

APRENDIZAJE AUTOMÁTICO Y MINERÍA DE DATOS

Examen final

2 de septiembre de 2015

Recuerda poner tu nombre y apellidos en todas las hojas

BUENA SUERTE

PROBLEMA 1 (2 puntos). Se dan los siguientes conjuntos de datos.

$$x_{train} = \begin{pmatrix} 1 & 2 \\ 0 & 2 \\ 1 & 4 \\ -2 & 3 \\ 2 & 3 \end{pmatrix} \quad y_{train} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} \quad x_{test} = \begin{pmatrix} 1 & 2 \\ 3 & -1 \\ 1 & 0 \end{pmatrix}$$

Suponiendo que al aplicar el algoritmo SVM con los datos de entrenamiento (x_{train}, y_{train}) utilizando un kernel polynomial $K(x, l) = (x^T l - 1)^2$ se obtiene $\theta = (1, 2, -2, 1, 2, -3)^T$, ¿cuál sería la predicción para x_{test} ?

PROBLEMA 2 (4 puntos). Se da el siguiente conjunto de items $\mathcal{D} = \{abcd, abcd, abcd, abd, bcd, bcd, ab, ab, bc, b\}$ sobre el alfabeto $\mathcal{I} = \{a, b, c, d\}$. Se pide calcular FC_τ , FG_τ , $RI_{\tau, \gamma}$ y $RR_{\tau, \gamma}$ para $\tau = 0, 4$ y $\gamma = 0, 8$.

PROBLEMA 3 (3 puntos). Se da el siguiente conjunto de datos.

	outlook	temp.	humid.	windy	play
t_1	sunny	hot	high	False	no
t_2	sunny	hot	high	True	no
t_3	overcast	hot	high	False	yes
t_4	rainy	mild	high	False	yes
t_5	rainy	cool	normal	False	yes
t_6	rainy	cool	normal	True	no
t_7	overcast	cool	normal	True	yes

	outlook	temp.	humid.	windy	play
t_8	sunny	mild	high	False	no
t_9	sunny	cool	normal	False	yes
t_{10}	rainy	mild	normal	False	yes
t_{11}	sunny	mild	normal	True	yes
t_{12}	overcast	mild	high	True	yes
t_{13}	overcast	hot	normal	False	yes
t_{14}	rainy	mild	high	True	no

Se pide calcular la ganancia normalizada para los atributos **outlook**, **temp** y **humidity**.

PROBLEMA 4 (1 punto). Señala las respuestas correctas (cada pregunta puede tener varias respuestas correctas). Las respuestas incorrectas se penalizan.

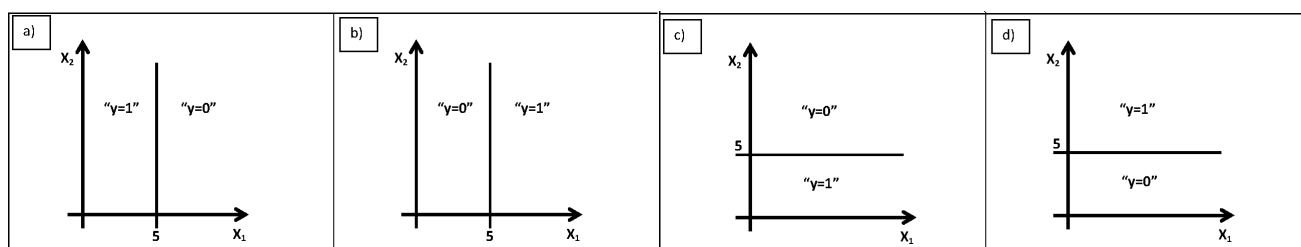
4.1 Se da f una función tal que $f(\theta_0, \theta_1)$ es un número real (la función f podría tener mínimos locales). Si utilizásemos el algoritmo del gradiente descendente para minimizar $f(\theta_0, \theta_1)$ en función de θ_0 y θ_1 , ¿cuál de las siguientes afirmaciones es verdadera?

- a) Si θ_0 y θ_1 se inicializan tal que $f(\theta_0, \theta_1)$ es el mínimo global, la primera iteración no cambiará sus valores.
- b) Indistintamente de como se inicializan θ_0 y θ_1 , mientras α es suficientemente pequeño, podemos esperar sin equivocarnos que el algoritmo del gradiente descendente convergerá a la misma solución.
- c) Si θ_0 y θ_1 se inicializan tal que $f(\theta_0, \theta_1)$ es un mínimo local, la primera iteración no cambiará sus valores.
- d) Si θ_0 y θ_1 se inicializan tal que $\theta_0 = \theta_1$, entonces por simetría (porque actualizamos de manera simultánea los dos parámetros), después de una iteración del algoritmo del gradiente descendente, seguiremos teniendo $\theta_0 = \theta_1$.

4.2 Nuestra tarea es identificar coches en imágenes de 100x100 píxeles (en escala de grises) utilizando como atributos el valor de intensidad de cada píxel. Si tuviésemos que entrenar un clasificador de regresión logística incluyendo todos los términos cuadráticos $(x_i x_j)$, ¿cuántos atributos tendríamos?

- a) $5 * 10^3$ b) 10^5 c) $5 * 10^7$ d) $5 * 10^9$

4.3 Queremos aplicar el modelo de regresión logística con dos variables x_1 y x_2 . Para $\theta_0 = 5, \theta_1 = -1, \theta_2 = 0$, ¿cuál de las siguientes gráficas representa el umbral de decisión para $h_\theta(x)$?



4.4 Para la regresión lineal con regularización, busquemos θ que minimiza la siguiente función:

$$J(\theta) = \frac{1}{2m} \left[\sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^p \theta_j^2 \right]$$

¿Qué pasa si λ tiene un valor muy grande (por ejemplo, $\lambda = 10^{10}$)?

- El algoritmo funciona perfectamente. Ponerle a λ un valor muy grande no le afecta.
- El algoritmo falla en eliminar el problema del sobreajuste.
- El algoritmo no se ajusta ni siquiera al conjunto de datos de entrenamiento.
- El algoritmo del gradiente descendente no converge.

DEFINICIÓN 1. Recordamos que en la fase de predicción, para un punto $x \in \mathbb{R}^{p+1}$ se calcula $f \in \mathbb{R}^{m+1}$ ($f_0 = 1$ y $f_j = K(x, x^{(j)}), \forall j \in \{1, \dots, m\}$) y predecimos “ $y = 1$ ” si y sólo si $\theta^T f > 0$.

DEFINICIÓN 2. Nota: la unión de conjuntos se denota por la justaposición de los mismos. Es decir, $abcd$ es el conjunto $\{a, b, c, d\}$ y XY es la unión del conjunto X con el conjunto Y .

$$F_\tau = \{X \subseteq \mathcal{I} \mid \sup(X) \geq \tau\},$$

$$FC_\tau = \{X \in F_\tau \mid \forall Z \supset X, \sup(Z) < \sup(X)\},$$

$$FG_\tau = \{X \in F_\tau \mid \forall Y \subset X, \sup(Y) > \sup(X)\},$$

$$RI_{\tau, \gamma} = \{X \in FC_\tau \mid \gamma * \max_{\tau, \gamma}(X) > \max_\tau(X)\}$$

$$RR_{\tau, \gamma} = \{X \rightarrow Y \mid Y \neq \emptyset, X \in FG_\tau, XY \in RI_{\tau, \gamma}, \max_\tau(XY) < \gamma * \sup(X) \leq \sup(XY) < \gamma * \min_\tau(X)\}$$

$$\max_\tau(X) = \max(\{\sup(Z) \mid Z \in FC_\tau, Z \supset X\} \cup \{0\}),$$

$$\min_\tau(X) = \min(\{\sup(Y) \mid Y \in FG_\tau, Y \subset X\} \cup \{\infty\}),$$

$$\max_{\tau, \gamma}(X) = \max(\{\sup(Y) \mid Y \in FG_\tau, Y \subset X, \sup(Y) \leq \frac{\sup(X)}{\gamma}\} \cup \{0\}).$$

DEFINICIÓN 3. Para valores C_1, \dots, C_k de la clase, sean q_1, \dots, q_k las frecuencias con que aparecen entre los datos de un conjunto S de tamaño s . La *información media* en S se define por $I(S) = -\sum_{i=1}^k q_i \log_2 q_i$. Al bifurcar por un atributo que genera la partición S_1, \dots, S_r de tamaños s_1, \dots, s_r podemos considerar la *ganancia de información* definida por $I(S) - \sum_{j=1}^r I(S_j) * \frac{s_j}{s}$.

La *ganancia normalizada* se calcula con la fórmula:

$$\frac{I(S) - \sum_{j=1}^r I(S_j) * \frac{s_j}{s}}{J(S)},$$

$$\text{donde } J(S) = -\sum_{j=1}^r \frac{s_j}{s} * \log_2 \frac{s_j}{s}.$$