APELLIDOS: PL:
NOMBRE: DNI:

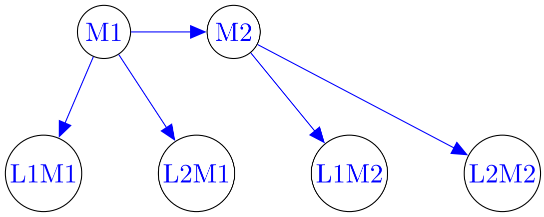
ESCUELA DE INGENIERÍA INFORMÁTICA
SISTEMAS INTELIGENTES
Segundo Examen Parcial. Miércoles 15 de diciembre de 2021.

II.Representación

Ejercicio 1. [2 puntos] Modela la red bayesiana para el siguiente juego, esto es el grafo y las probabilidades. En el juego existen tres monedas, A, B y C, que están trucadas de forma que la frecuencia con que sale cara con cada una de ellas es 0.3, 0.6 y 0.8 respectivamente. El juego se gana si se determina cuál es la moneda C, la que da cara con mayor frecuencia. Para esto el jugador puede elegir dos monedas distintas al azar y puede lanzar cada una dos veces. En base al resultado de los cuatro lanzamientos debe decidir si la moneda C es la que escogió y lanzó en primer lugar, en segundo o la que no se lanzó.

Solución Ejercicio 1

En este problema existen varias variables con incertidumbre que deben modelarse en la red bayesiana. Una de ellas es la moneda que se elige primero, que puede ser A, B o C, y también la moneda que se escoge en segundo lugar que solo podrá ser una de las dos que no se escogieron. También hay incertidumbre sobre el resultado de los lanzamientos. Por todo ello, la red bayesianas tiene 6 nodos con el siguiente grafo.



En el grafo, M indica moneda y L indica lanzamiento. El resultado del lanzamiento de cada moneda depende de la moneda que se eligió y la moneda que se escoge en segundo lugar depende de la primera elección porque no se puede volver a elegir la misma moneda.

Las tablas de probabilidad son las siguiente:

M1

A	0.33
B	0.33
C	0.33

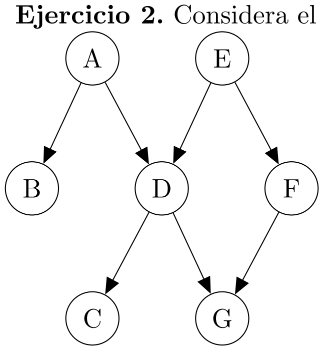
M2

M1	A	B	C
A	0	0.5	0.5
B	0.5	0	0.5
C	0.5	0.5	0

L?M?

M1/M2	A	B	C
Cara	0.3	0.6	0.8
Cruz	0.7	0.4	0.2

Ejercicio 2. Considera el siguiente grafo de una red bayesiana



a) **[0.5 puntos]** Indica razonadamente todas la variables que son independientes de B cuando no hay ninguna evidencia.

b) **[0.5 puntos]** Indica razonadamente todas la variables que son independientes de B cuando se conoce el valor de C.

Solución Ejercicio 2

a) Si no hay ningún nodo observado, los caminos solo se bloquean cuando se atraviesa por una configuración de efecto común. Desde el nodo B, solo los nodos E y F tienen todos sus caminos en los que aparece esta configuración.

b) Si el nodo C está observado, la configuración de efecto común A→D←E ya no bloquea por lo tanto el nodo E pasa a ser dependiente de B. También el no F es ahora dependiente de B porque se puede llegar por el camino B←A→D←E→F a pesar de que el nodo G bloquea el otro camino. Por tanto, todas las variables son dependiente de B dada C.

Ejercicio 3. [1 punto] Sobre la red del ejercicio anterior, suponiendo que todas la variables son booleanas, explica como funciona el algoritmo del muestreo estocástico para calcular $P(E, \neg C|G)$.

Solución Ejercicio 3

Para realizar el proceso de muestreo, en primer lugar hay que establecer un orden topológico. Para esta red un orden posible es {A,E,B,D,F,C,G}. Con este orden hay que muestrear cada variable según su tabla de probabilidad condicional teniendo en cuenta el valor de sus padres, si los tiene, que ya han sido asignados para la muestra actual.

La aproximación de la probabilidad pedida es la proporción de las muestras en las que se cumple que E es cierto, C es falso y G es cierto, el valor N_S, en relación a las muestras en las que G es cierto, el valor N_C. Es decir:

$$P(E, \neg C|G) = \frac{P(E, \neg C, G)}{P(G)} \approx \frac{N_S}{N_C}$$

III. Aprendizaje

Ejercicio 4. [1 punto] Explica como funciona el esquema de validación bootstrap, cuando es aconsejable usarlo y describe un par de variantes.

Solución Ejercicio 4

El esquema de validación bootstrap consiste en usar el mismo número de instancias para el entrenamiento que el número total de las que se disponen. Sin embargo, para tener la posibilidad de validar el modelo entrenado con instancias no usadas en el entrenamiento, se realiza un muestreo con reemplazo de forma que varias instancias aparecerán repetidas.

Como con este esquema se usan para el aprendizaje el mismo número de instancias que el total, es muy indicado para situaciones en las que el conjunto de datos es pequeño.

Algunas de las variantes que se pueden usar del esquema bootstrap son por ejemplo validar solamente con las instancias que no se usaron para entrenar o usar todas la instancias originales pero ponderadas de forma que las instancias usadas en el entrenamiento tendrán menos pesos que las que no.

Ejercicio 5. [1 punto] Para un problema de clasificación en el que la variable de interés es Y, que tiene los valores "sí" y "no", contamos con un conjunto de entrenamiento con 100 instancias y las variables A, B y C. La siguiente tabla muestra la distribución de valores de cada una de estas tres variables en función de la variable de interés.

	Y=sí	Y=no
A	a1:30 a2:20	a1:20 a2:30
B	b1:40 b2:10	b1:5 b2: 45
C	c1:15 c2:15 c3:20	c1:15 c2:20 c3:15

Razona que variable será escogida como nodo raíz de un árbol de decisión.

Solución Ejercicio 5

Cualquier algoritmo para construir un árbol de decisión elegirá como raíz la variable que mejor separe las instancias para cada valor de la variable de interés. Independientemente de la métrica que se pueda usar, en vista de la tabla dada se puede observar que la variable B es la que mejor determina la variable Y ya que para el valor b1 casi todos los casos son con Y=sí, mientras que para b2 casi todos los casos son con Y=no. Las otras dos variables, A y C, con cada uno de sus valores dejan la distribución de Y casi de manera uniforme.

Ejercicio 6. [2 puntos] Considera un perceptrón y dos entradas x_1 y x_2 donde el valor de la entrada de sesgo es 1 y los pesos asignados a cada entrada son $(w_0 = 2, w_1 = 1, w_2 = -1)$. La función de activación de la neurona es la función signo, de manera que los ejemplos se clasifican como positivos si el valor en el hiperplano es positivo o cero y como negativo en otro caso.

a) Escribe la ecuación del hiperplano separador inicial e indica cual es la clasificación que daría para el punto (-1, 2).

b) Supongamos que el punto (-1, 0) es de la clase negativa. Dados los pesos anteriores aplica la regla del perceptrón para que se clasifique correctamente y da la ecuación del hiperplano resultante usando una tasa de aprendizaje de 0.5.

Solución Ejercicio 6

La ecuación del hiperplano de separación inicial viene dada por $h(x_1, x_2) = x_1 - x_2 + 2$

a) Evaluamos esta ecuación en el punto (-1,2), obteniendo $h(-1, 2) = -1-2+2 < 0$, así que el ejemplo se clasifica como negativo.

b) $h(-1,0) = -1 + 2 = 1 > 0$. Luego el perceptrón diría que es de la clase positiva.
Actualizamos pesos (el superíndice 1 indica los pesos de la iteración anterior)

$$w_1^2 = w_1^1 + \eta \cdot (y - h(-1,0)) \cdot x_1 = 1 + 0{,}5 \cdot (-2) \cdot (-1) = 2$$

$$w_2^2 = -1 + 0{,}5 \cdot (-2) \cdot 0 = -1$$

$$w_0^2 = 2 + 0{,}5 \cdot (-2) \cdot 1 = 2 - 1 = 1$$

Ejercicio 7. [2 puntos] Dado el siguiente conjunto de entrenamiento, al que se le ha aplicado SVM con kernel lineal para predecir la variable Y, calcula los parámetros del hiperplano obtenido.

X1	3	6	8	7	1	5	9	3
X2	6	6	5	9	3	4	7	1
Y	1	-1	-1	-1	1	1	-1	1
α	0.111	0.445	0	0	0	0.333	0	0

Solución Ejercicio 7

$$\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i = 0{,}111 \cdot 1 \cdot \begin{pmatrix} 3 \\ 6 \end{pmatrix} + 0{,}445 \cdot (-1) \cdot \begin{pmatrix} 6 \\ 6 \end{pmatrix} + 0{,}333 \cdot 1 \cdot \begin{pmatrix} 5 \\ 4 \end{pmatrix} = \begin{pmatrix} -0{,}672 \\ -0{,}672 \end{pmatrix}$$

$$\begin{aligned} b &= \frac{1}{|VS|} \sum_i y_i - \mathbf{w}\mathbf{x}_i \\ &= \frac{1}{3} \cdot \left(1 - \begin{pmatrix} -0{,}672 & -0{,}672 \end{pmatrix} \begin{pmatrix} 3 \\ 6 \end{pmatrix} + (-1) - \begin{pmatrix} -0{,}672 & -0{,}672 \end{pmatrix} \begin{pmatrix} 6 \\ 6 \end{pmatrix} + 1 - \begin{pmatrix} -0{,}672 & -0{,}672 \end{pmatrix} \begin{pmatrix} 5 \\ 4 \end{pmatrix} \right) \\ &= 7{,}05 \end{aligned}$$

$P(x_1, \ldots, x_i) = \sum_{x_{i+1}, \ldots, x_n} P(x_1, \ldots, x_i, x_{i+1}, \ldots, x_n) = \sum_{x_{i+1}} P(x_1, \ldots, x_i, x_{i+1})$

$P(x|y) \equiv P(X = x|Y = y) = \frac{P(x,y)}{P(y)}$ // $P(X,Y) = P(X|Y)P(Y) = P(Y|X)P(X)$

$P(Y) = \sum_{i=1}^{m} P(Y|x_i)P(x_i)$ // $P(A|MB(A), B) = P(A|MB(A))$

$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i | X_{i+1} = x_{i+1}, \ldots, X_n = x_n)$

$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(X_i = x_i | Padres(X_i))$

$P(X|Padres(X)) = P(X|Padres(X), NoDescendientes(X))$

$H(X) = -\sum_{i=1}^{k} p_i \log_2 p_i$ // $H(X|Y) = \sum_{y \in Y} p(y)\, H(X|Y = y)$ // $IG(X,Y) = H(X) - H(X|Y)$

$GR(X,Y) = \dfrac{IG(X,Y)}{H(Y)}$ // $Gini(X) = 1 - \sum_{x \in X} p(x)^2$ // $Gini(X,Y) = Gini(X) - Gini(X|Y)$

$dist(\mathbf{x}, \mathbf{y}) = \left(\sum_i |x_i - y_i|^p \right)^{1/p}$ // $g(\cdot) = \begin{cases} 1 \text{ si } wx > 0 \\ -1 \text{ en otro caso} \end{cases}$ // $E(w) = \frac{1}{2} \sum_d \sum_k (y_{kd} - o_{kd})^2$

$\Delta w_i = \eta(y_d - wx_d)x_{di}$ // $\delta_k = o_k(1 - o_k)(y_k - o_k)$ // $\delta_h = o_h(1 - o_h) \sum_k w_{kh} \delta_k$

$E(w) = \frac{1}{2}(y_d - wx_d)^2$ // $d = \frac{2}{\|w\|}$ // $K(x,y) = (xy + 1)^d$ // $g(x) = \frac{1}{1+e^{-x}}$

$b = \frac{1}{|VS|} \sum_{i \in VS} y_i - wx_i$ // $w = \sum_{i \in VS} \alpha_i y_i x_i$ // $g(\mathbf{x}) = \sum_{i \in VS} \alpha_i y_i K(\mathbf{x_i}, \mathbf{x}) + b$