



APELLIDOS:
NOMBRE:

DNI:

ESCUELA DE INGENIERÍA INFORMÁTICA

SISTEMAS INTELIGENTES

Examen Parcial de Teoría (Aprendizaje). Jueves 12 de diciembre de 2019.

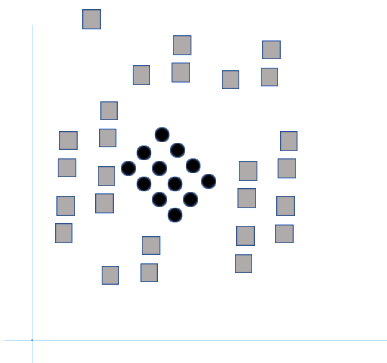
1. Tenemos un conjunto de datos que vamos a utilizar para predecir si un individuo padece una enfermedad. La variable a predecir toma dos valores (SI, NO). En nuestro conjunto de entrenamiento tenemos 25 individuos que padecen la enfermedad y 275 que no la padecen. Para resolver este problema utilizamos un método de aprendizaje supervisado, obteniendo un porcentaje de acierto del 91.6%. ¿Es el método elegido adecuado para resolver este problema? ¿Por qué?

No es adecuado porque su porcentaje de acierto es el mismo que si eligiéramos el clasificador que asigna la clase mayoritaria.

2. En el eje x de la curva ROC se representa FPR que se define como $FP/FP+TN$.
En el eje y se representa TPR que se define como $TP/TP+FN$.
3. Dada una variable, ¿qué significa que su entropía sea alta?

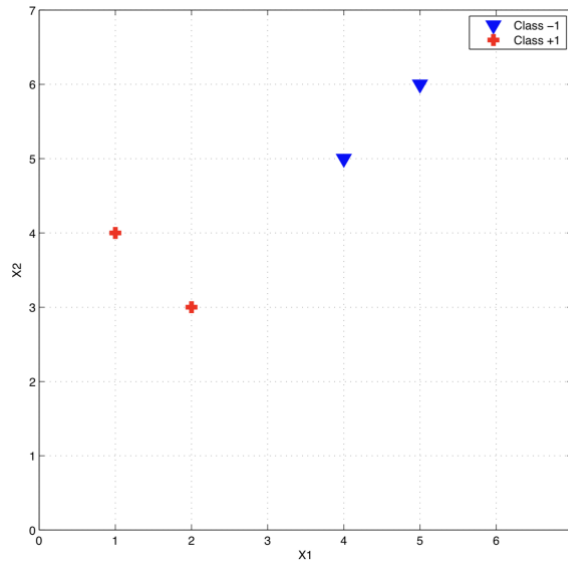
Significa que la variable no es útil como variable predictora.

4. Un árbol de decisión más profundo es más proclive al sobreajuste que un árbol menos profundo. Completa poniendo las palabras *más* o *menos*.
5. Dado el siguiente conjunto de datos, indica cual de los siguientes métodos es capaz de identificar las dos clases representadas.



Método	SI	NO
C4.5	X	
KNN	X	
Red Neuronal con función de activación salto, sin capas ocultas		X
Red Neuronal con función de activación lineal, sin capas ocultas		X
Red Neuronal con función de activación lineal, con una capa oculta		X
Red Neuronal con función de activación sigmoidea, sin capas ocultas		X
Red Neuronal con función de activación sigmoidea, con al menos una capa oculta	X	
SVM con margen duro		X
SVM con margen blando		X
SVM con Kernel polinomial de grado 2		X
SVM con kernel radial de ancho 0.5	X	

6. Dado el siguiente conjunto de datos, indica las coordenadas de los vectores soporte, determina w y b , y calcula la ecuación del hiperplano de separación.



SVM trata de maximizar el margen entre dos clases. El hiperplano de decisión pasa por el punto (3,4) y es perpendicular a la recta que pasa por los vectores (4,5) y (2,3). Por tanto su pendiente es $m = -1$.

Entonces la ecuación de la recta es $x_2 - 4 = -1(x_1 - 3)$. Es decir $x_1 + x_2 = 7$. Por tanto, podemos deducir que $w_1 = w_2$.

Por otra parte

$$2w_1 + 3w_2 + b = 1 \text{ y } 4w_1 + 5w_2 + b = -1$$

Luego

$$1 - 2w_1 - 3w_2 = -1 - 4w_1 - 5w_2$$

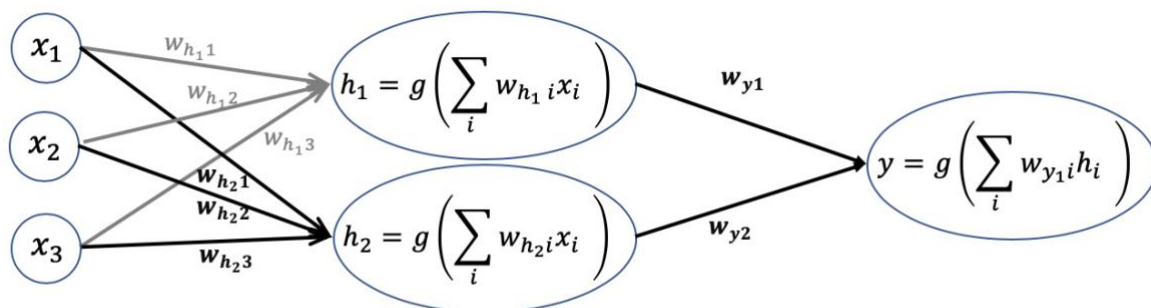
Pero como $w_1 = w_2$,

$$1 - 5w_1 = -1 - 9w_1, \text{ así que } 2 = -4w_1, \text{ luego } w_1 = w_2 = -0.5$$

Despejando en $2w_1 + 3w_2 + b = 1$ o en $4w_1 + 5w_2 + b = -1$, obtenemos $b = -7/2$.

Los vectores soporte son (1,4), (2,3) y (4,5).

7. Considera la siguiente red neuronal, con $g(x) = \frac{1}{1+e^{-x}}$. Considera el ejemplo de entrenamiento $(x_1, x_2, x_3, y) = (0, 1, 1, 2)$, de tal modo que cuando se presenta este ejemplo a la red se obtiene una salida con valor 1.1. Calcula la actualización del peso $w_{h_1 2}$ utilizando el algoritmo de propagación hacia atrás. El valor de la tasa de aprendizaje y la inicialización de pesos aparecen debajo de la red neuronal.



$$w_{h_1 j} = 0.1, \quad w_{h_2 j} = 0.2, \quad w_{y j} = 0.3, \quad \eta = 0.2$$

Debemos calcular en primer lugar calculamos $\partial_y = O_y(1 - O_y)(y - O_y) = 1.1(1 - 1.1)(2 - 1.1) = -0.099$.

Ahora debemos calcular ∂_{h_1} . Para ello, consideramos que

$$O_{h_1} = g(w_{h_11}x_1 + w_{h_12}x_2 + w_{h_13}x_3) = g(0.1 * 0 + 0.1 * 1 + 0.1 * 1) = g(0.2) = \frac{1}{1 - e^{-0.2}} = 5.52$$

Luego

$$\partial_{h_1} = O_{h_1}(1 - O_{h_1})w_{y_1}\partial_y = 5.52 * (1 - 5.52) * 0.3 * (-0.099) = 0.74.$$

Por tanto,

$$w_{h_12} = w_{h_12} + \eta x_2 \partial_{h_1} = 0.1 + 0.2 * 0.74 * 1 = 0.248$$

8. A partir del siguiente conjunto de datos queremos determinar si un estudiante aprobará en función de su calificación de prácticas (B,M,A) y de si ha estudiado o no. Calcula $H(\text{Aprueba}/\text{Estudio})$. Sabiendo que $H(\text{Aprueba})=0.92$ y $H(\text{Aprueba}/\text{Practicas})=0.66$, construye el árbol de decisión que se obtiene cuando la métrica de calidad es la Ganancia de Información.

Practicas	Estudio	Aprueba
B	No	No
B	Si	Si
M	No	No
M	Si	Si
A	No	Si
A	Si	Si

$H(\text{Aprueba}/\text{Estudio})=0.46$. Por tanto, la variable con la que se obtendrá mayor IG es la variable estudio. Así la raíz del árbol es la variable Estudio y el árbol que se obtiene es el siguiente:

