Examen de Aprendizaje Automático

ETSINF, Universitat Politècnica de València, 24 de enero de 2022

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 $\boxed{\text{A}}$ 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: (1,1,1,1,1,1,1,1,1,1). Indicar cual de las siguientes afirmaciones es incorrecta.

A) El test efectivo es de 100 muestras
B) El error estimado es $p_e = 1.0\,\%$
C) El intervalo de confianza al 95 % es ± 0.6
D) El tamaño de entrenamiento efectivo es de 900 muestras

2 $\boxed{\text{C}}$ En el problema de optimización con restricciones

$$\begin{array}{ll} \text{minimizar} & q(\boldsymbol{\Theta}), \quad \boldsymbol{\Theta} \in \mathbb{R}^D \\ \text{sujecto a} & v_i(\boldsymbol{\Theta}) \leq 0, \quad 1 \leq i \leq k \end{array}$$

se cumplen las condiciones complementarias de Karush-Kuhn-Tucker $\alpha_i^\star v_i(\boldsymbol{\Theta}^\star) = 0$ para $1 \leq i \leq k$. Indicar cuál de las siguientes afirmaciones se deduce de ellas:

A) Si para un i, $\alpha_i^\star = 0$, entonces $v_i(\boldsymbol{\Theta}^\star) \ > \ 0$
B) Si para un i, $\alpha_i^\star = 0$, entonces $v_i(\boldsymbol{\Theta}^\star) = 0$,
C) Si para un i, $\alpha_i^\star > 0$, entonces $v_i(\boldsymbol{\Theta}^\star) \ = \ 0$
D) Si para un i, $v_i(\boldsymbol{\Theta}^\star) = 0$, entonces $\alpha_i^\star = 0$

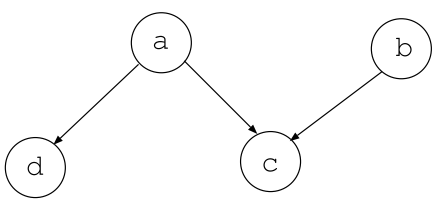
3 $\boxed{\text{B}}$ En el problema de aprendizaje de modelos probabilísticos con variables observables $\boldsymbol{x}_n$ y latentes $\boldsymbol{z}_n$, la log-verosimilitud se expresa como:

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

y 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 una estimación de los valores de las variables latentes $\boldsymbol{z}_n$, y el paso M, obtener los parámetros óptimos de $\boldsymbol{\Theta}$ utilizando los estimadores de $\boldsymbol{z}_n$ obtenidas en el paso E.
C) 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.
D) 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.

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



¿cuál de las relaciones siguientes es falsa en general?

A) $P(a, b) \ = \ P(a) \ P(b)$
B) $P(a, d) \ = \ P(a) \ P(d)$
C) $P(d, c \mid a) \ = \ P(d \mid a) \ P(c \mid a)$
D) $P(b, d) \ = \ P(b) \ P(d)$

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

Se dispone de un conjunto de entrenamiento en $\mathbb{R}^2$ con el que se prentende entrenar un modelo basado en máquinas de vectores soporte. Los multiplicadores de Lagrange óptimos obtenidos para $C = 10$ son:

i	1	2	3	4	5	6	7	8
x_{i1}	3	2	4	2	2	4	1	4
x_{i2}	4	5	4	2	3	2	4	3
Clase	–1	–1	–1	+1	–1	+1	+1	–1
$\alpha_i^\star$	0	0	0	0	10	9.11	7.11	6.22

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 $(6, 5)^t$.

a) Pesos de la función discriminante:

$\boldsymbol{\theta}^\star = c_5\ \alpha_5^\star\ \mathbf{x_5} + c_6\ \alpha_6^\star\ \mathbf{x_6} + c_7\alpha_4^\star\ \mathbf{x_7} + c_8\alpha_8^\star\ \mathbf{x_8}$

$$\theta_1^* = (-1)\ (2)\ (10) + (+1)\ (4)\ (9.11) + (+1)\ (1)\ (7.11) + (-1)\ (4)\ (6.22) = -1.33$$
$$\theta_2^* = (-1)\ (3)\ (10) + (+1)\ (2)\ (9.11) + (+1)\ (4)\ (7.11) + (-1)\ (3)\ (6.22) = -2.00$$

Usando el vector soporte $\mathbf{x_7}$ (que verifica la condición : $0 < \alpha_1^* < C$)

$\theta_0^\star = c_7 - \boldsymbol{\theta}^{\star t}\mathbf{x_7} = 1 - ((-1.33)\ (1) - (2.00)\ (4)) = 10.33$

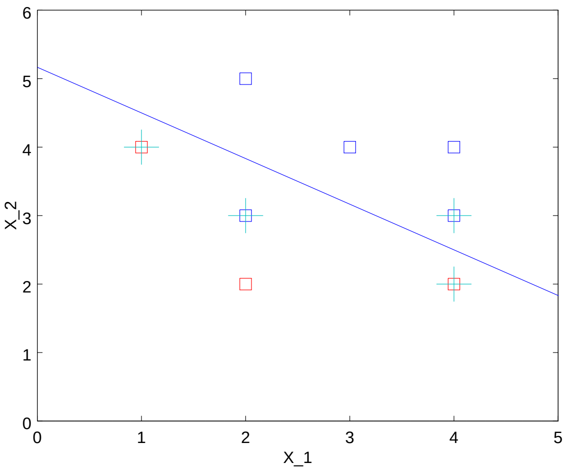
Función discriminante: $\phi(\mathbf{x}) = 10.33 - 1.33x_1 - 2x_2$

b) Frontera de separación y representación gráfica:

Ecuación de la frontera lineal de separación: $10.33 - 1.33\ x_1 - 2.00\ x_2\ =\ 0\ \ \rightarrow\ \ x_2 = -0.665\ x_1 + 5.165\,.$

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

Representación gráfica:

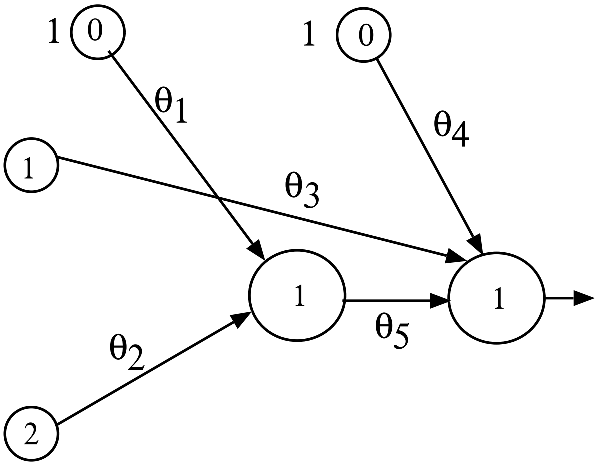


c) Clasificación de la muestra $(6, 5)^t$:

El valor de la función discriminante para $(6, 5)^t$ es: $\theta_0^* + 6 * \theta_1^* + 5 * \theta_2^* = 10.33 + 6 * (-1.33) + 5 * (-2.00) = -7.65 < 0 \Rightarrow$ 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 oculta de tipo *sigmoid* y lineal en el nodo de la capa de salida, y factor de aprendizaje $\rho = 1.0$.



Dados unos pesos iniciales $\theta_1 = \theta_2 = \theta_3 = \theta_4 = \theta_5 == 1.0$, un vector de entrada $\boldsymbol{x}^t = (0, 1)$ 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

Pista: La actualización de pesos en esta red sigue la misma formulación que en el BackProp para el perceptrón multicapa convencional: el incremento de peso es $\Delta\theta \;=\; \rho\; z\; \delta$, donde ρ es el factor de aprendizaje, z es la entrada del arco asociado al peso θ, y δ es el error que se observa en la salida de la unidad a la que llega ese arco, multiplicado por la derivada de la función de activación.

a) Las salidas de todas las unidades

$\phi_1^1 = \theta_1 + \theta_2\; x_2 \;=\; 1.0 + 1.0\; 1.0 \;=\; 2.0$ $\qquad s_1^1 = \dfrac{1}{1 + \exp(-\phi_1^1)} \;=\; 0.880797$

$\phi_1^2 = \theta_4 + \theta_3\; x_1 + \theta_5\; s_1^1 \;=\; 1.0 + 1.0\; 0.0 + 1.0\; 0.880797 \;=\; 1.880797$ $\qquad s_1^2 = \phi_1^2 \;=\; 1.880797$

b) El error en la capa de salida es:

$\delta_1^2 = (t_1 - s_1^2) \;=\; (1.0 - 1.880797) \;=\; -0.880797$

El error en la capa de oculta es:

$\delta_1^1 = (\delta_1^2\; \theta_5)\; s_1^1\; (1 - s_1^1) \;=\; ((-0.880797)\; 1.0)\; 0.880797\; (1 - 0.880797) \;=\; -0.092478$

c) Los nuevos pesos son:

$\theta_1 = \theta_1 + \rho\; \delta_1^1\; (+1) \;=\; 1.0 + 1.0\; (-0.092478)\; 1.0 \;=\; 0.90752$

$\theta_2 = \theta_2 + \rho\; \delta_1^1\; x_2 \;=\; 1.0 + 1.0\; (-0.092478)\; 1.0 \;=\; 0.90752$

$\theta_3 = \theta_3 + \rho\; \delta_1^2\; x_1 \;=\; 1.0 + 1.0\; (-0.880797)\; 0.0 \;=\; 1.0$

$\theta_4 = \theta_4 + \rho\; \delta_1^2\; (+1) \;=\; 1.0 + 1.0\; (-0.880797)\; 1.0 \;=\; 0.119203$

$\theta_5 = \theta_5 + \rho\; \delta_1^2\; s_1^1 \;=\; 1.0 + 1.0\; (-0.880797)\; 0.880797 \;=\; 0.224196$

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

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

- $P(X)$ es uniforme: $P(X = \mathrm{a}) = P(X = \mathrm{b})$,
- $P(Y \mid X)$, $P(Z \mid X)$, $P(V \mid Y)$ y $P(W \mid Z)$ son idénticas a la siguiente tabla $P(A \mid B)$:

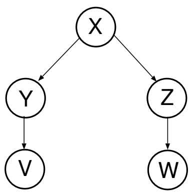
$P(A \mid B)$	$A:$ a	b
$B:$ a	1/3	2/3
b	1/4	3/4

a) Representar gráficamente la red.

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

c) Calcular $P(V = b \mid X = a)$,

a) Representación gráfica de la red:



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

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

$$P(W=\mathrm{b},\ Y=\mathrm{b},\ Z=\mathrm{b},\ V=\mathrm{b} \mid X=\mathrm{a}) = \frac{2}{3}\cdot\frac{2}{3}\cdot\frac{3}{4}\cdot\frac{3}{4} = \frac{36}{144} = 0.25$$

c) Calcular $P(V = b \mid X = a)$:

$$P(V \mid X) = \frac{P(V,X)}{P(X)} = \sum_{y,z,w\in\{a,b\}} P(Y=y \mid X)\ P(Z=z \mid X)\ P(W=w \mid Z=z)\ P(V \mid Y=y)$$

$$\begin{aligned} P(V=b \mid X=a) &= \sum_{y,z,w\in\{a,b\}} P(Y=y \mid X=a)\ P(Z=z \mid X=a)\ P(W=w \mid Z=z)\ P(V=b \mid Y=y) \\ &= \sum_{y,z\in\{a,b\}} P(Y=y \mid X=a)\ P(Z=z \mid X=a)\ P(V=b \mid Y=y) \\ &= \sum_{y\in\{a,b\}} P(Y=y \mid X=a)\ P(V=b \mid Y=y) \\ &= P(Y=a \mid X=a)\ P(V=b \mid Y=a) + P(Y=b \mid X=a)\ P(V=b \mid Y=b) \\ &= \frac{1}{3}\cdot\frac{2}{3} + \frac{2}{3}\cdot\frac{3}{4} = 0.8333 \end{aligned}$$