20}_{12} = \theta^{21}_{10} = \theta^{21}_{11} = 1.0$, un vector de entrada $\boldsymbol{x}^t = (1, 1)$ 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 el de la capa oculta.

c) Los nuevos valores de los pesos de las conexiones

a) Las salidas de todas las unidades

$\phi^1_1 = \theta^{10}_{10} + \theta^{10}_{11}\ x_1 = 2.0$ $\qquad s^1_1 = \dfrac{\exp(\phi^1_1) - \exp(-\phi^1_1)}{\exp(\phi^1_1) + \exp(-\phi^1_1)} = 0.96402$

$\phi^2_1 = \theta^{21}_{10} + \theta^{20}_{12}\ x_2 + \theta^{21}_{11}\ s^1_1 = 2.96402$ $\qquad s^2_1 = \dfrac{\exp(\phi^2_1) - \exp(-\phi^2_1)}{\exp(\phi^2_1) + \exp(-\phi^2_1)} = 0.99469$

b) El error en la capa de salida es:

$\delta^2_1 = (t_1 - s^2_1)\ (1 - (s^2_1)^2) = 0.0000562$

El error en la capa de oculta es:

$\delta^1_1 = (\delta^2_1\ \theta^{21}_{11})\ (1 - (s^1_1)^2) = 0.000004$

c) Los nuevos pesos son:

$\theta^{21}_{10} = \theta^{21}_{10} + \rho\ \delta^2_1\ (+1) = 1.0 + 1.0 * 0.0000562 * 1.0 = 1.0000562$

$\theta^{21}_{11} = \theta^{21}_{11} + \rho\ \delta^2_1\ s^2_1 = 1.0 + 1.0 * 0.0000562 * 0.96402 = 1.0000542$

$\theta^{20}_{12} = \theta^{20}_{12} + \rho\ \delta^2_1\ x_2 = 1.0 + 1.0 * 0.0000562 * 1.0 = 1.0000562$

$\theta^{10}_{10} = \theta^{10}_{10} + \rho\ \delta^1_1\ (+1) = 1.0 + 1.0 * 0.000004 * 1.0 = 1.0000040$

$\theta^{10}_{11} = \theta^{10}_{11} + \rho\ \delta^1_1\ x_1 = 1.0 + 1.0 * 0.000004 * 1.0 = 1.0000040$

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

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

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

$P(w \mid y,z)$		w: a	b
Y: a	Z: a	1/2	1/2
a	b	1/4	3/4
b	a	1/5	4/5
b	b	3/5	2/5

$P(y \mid x)$	y: a	b
x: a	1/3	2/3
b	1/4	3/4

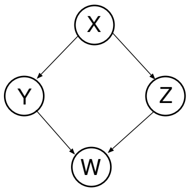
$P(z \mid x)$	z: a	b
x: a	1/3	2/3
b	1/4	3/4

a) Representar gráficamente la red

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

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

a) Representación gráfica de la red:



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

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

$$
P(W=\mathrm{b},\ Y=\mathrm{b},\ Z=\mathrm{b} \mid X=\mathrm{b}) = \frac{3}{4}\cdot\frac{3}{4}\cdot\frac{2}{5} = \frac{9}{40} = 0.225
$$

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

$$
\begin{aligned}
P(X \mid W,Y,Z) &= \frac{P(W,X,Y,Z)}{P(W,Y,Z)} = \frac{P(X)\ P(Y \mid X)\ P(Z \mid X)\ \cancel{P(W \mid Y,Z)}}{\sum_{x\in\{\mathrm{a,b}\}} P(X=x)P(Y \mid X=x)P(Z \mid X=x)\ \ \cancel{P(W \mid Y,Z)}} \\
&= \frac{\cancel{(1/2)}\,P(Y \mid X)P(Z \mid X)}{\cancel{(1/2)}\sum_{x\in\{\mathrm{a,b}\}} P(Y \mid X=x)P(Z \mid X=x)} = \frac{P(Y \mid X)P(Z \mid X)}{\sum_{x\in\{\mathrm{a,b}\}} P(Y \mid X=x)P(Z \mid X=x)}
\end{aligned}
$$

$$
P(X=\mathrm{b} \mid W=\mathrm{b},\ Y=\mathrm{b},\ Z=\mathrm{b}) = \frac{P(Y=\mathrm{b} \mid X=\mathrm{b})P(Z=\mathrm{b} \mid X=\mathrm{b})}{\sum_{x\in\{\mathrm{a,b}\}} P(Y=\mathrm{b} \mid X=x)P(Z=\mathrm{b} \mid X=x)} = \frac{3/4\cdot 3/4}{3/4\cdot 3/4 + 2/3\cdot 2/3} = 0.5586
$$