Examen de Aprendizaje Automático

ETSINF, Universitat Politècnica de València, 23 de enero de 2019

Apellidos: ______________________ **Nombre:** ______________

Cuestiones (2 puntos; tiempo estimado: 30 minutos)

Marca cada recuadro con una única opción de entre las dadas. Cada acierto suma 0.4 puntos y cada fallo resta 1/6 puntos.

1 $\boxed{D}$ Si aplicamos el método de partición de datos denominado exclusión individual (Leaving One Out) a un conjunto de 1000 muestras, indicar qué opción es *correcta*:

A) La talla de entrenamiento efectiva es de 1000 muestras y la de test de 1000 muestras.
B) La talla de entrenamiento efectiva es de 1000 muestras y la de test de 999 muestras.
C) La talla de entrenamiento efectiva es de 999 muestras y la de test de 999 muestras.
D) La talla de entrenamiento efectiva es de 999 muestras y la de test de 1000 muestras.

2 $\boxed{A}$ En un clasificador en 3 clases resulta que la probabilidad a-posteriori de cada clase, y, dada una muestra $x \in \{1, 2, 3\}$ y la probabilidad a-priori de $x \in \{1, 2, 3\}$ son:

$P(Y = y \mid X = x)$

y \ x	1	2	3
A	0.1	0.2	0.5
B	0.6	0.2	0.4
C	0.3	0.6	0.1

x	$P(X = x)$
1	0.2
2	0.3
3	0.5

Indicar cuál es la opción *correcta*:

A) El mínimo riesgo total es 0.45
B) El mínimo riesgo total es 0.55
C) El mínimo riesgo total es 0.5
D) El mínimo riesgo total es 0.4

3 $\boxed{D}$ La técnica *descenso por gradiente* aplicada a un problema de optimización da lugar a un algoritmo. Se pide identificar qué afirmación sobre este algoritmo es *falsa*.

A) Se aplica a problemas de minimización
B) Converge asintóticamente en determinadas condiciones
C) Es más rápido si el tamaño del paso de descenso es grande pero no se garantiza la convergencia
D) Siempre converge independientemente del tamaño del paso de descenso

4 $\boxed{B}$ Identificar qué afirmación es *cierta* en una máquina de vectores soporte:

A) En el caso de muestras no linealmente separables, los vectores soporte son las muestras que presentan un multiplicador de Lagrange óptimo cero
B) En el caso de muestras no linealmente separable, los vectores soporte son las muestras que presentan un multiplicador de Lagrange óptimo distinto cero aunque estén mal clasificados
C) En el caso de muestras no linealmente separable, los vectores soporte son las muestras que presentan un multiplicador de Lagrange óptimo distinto de cero y tolerancia cero
D) En el caso de muestras no linealmente separable, los vectores soporte son las muestras que presentan un multiplicador de Lagrange óptimo distinto de cero y la tolerancia menor que 1

5 $\boxed{C}$ En un modelo gráfico, sea $\mathcal{A}$ un conjunto de variables aleatorias y G el grafo que establece las dependencias entre las variables de $\mathcal{A}$. Identificar cuál de las siguientes afirmaciones es *cierta*.

A) Solo si el grafo es dirigido, el modelo gráfico define una distribución condicionale en las variables del grafo
B) Si el grafo es no dirigido el número de factores en la distribución conjunta es igual al número de nodos del grafo siempre
C) Cualquier distribución condicional o conjunta en la que participen todas o cualquier subconjunto de las variables de A, se puede deducir a partir de la distribución conjunta definida por G.
D) Si el grafo es dirigido el número de factores en la distribución conjunta es menor que el número de nodos del grafo

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	7
x_{i1}	1	1	4	4	2	1	3
x_{i2}	4	3	1	3	1	2	2
Clase	+1	+1	−1	−1	−1	+1	+1
$\alpha_i^\star$	0	0	0	5	6	1	10

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) Calcular las tolerancias óptimas.

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

a) Pesos de la función discriminante:

$$\boldsymbol{\theta}^\star = c_4\ \alpha_4^\star\ \mathbf{x_4} + c_5\ \alpha_5^\star\ \mathbf{x_5} + c_6\ \alpha_6^\star\ \mathbf{x_6} + c_7\ \alpha_7^\star\ \mathbf{x_7}$$

$$\theta_1^* \ = \ (-1)\ (5)\ (4) + (-1)\ (6)\ (2) + (+1)\ (1)\ (1) + (+1)\ (10)\ (3) \ = \ -1.0$$
$$\theta_2^* \ = \ (-1)\ (5)\ (3) + (-1)\ (6)\ (1) + (+1)\ (1)\ (2) + (+1)\ (10)\ (2) \ = \ +1.0$$

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

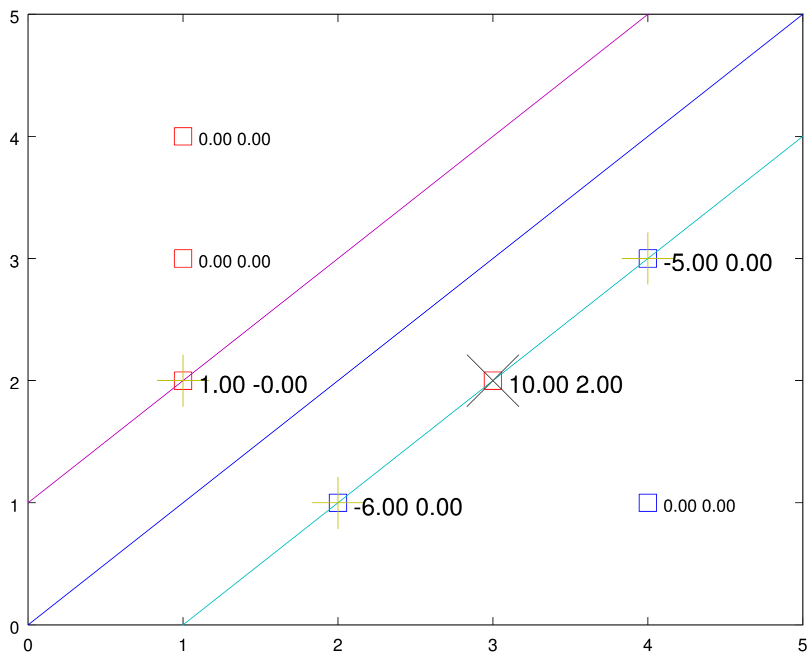
$$\theta_0^\star = c_5 - \boldsymbol{\theta}^{\star t}\mathbf{x_5} = -1 - ((-1.0)\ (2) + (1.0)\ (1)) \ = \ 0.0$$

b) Frontera de separación y representación gráfica:

Ecuación de la frontera lineal de separación: $(-1.0)\ x_1 + (1.0)\ x_2 + 0.0 \ = \ 0.0$

Loss vectores soporte son: $(4, 3)^t, (2, 1)^t, (1, 2)^t, (3, 2)^t$.

Representación gráfica:



c) Las tolerancias de todas las muestras excepto la séptima son cero y la séptima que es $\zeta_7^* \ = \ 1 - c_7\ (\boldsymbol{\theta}^{*t}\ \boldsymbol{x}_7 + \theta_0^*) \ = \ 2.0$;

d) Clasificación de la muestra $(3, 1)^t$:

El valor de la función discriminante para este vector es: $\theta_0^* + \theta_1^*\ 1 + \theta_2^*\ 1 \ = -2.0 < 0 \ \Rightarrow$ clase -1.

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

La solución para un determinado problema de regresión viene dado por el perceptrón multicapa, donde la función de activación de todos los nodos de la red son de tipo sigmoid y los pesos en una iteración dada del algoritmo BackProp son:

$$\boldsymbol{\theta}_1^1 = (1.0, -1.0, 0.0, 1.0)^t \qquad \boldsymbol{\theta}_1^2 = (0.0, 0.0)^t \quad \boldsymbol{\theta}_2^2 = (-1.0, -1.0)^t \quad \boldsymbol{\theta}_3^2 = (1.0, 1.0)^t$$

Supongamos que se dan la circunstancias siguientes:

Un vector de entrada	$x_1 = 0.0$	$x_2 = 1.0$	$x_3 = -1.0$
Las salidas de la capa oculta	$s_1^1 = 0.5$		
Las salidas de la capa de salida	$s_1^2 = 0.5$	$s_2^2 = 0.182$	$s_3^2 = 0.817$
Los valores deseados de la capa de salida	$t_1 = 0.1$	$t_2 = 0.9$	$t_3 = -0.9$

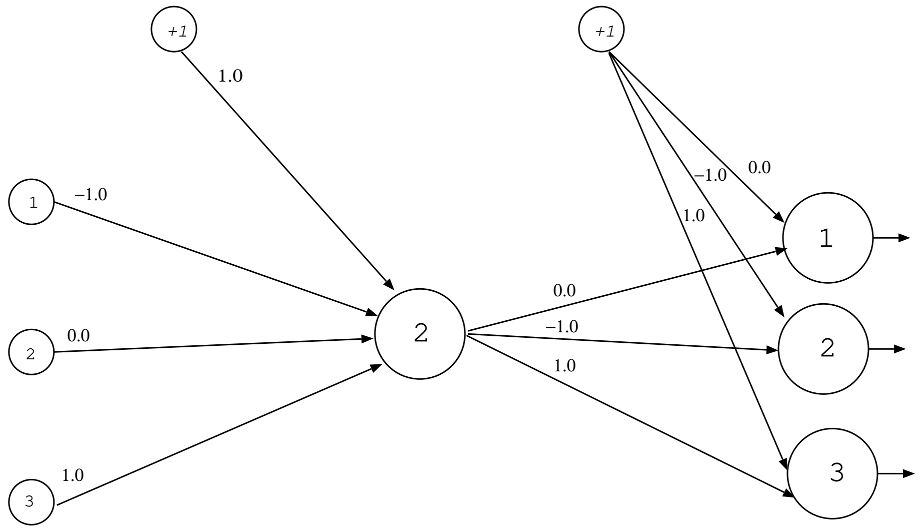
Se pide:

a) Dibujar el perceptrón multicapa descrito al principio del enunciado.

b) Calcular los errores (δ's) en los nodos de la capa de salida y en los nodos de la capa oculta.

c) Calcular los nuevos valores de los pesos $\theta_{3,1}^2$ y $\theta_{1,3}^1$ asumiendo que el factor de aprendizaje ρ es 1.0

a) Dibujo del perceptrón multicapa



b) Errores (δ's) en la capa de salida:

$\delta_1^2 = (t_1 - s_1^2)\ s_1^2\ (1 - s_1^2) = -0.1$
$\delta_2^2 = (t_2 - s_2^2)\ s_2^2\ (1 - s_2^2) = 0.107$
$\delta_3^2 = (t_3 - s_3^2)\ s_3^2\ (1 - s_3^2) = -0.256$

Errores en la capa de oculta:

$\delta_1^1 = (\delta_1^2\ \theta_{11}^2 + \delta_2^2\ \theta_{21}^2 + \delta_3^2\ \theta_{31}^2) s_1^1\ (1 - s_1^1) = -0.091$

c) Nuevo peso $\theta_{3,1}^2 = \theta_{3,1}^2 + \rho\ \delta_3^2\ s_1^1\ =\ 0.872$

Nuevo peso $\theta_{1,3}^1 = \theta_{1,3}^1 + \rho\ \delta_1^1\ x_3\ =\ 1.091$

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

Considerar la red bayesiana $\mathcal{R}$ definida como $P(R,X,Y,Z) = P(X)\ P(R \mid X)\ P(Y \mid R)\ P(Z \mid R)$, cuya variable R toma valores en $\{1,2,3\}$ y las variables X, Y, Z, en $\{\texttt{"a"},\texttt{"b"},\texttt{"c"}\}$. Las distribuciones de probabilidad asociadas son como sigue:

- $P(X)$ es uniforme: $P(X = \texttt{"a"}) = P(X = \texttt{"b"}) = P(X = \texttt{"c"})$
- $P(R \mid X)$ es uniforme: $P(R = 1 \mid x) = P(R = 2 \mid x) = P(R = 3 \mid x)$, $\forall x \in \{\texttt{"a"},\texttt{"b"},\texttt{"c"}\}$
- $P(Y \mid R)$ y $P(Z \mid R)$ son idénticas y vienen dadas en la siguiente tabla

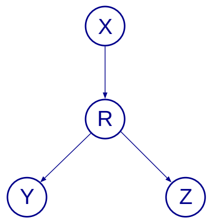
	"a"	"b"	"c"
1	1/3	0	2/3
2	1/4	1/2	1/4
3	0	3/5	2/5

a) Representar gráficamente la red

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

c) Calcular $P(R = 3 \mid X = \texttt{"b"}, Y = \texttt{"b"}, Z = \texttt{"b"})$

a) Representación gráfica de la red:



b)
$$\begin{aligned} P(R \mid X,Y,Z) = \frac{P(R,X,Y,Z)}{P(X,Y,Z)} &= \frac{P(X)\ P(R \mid X)\ P(Y \mid R)\ P(Z \mid R)}{\sum_{r=1}^{3} P(X)\ P(R=r \mid X)\ P(Y \mid R=r)\ P(Z \mid R=r)} \\ &= \frac{P(R \mid X)\ P(Y \mid R)\ P(Z \mid R)}{\sum_{r=1}^{3} P(R=r \mid X)\ P(Y \mid R=r)\ P(Z \mid R=r)} \\ &= \frac{\frac{1}{3}\ P(Y \mid R)\ P(Z \mid R)}{\sum_{r=1}^{3} \frac{1}{3}\ P(Y \mid R=r)\ P(Z \mid R=r)} \\ &= \frac{P(Y \mid R)\ P(Z \mid R)}{\sum_{r=1}^{3} P(Y \mid R=r)\ P(Z \mid R=r)} \end{aligned}$$

c)
$$\begin{aligned} P(R = 3 \mid X = \texttt{"b"}, Y = \texttt{"b"}, Z = \texttt{"b"}) &= \frac{P(Y = \text{"}b\text{"} \mid R = 3)\ P(Z = \text{"}b\text{"} \mid R = 3)}{\sum_{r=1}^{3} P(Y = \text{"}b\text{"} \mid R = r)\ P(Z = \text{"}b\text{"} \mid R = r)} \\ &= \frac{\frac{3}{5} \cdot \frac{3}{5}}{0 \cdot 0 + \frac{1}{2} \cdot \frac{1}{2} + \frac{3}{5} \cdot \frac{3}{5}} = \frac{9/5}{61/100} = \frac{36}{61} \approx 0.590 \end{aligned}$$