

## RECAP EJERCICIOS REDES BAYESIANAS

2022-01-17

**Ejercicio 3. [0.75 puntos]** Considera el caso en que la probabilidad conjunta de las variables  $\{A, B, C, D, E, F\}$  se puede escribir de la siguiente forma:

$$P(A, B, C, D, E, F) = P(A) \cdot P(B) \cdot P(C|A, B) \cdot P(D|A) \cdot P(E|C) \cdot P(F|D, E)$$

Dibuja el grafo de la red bayesiana que modela la relación entre estas 6 variables.

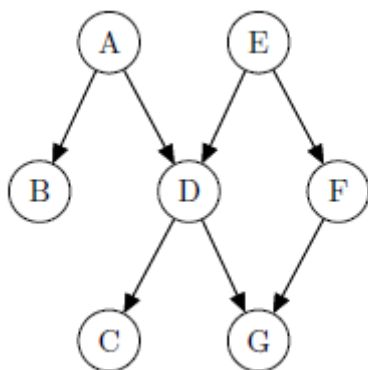
**Ejercicio 4.** Considera el siguiente grafo de una red bayesiana

- a) **[0.25 puntos]** Usando el criterio de D-separación, indica un par de variables que sean marginalmente dependientes, pero se necesite conocer exactamente el valor de otras dos para que sean independientes.
- b) **[0.25 puntos]** Indica el manto de Markov para la variable  $X_4$ .
- c) **[0.25 puntos]** Indica un orden topológico para este grafo.

2021-12-15

**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.

**Ejercicio 2.** Considera el siguiente grafo de una red bayesiana

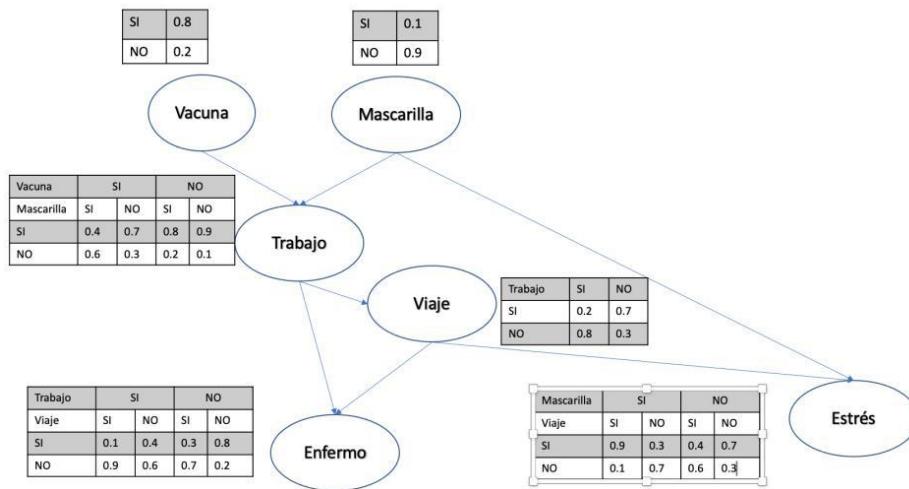


- a) **[0.5 puntos]** Indica razonadamente todas las variables que son independientes de B cuando no hay ninguna evidencia.
- b) **[0.5 puntos]** Indica razonadamente todas las variables que son independientes de B cuando se conoce el valor de C.

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

2021-05-24

4.- Dada la red bayesiana de la figura, utilizando el criterio de D-separación o la condición de Markov

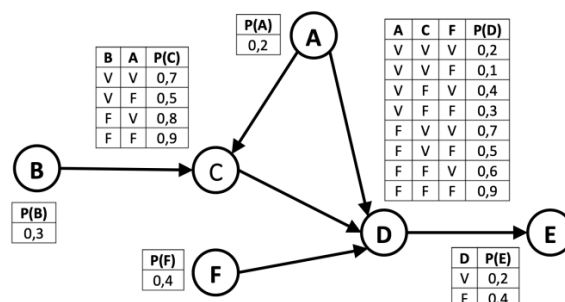


- [0.75 puntos]** Indica, utilizando el criterio de D-separación, el manto de Markov o la condición de Markov, si son ciertas o falsas las siguientes igualdades:
  - $P(\text{Vacuna}, \text{Mascarilla}) = P(\text{Vacuna}) \cdot P(\text{Mascarilla})$
  - $P(\text{Mascarilla} | \text{Trabajo}, \text{Viaje}, \text{Enfermo}) = P(\text{Mascarilla} | \text{Trabajo}, \text{Viaje})$
  - $P(\text{Vacuna} | \text{Estrés}) = P(\text{Vacuna})$
- [0.75 puntos]** Calcula  $P(\neg \text{Trabajo} = X3, \text{Viaje} = X4, \text{Estrés} = X6 | \text{Vacuna} = X1, \neg \text{Mascarilla} = X2)$
- [0.75 puntos]** Utilizando los números aleatorios siguientes: 0.04, 0.54, 0.65, 0.17, 0.32, 0.83, 0.97, 0.58, 0.09, 0.25; indica paso a paso cuál sería la muestra generada por el método de muestreo estocástico para calcular la probabilidad  $P(\text{Viaje}, \neg \text{Enfermo} | \neg \text{Vacuna}, \text{Estrés})$ .

2021-01-22

4.- Dada la red bayesiana de la figura, utilizando el criterio de D-separación o la condición de Markov

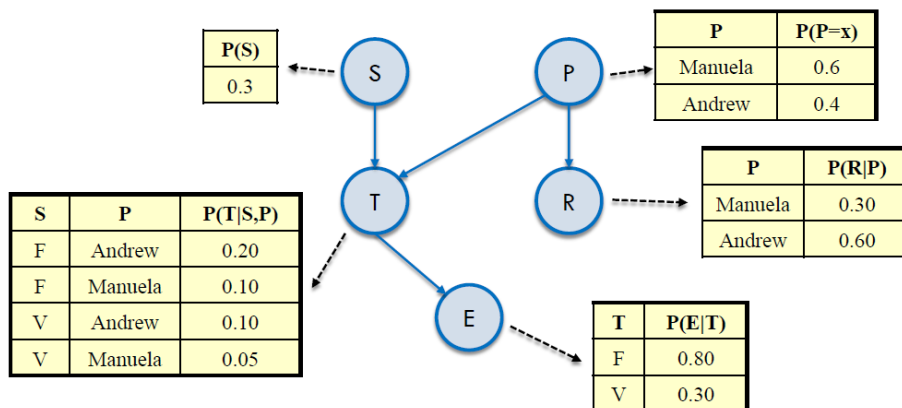
- [0.5 puntos]** Estudia que variables son condicionalmente independientes dadas otras variables y que variables son independientes si ninguna otra variable está observada.
- [0.5 puntos]** Calcula  $P(\neg A, \neg B, \neg C, D, E, F)$
- [0.5 puntos]** Calcula un ejemplo de muestra aplicando la ponderación de la verosimilitud para calcular  $P(B/A, \neg D)$ .



2020-06-26

**6.- [1,5 puntos]** Una de las pruebas de un concurso consiste en que el concursante tiene que elegir entre dos cajas, y solo una de ellas tiene el premio. Para elegir, puede sacar sin mirar dos bolas de forma consecutiva y sin reemplazo de solo una de las cajas. La caja con premio contiene 6 bolas rojas y 4 negras, mientras que la caja sin premio contiene 3 bolas rojas y 7 negras. Modela este problema mediante una red bayesiana, con su estructura y sus probabilidades condicionadas. Calcula, en base a la red creada, cuál será la probabilidad de haber sacado bolas de la caja buena si se han sacado en un caso primero una bola negra y después una roja, y en otro caso primero una bola roja y después una negra.

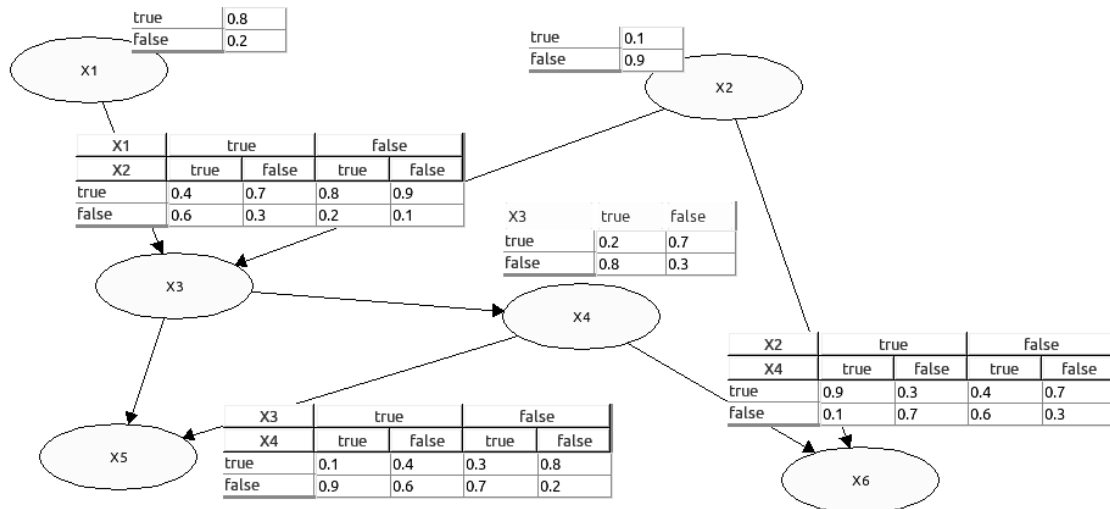
**7.- [1,5 puntos]** Dada la siguiente red bayesiana y usando las muestras que vienen debajo, obtén  $P(E|S,R)$  mediante el método de inferencia por muestreo con ponderación de la verosimilitud.



Muestra	P	R	S	T	E
1	Manuela	V	V	F	V
2	Manuela	V	V	F	F
3	Manuela	V	V	F	V
4	Andrew	V	V	F	F
5	Andrew	V	V	F	F
6	Manuela	V	V	F	V
7	Andrew	V	V	F	V
8	Andrew	V	V	V	V
9	Manuela	V	V	F	V
10	Manuela	V	V	F	V

2020-01-14

5.- Dada la siguiente red bayesiana



- [0.7 puntos]** Indica, utilizando el criterio de D-separación, el manto de Markov o la condición de Markov, si son ciertas o falsas las siguientes igualdades:
  - $P(X1, X2) = P(X1) \cdot P(X2)$
  - $P(X2 | X3, X4, X5) = P(X2 | X3, X4)$
  - $P(X1 | X6) = P(X1)$
- [0.7 puntos]** Calcula  $P(\neg X3, X4, X6 | X1, \neg X2)$
- [0.7 puntos]** Utilizando los números aleatorios siguientes: 0.04, 0.54, 0.65, 0.17, 0.32, 0.83, 0.97, 0.58, 0.09, 0.25; indica paso a paso cuál sería la muestra generada por el método de muestreo estocástico para calcular la probabilidad  $P(X4, \neg X5 | \neg X1, X6)$ .
- [0.5 punto]** Para la probabilidad del punto anterior, ¿cuántas muestras de las generadas por el método de muestreo estocástico se descartarían en general?

2019-11-18

2.- Tenemos tres variables aleatorias booleanas A, B y C, y la siguiente tabla de distribución de probabilidad conjunta:

A	B	C	P(A,B,C)
T	T	T	0.042
T	T	F	0.018
T	F	T	0.168
T	F	F	0.072
F	T	T	0.336
F	T	F	0.084
F	F	T	0.224
F	F	F	0.056

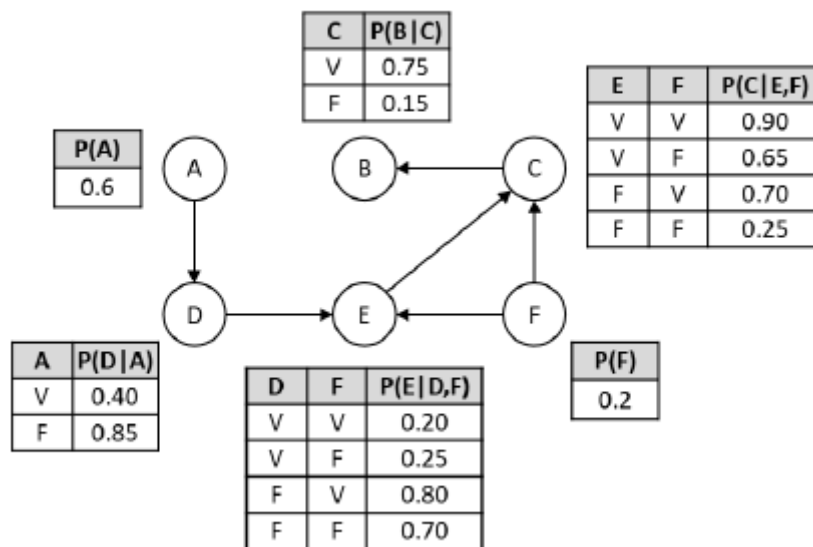
Se pide responder de forma razonada (es decir indicando en cada paso cuál es la definición, regla, teorema, criterio, etc. de los vistos en las clases de teoría) a las siguientes cuestiones.

- [0.5 puntos]** Demostrar que las variables B y C no son independientes.
- [2 puntos]** Demostrar que las variables B y C son independientes si conocemos el valor de la variable A.

- c) **[3 puntos]** A partir de las dependencias/independencias anteriores, diseñar una red bayesiana para las variables A, B y C. Esto incluye la estructura de la red y las tablas de probabilidad de cada nodo
- d) **[1,5 puntos]** A partir de la red bayesiana anterior, calcular la tabla de probabilidad conjunta y comprobar que es la misma que la tabla anterior.
- e) **[1 punto]** A partir de la red bayesiana anterior, obtener dos muestras por el método de muestreo estocástico. La primera suponiendo que en cada paso sale el valor más probable y la segunda suponiendo que en cada paso sale el valor menos probable. ¿Son coherentes estas muestras con los valores de la tabla de distribución conjunta?

2019-06-26

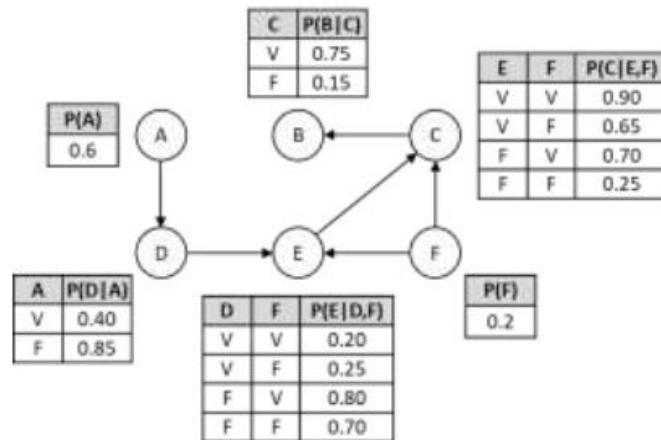
4.- Dada la siguiente red bayesiana, responder razonadamente a las siguientes preguntas:



- a) **[0,5 puntos]** Sabiendo que la variable E es falsa, ¿podemos afirmar que las variables A y B son independientes? Justifica la respuesta utilizando el criterio de D-separación, indicando todos los posibles caminos y por qué están, o no están, bloqueados.
- b) **[1 punto]** Sabiendo que A y F se cumplen, pero B y D no, ¿cuál es la probabilidad de que se cumpla C?.

2019-05-27

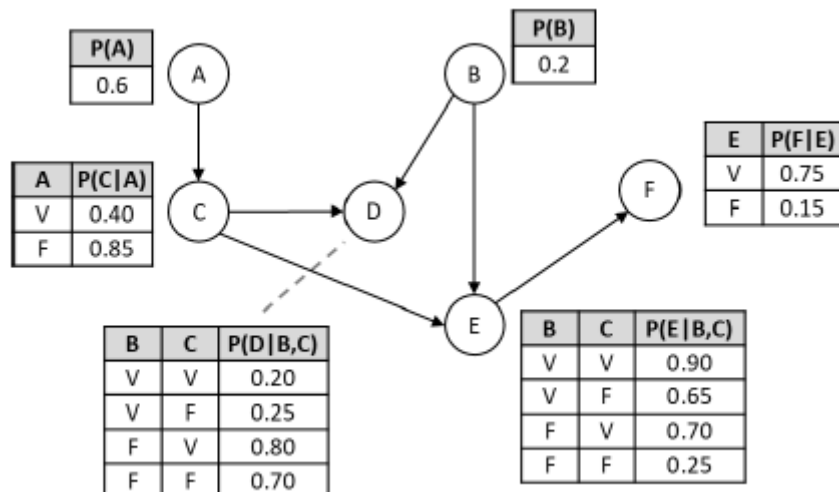
4.- Dada la siguiente red bayesiana, responder razonadamente a las siguientes preguntas:



- [0,5 puntos]** Sabiendo que la variable B es falsa, ¿podemos afirmar que las variables A y F son independientes? Justifica la respuesta utilizando el criterio de D-separación.
- [0,75 puntos]** Se desea calcular la probabilidad  $P(E \mid \neg D, \neg B)$  mediante el método de muestreo estocástico. Haz una iteración de este método utilizando la siguiente secuencia de números aleatorios: 0.9, 0.45, 0.6, 0.5, 0.15, 0.8.

2019-01-11

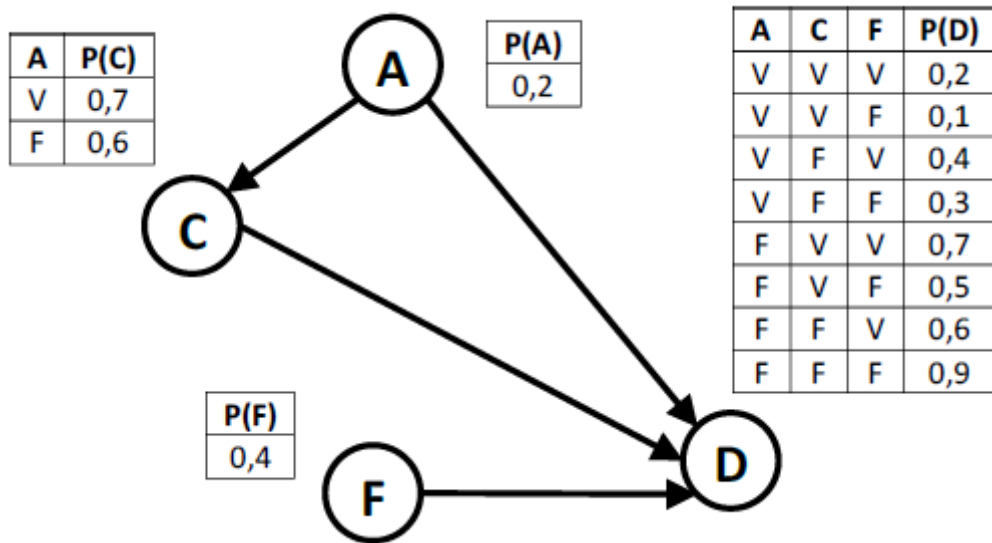
4.- Dada la siguiente red bayesiana, responder razonadamente a las siguientes preguntas:



- [0,75 puntos]** ¿Son las variables aleatorias A y B independientes? ¿Y las variables F y A? Justifica ambas respuestas utilizando el criterio de D-separación.
- [0,75 puntos]** Sabiendo que A y F se cumplen, pero B y E no, ¿cuál es la probabilidad de que se cumpla D?
- [0,25 puntos]** Indica las ventajas del método de ponderación de la verosimilitud sobre el muestro estocástico.

2018-01-11

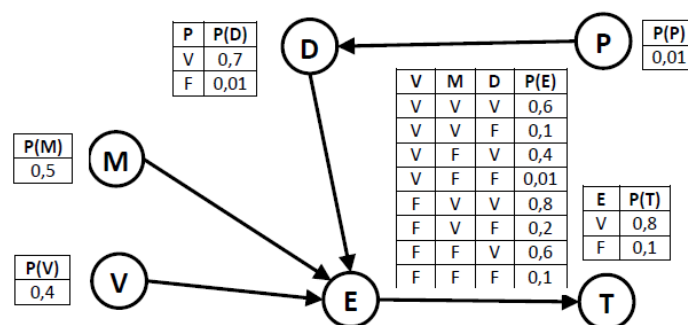
5.- Considera la siguiente red bayesiana:



- [0,7 puntos]** Calcula de forma exacta la siguiente probabilidad condicionada:  $P(D | \neg A, F)$ .  
Nota: no es necesario dar el resultado final, sino que puedes dejar las operaciones indicadas, por ejemplo:  $\frac{(0,5 \times 0,3) + 0,4}{0,8 \times 0,4}$
- [0,4 puntos]** ¿A y F son independientes, si conocemos el valor de D? Debes razonar la respuesta utilizando el criterio de D-separación, o bien la condición de Markov si ésta fuese suficiente.
- [0,4 puntos]** Supón que quieres calcular  $P(D | \neg C, F)$  utilizando el método de ponderación de la verosimilitud, y que una de las muestras que se generan es la siguiente:  $\{A=F, C=F, D=V, F=V\}$ . ¿Qué peso asignaría el método a esa muestra al generarla?

2017-05-22

5.- Considera la siguiente red bayesiana, en donde hay una serie de factores que influyen en la probabilidad de tener una determinada enfermedad (variable E). Dichos factores son: que el individuo tenga hábitos de vida saludables o no (variable V), que el sexo del individuo sea masculino o no (variable M), y que el individuo tenga o no un determinado defecto genético (variable D). La variable P indica si los padres del individuo son portadores del defecto genético, lo cual influye en la probabilidad de que el individuo tenga dicho defecto. Por último la variable T representa el resultado de un test médico que determina si el individuo tiene la enfermedad.



- a) [0,3 puntos] Calcula  $P(D, \neg V, P, M, T)$  (nota: si no tienes calculadora puedes dejar las operaciones indicadas, por ejemplo  $0,3 * 0,2 + 0,5 + 0,8 * 0,3 * 0,4$ )
- b) [0,3 puntos] Dada la estructura de la red, ¿podemos decir que P y M son independientes, si conocemos el valor de T? Debes razonar la respuesta utilizando el criterio de D-separación. También puedes utilizar la condición de Markov si ésta fuese suficiente.
- c) [0,4 puntos] Imagina que quieres añadir a la red la siguiente información:
- La máquina que realiza el test puede estar defectuosa (variable F), y en ese caso falla más a la hora de determinar si el individuo tiene o no tiene la enfermedad.
  - La enfermedad puede producir un extraño síntoma (variable S) con una alta probabilidad, pero es poco probable que se manifieste si no se tiene esta enfermedad. Además, el síntoma aparece con más frecuencia en individuos de sexo femenino.
- ¿Cómo modificarías la red? Es suficiente con que indiques la nueva estructura (nodos y arcos), no hace falta que indiques ninguna tabla de probabilidad.

2016-12-19

5.- [1 punto] Tenemos dos urnas: en la urna A hay dos bolas verdes y tres bolas azules. En la urna B hay tres bolas verdes y una bola azul. Alguien elige una urna al azar, saca una bola al azar y resulta que es verde. Luego saca otra bola (sin devolver a la urna la primera bola que sacó) y resulta que también es verde. ¿Cuál es la probabilidad de que sea la urna A?

Notas: Para resolver el ejercicio, construye la red bayesiana asociada al problema y calcula la inferencia sobre ella. Hacer bien la red bayesiana cuenta 0,5 puntos y hacer bien los cálculos correspondientes cuenta 0,5 puntos. Si no dispones de calculadora puedes dejar los cálculos indicados (por ejemplo  $0,25 + 0,5 / 0,25 + 0,8$ ).

2016-07-01

4.- [2 puntos] Razonamiento con incertidumbre

a) [1,2 puntos] Considerar el problema de transmitir palabras de longitud 3 del alfabeto  $A=\{a,b\}$  sobre un canal de transmisión. Las palabras se transmiten símbolo a símbolo. La transmisión tiene ruido y algunas veces no se recibe el símbolo emitido. Si se emite una 'a' se recibe una 'a' con probabilidad 0.8 y una 'b' con probabilidad 0.2. Si se emite una 'b' se recibe una 'b' con probabilidad 0.9 y una 'a' con probabilidad 0.1. La probabilidad de error en la recepción solo depende del símbolo emitido y no de la presencia de error en cualquier otro símbolo. Por otra parte, las palabras emitidas no son completamente aleatorias: el primer símbolo sí que es aleatorio, pero el segundo y el tercero dependen de si hubo algún error en la transmisión del símbolo que lo precede (se entiende que hubo un error si el símbolo recibido fue diferente al símbolo emitido) de la siguiente forma: si hubo un error en la transmisión del símbolo anterior entonces se emitirá una 'a' con probabilidad 0.7 y una 'b' con probabilidad 0.3. Si no hubo un error en la transmisión del símbolo anterior, entonces el símbolo emitido será aleatorio. Dibuja una red bayesiana que represente este problema, con sus correspondientes tablas de probabilidad condicionada. Si para alguna tabla necesitas alguna probabilidad no especificada en este enunciado, puedes proponer cualquier valor que consideres razonable.

b) [0,4 puntos] En la red bayesiana que hayas construido en el apartado a), razona si se cumplen o no las siguientes independencias, dada la estructura de la red:

- ¿Son independientes el primer símbolo emitido y el tercer símbolo recibido?



- ¿Son independientes el primer símbolo emitido y el tercer símbolo recibido, si conocemos el valor del segundo símbolo emitido?

c) [0,4 puntos] En una determinada red bayesiana compuesta por 5 variables A, B, C, D y E, queremos calcular la probabilidad  $P(A, \neg B \mid C)$  mediante el método del muestreo estocástico estándar. Para ello, hemos realizado 8 muestras aleatorias, cuyos valores resumimos en la siguiente tabla. Indica cuál sería el resultado final de dicha probabilidad, utilizando únicamente las muestras realizadas.

A	B	C	D	E
V	V	F	F	V
V	F	V	F	V
V	F	F	F	F
V	V	F	V	V
F	V	F	F	V
V	V	V	F	F
F	F	F	V	F
F	V	V	F	V

2016-05-23

#### 4.- [2 puntos] Razonamiento con incertidumbre

a) [0,5 puntos] Considera el siguiente problema: Tenemos tres variables booleanas A, B y C que pueden valer 0 o 1, y se cumple la siguiente relación entre las variables:  $C = \text{XOR}(A, B)$  (es decir, que  $C = (A+B) \bmod 2$ ). A priori, se eligen aleatoriamente los valores para A y para B. Dibuja una red bayesiana que represente este problema, con sus correspondientes tablas de probabilidad condicionada.

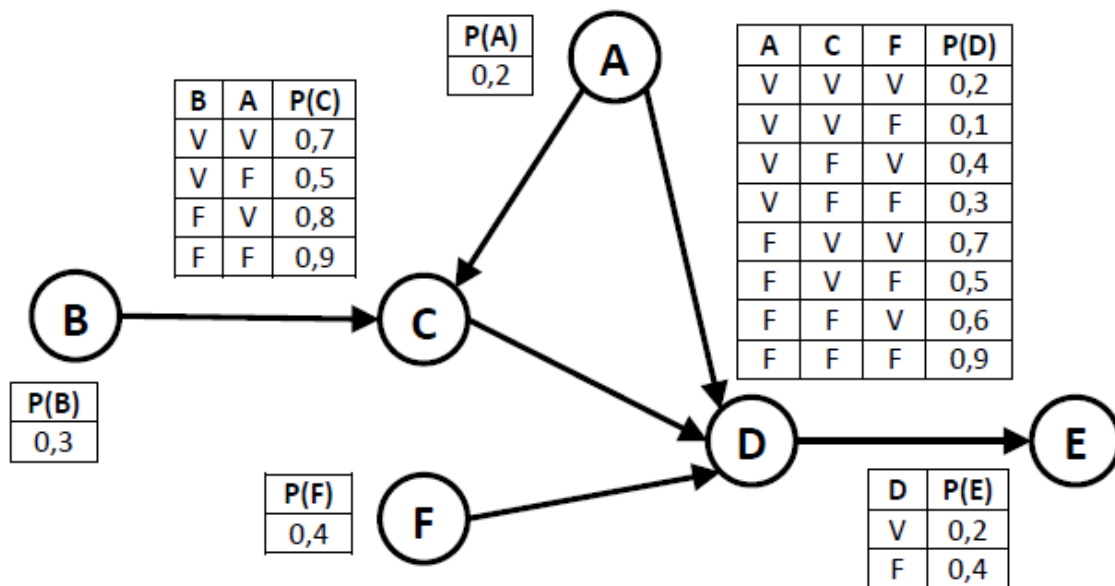
b) [0,9 puntos] Ampliación del apartado anterior: estamos enseñándole informática a un mono y queremos comprobar si se sabe la función XOR. Para ello, le damos al mono los valores de A y de B y el mono responde cuál piensa que es el resultado de  $\text{XOR}(A, B)$ . La respuesta del mono es la variable D. Lo ideal sería que D fuese igual que C. Pero puede ser diferente, o bien porque el mono no se sepa la función XOR, o bien porque se la sabe pero se ha equivocado al pulsar el botón. La variable  $K=1$  indica que el mono conoce la función XOR, y si no la conoce es  $K=0$ . Si el mono conoce la función XOR, responderá correctamente el 99% de las veces. Por el contrario, si no la conoce, el mono responderá de forma aleatoria. Supón que, a priori, la probabilidad de que el mono conozca la función XOR es de 0,4. Amplia la red bayesiana del apartado anterior (con sus correspondientes tablas de probabilidad condicionada) para representar este problema. Si para alguna tabla necesitas alguna probabilidad no especificada en este enunciado, puedes proponer cualquier valor que consideres razonable.

c) [0,6 puntos] En una determinada red bayesiana compuesta por 5 variables A, B, C, D y E, queremos calcular la probabilidad  $P(A, \neg B \mid C)$  mediante el método del muestreo estocástico estándar. Para ello, hemos realizado 8 muestras aleatorias, cuyos valores resumimos en la siguiente tabla. Indica cuál sería el resultado final de dicha probabilidad, utilizando únicamente las muestras realizadas.

A	B	C	D	E
V	V	F	F	V
V	F	V	F	V
V	F	F	F	F
V	V	F	V	V
F	V	F	F	V
V	V	V	F	F
F	F	F	V	F
F	V	V	F	V

2016-01-11

5.- Considera la siguiente red bayesiana:



- [0,6 puntos]** Calcula  $P(\neg A, \neg B, \neg C, D, E, F)$
- ¿Cuáles de las siguientes independencias se cumplen dada la estructura de la red? Debes razonar la respuesta utilizando el criterio de D-separación. También puedes utilizar la condición de Markov si ésta fuese suficiente.
  - (0,4 puntos)** ¿B y E son independientes, si conocemos el valor de D?
  - (0,4 puntos)** ¿B y F son independientes, si conocemos el valor de C y E?
- [0,6 puntos]** Supón que quieres calcular  $P(B|A, \neg D)$  utilizando el método de ponderación de la verosimilitud. Construye paso a paso una posible muestra, e indica qué peso le asignarías a dicha muestra.

**Nota:** en el apartado a) no es necesario dar el resultado final exacto, es suficiente con dejar indicadas las operaciones necesarias, por ejemplo  $0,4 \times 0,5 + 0,9 \times 0,2 \times 0,1 + 0,4 \times 0,9$

2015-06-15

5.- [2 puntos] Tenemos una enfermedad que queremos diagnosticar, y tres máquinas tales que cada una realiza un determinado test a un paciente. El resultado de dichos test puede ser positivo o negativo, y sabemos que los resultados de los test son condicionalmente independientes conocido si se tiene o no se tiene la enfermedad.

La enfermedad es más común en los fumadores (la tienen con probabilidad 0.005) que entre los no fumadores (la tienen con probabilidad 0.00001).

Por otra parte, sabemos que la probabilidad de que un individuo cualquiera sea fumador es de 0.3. Si las máquinas funcionan bien, las probabilidades de que los test den positivo según se tenga o no la enfermedad son las siguientes:

	Test 1 positivo	Test 2 positivo	Test 3 positivo
Enfermo	0.9	0.8	0.75
No enfermo	0.2	0.05	0.06

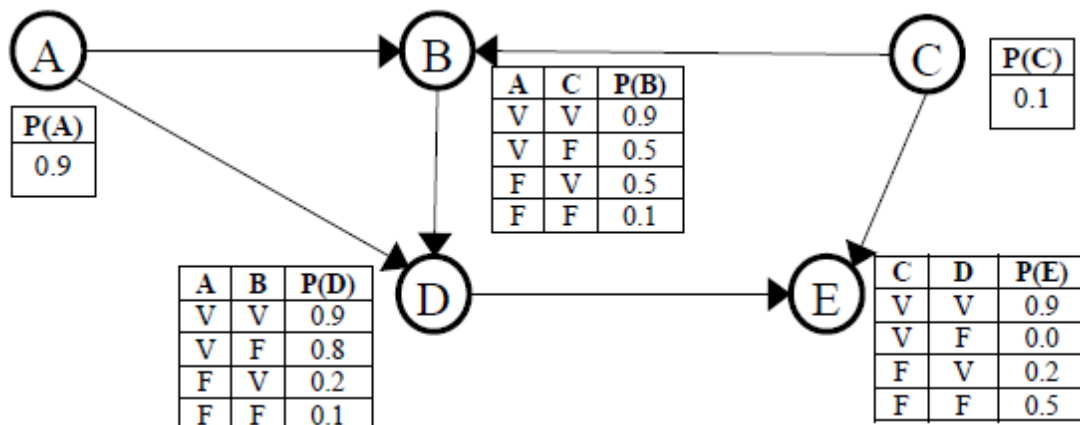
Por desgracia, las máquinas no son perfectas, sino que cada una de las máquinas funciona incorrectamente con probabilidad 0.001, y en ese caso devuelve positivo o negativo de forma aleatoria en su correspondiente test.

En esta pregunta se pide determinar una red bayesiana, con sus tablas de probabilidad, para representar el conocimiento anterior sobre el problema de diagnóstico de la enfermedad. Si en alguna tabla de probabilidad te hiciese falta algún dato no especificado en el enunciado del problema, puedes poner cualquier valor que consideres razonable.

2015-01-21

#### 4.- [2 puntos. Incertidumbre]

Considera la siguiente red bayesiana:



- (0.4 puntos)** Calcula  $P(\neg A, \neg B, \neg C, D, E)$
- (0.4 puntos)** Calcula  $P(\neg B, \neg C, D, E)$
- (0.5 puntos)** ¿Cuáles de las siguientes afirmaciones se cumplen, dada la estructura de la red? No es suficiente decir sí o no, debes razonar la respuesta.
  - $P(A, C) = P(A)P(C)$
  - $P(D|B) = P(D|B, A)$
- (0.7 puntos)** Supón que quieres calcular  $P(\neg B, \neg C, E | D)$ , pero no de forma exacta sino de forma aproximada, mediante el algoritmo estándar de muestreo. Explica brevemente cómo tendrías que proceder.

**Nota:** En los apartados a) y b) no es necesario dar el resultado final exacto, es suficiente con dejar indicadas las operaciones necesarias, por ejemplo  $0,4 \times 0,5 + 0,9 \times 0,2 \times 0,1 + 0,4 \times 0,9$ .



## RECAP APRENDIZAJE

2022-01-17

**Ejercicio 5. [0.75 puntos]** Dibuja la curva ROC para el clasificador del cual se han obtenido los siguientes resultados:

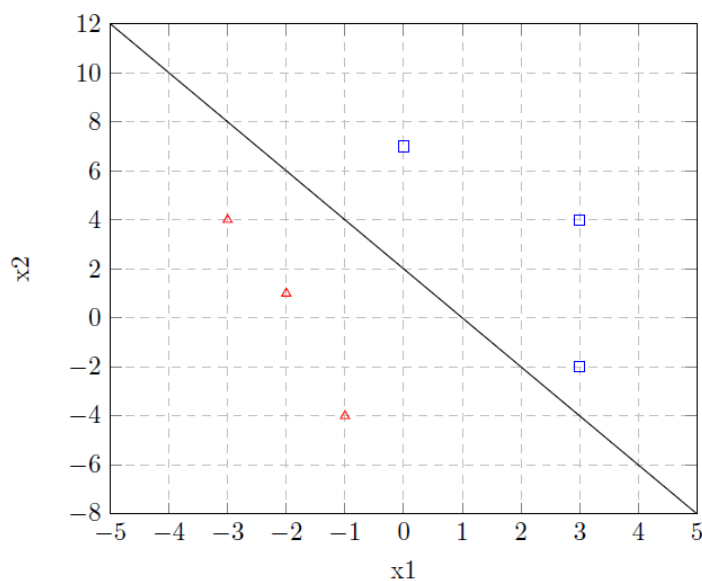
Real	pos	neg	pos	pos	neg	neg	neg	pos	neg	pos
Score(pos)	0.6	0.1	0.4	0.8	0.6	0.3	0.4	0.7	0.2	0.7

**Ejercicio 6. [0.75 puntos]** 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  $A$  y  $B$ . La siguiente tabla muestra la distribución de valores de cada una de estas variables en función de la variable de interés.

	$Y=\text{sí}$	$Y=\text{no}$
$A$	$a1:20 \ a2:40$	$a1:x \ a2:y$
$B$	$b1:40 \ b2:20$	$b1:5 \ b2:35$

Razona qué valores dar a  $x$  e  $y$ , es decir cuantas instancias para  $A=a1$  y  $A=a2$  cuando  $Y=\text{no}$ , de forma que la variable  $A$  sea elegida por delante de  $B$  como nodo de un árbol de decisión.

**Ejercicio 7. [1 punto]** Calcula los valores de los pesos  $w_0, w_1, w_2$  para el perceptrón cuya frontera de decisión se ilustra en la siguiente figura.



**Ejercicio 8. [1 punto]** Considera un modelo SVM lineal con margen duro cuyos parámetros son  $w = \begin{pmatrix} -0.8 \\ -0.4 \end{pmatrix}$  y  $b = 6.2$ . Este modelo se ha obtenido del conjunto de entrenamiento siguiente. Razona cuáles son los vectores soporte y a qué clase pertenecen.

$x_1$	3	6	8	8	3	5	9	5
$x_2$	7	6	5	5	3	3	6	1
$y$	1	-1	-1	-1	1	1	-1	1

2021-12-15

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

**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.

**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.

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

**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
$\alpha$	0.111	0.445	0	0	0	0.333	0	0

2021-05-24

**5.- [0.5 puntos]** Explica si el error en entrenamiento de un clasificador es una estimación optimista o pesimista.

**6.- [0.5 puntos]** Juan odia ver mensajes de spam en su bandeja de entrada. Sin embargo, no le importa comprobar periódicamente la carpeta *Correos no deseados* en busca de correos marcados incorrectamente como spam.

Patricia nunca consulta la carpeta *Correos no deseados* Prefiere ver correos electrónicos no deseados en su bandeja de entrada que perder correos auténticos sin saberlo.

Si construimos un clasificador de SPAM cuya precisión es 80% y la cobertura es del 60%, ¿quién estará más satisfecho con el clasificador? ¿Por qué?

**7.- [0.5 puntos]** Indica razonadamente si incrementar el número nodos en una capa oculta de un perceptrón multicapa siempre repercute en una mejora de la generalización.

8.- [1 punto] Tenemos un perceptrón con 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.

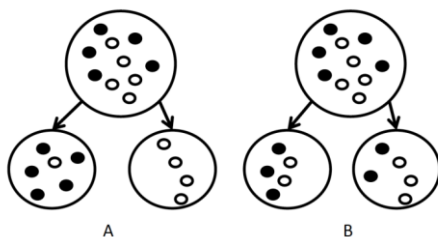
- Escribe la ecuación del hiperplano separador e indica cual es la clasificación que daría para los puntos (0, 0) y (-1, 2).
- Supongamos que el punto (-1, 0) es de la clase negativa. Dados los pesos anteriores aplica la regla del perceptrón para que se clasifique correctamente y da la ecuación del hiperplano resultante usando una tasa de aprendizaje de 0.5.

9.- [0.5 puntos] En relación al clasificador SVM

- Supón que tienes un clasificador binario SVM lineal. Considera un punto que está clasificado correctamente, y que está lejos de la frontera de decisión. Si eliminamos el ejemplo del conjunto de entrenamiento y volvemos a entrenar el clasificador, ¿cambiará la frontera de decisión o permanecerá igual? Explica tu respuesta
- Si los ejemplos de entrenamiento son linealmente separables, ¿cuántas fronteras de decisión pueden separar los ejemplos positivos de los negativos en un problema de clasificación binario? ¿Cuántas produce un SVM lineal con margen duro? ¿Qué características tienen, es decir, que propiedad verifican los hiperplanos que produce SVM?

10.- [0.5 puntos]. Supongamos que tenemos N ejemplos definidos mediante k variables binarias para predecir si este verano va a hacer calor o no. ¿Cuál es la profundidad un árbol de decisión que construyamos a partir de ese dataset (la profundidad de un árbol es la longitud del camino más largo desde la raíz a cada hoja)? ¿Y si las k variables son continuas?

La siguiente figura representa la división del conjunto de entrenamiento que se produce cuando utilizo como raíz la variable A o la B. Los círculos blancos y negros representan la etiqueta de la clase. ¿Cuál es la entropía de cada división? ¿Qué división produce más nodos puros?



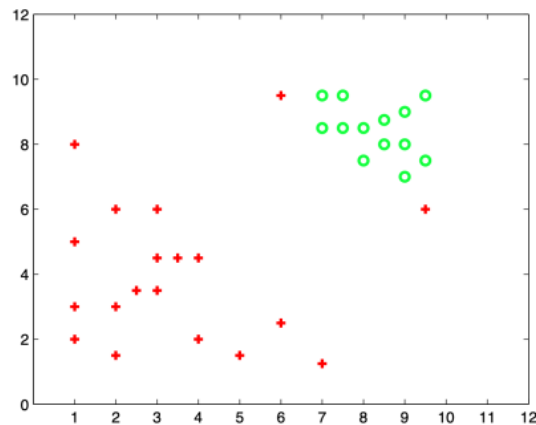
2021-01-22

5.- [0.5 puntos] Tenemos un conjunto de datos con 100 instancias para resolver un problema de clasificación binario. Explica **razonadamente** que esquema de validación utilizarías.

6.- [1 punto] Considera una red neuronal con **función de activación lineal**. Tiene dos entradas (a, b), una neurona oculta (c) y una unidad de salida (d). Habrá un total de cinco pesos ( $w_{ca}, w_{cb}, w_{c0}, w_{dc}, w_{d0}$ ) que vamos a inicializar todos a 0.1 y con una tasa de aprendizaje  $\eta=0.3$ . Calcula el valor de los pesos utilizando el algoritmo de propagación hacia atrás para la siguiente instancia (a,b,d)=(1,1,0). En este caso:  $\delta_k=(y_k-o_k)$  y  $\delta_h=\sum \delta_k w_{kh}$

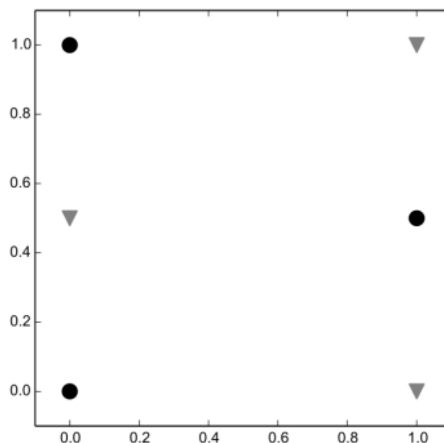
7.- [0.5 puntos] Dado el siguiente conjunto de datos que representa un problema de clasificación binario, pretendemos resolverlo utilizando SVM con kernel polinómico de grado 2. Dibuja una

posible frontera de separación cuando el valor de C es tan grande como sea posible y justifica por qué has dibujado esa frontera de decisión.



**8.- [0.5 puntos]** Dado el siguiente conjunto de datos, indica **razonadamente** cual o cuales de los siguientes métodos podrían resolverlo:

- SVM con Kernel lineal
- SVM con kernel radial
- Árboles de Decisión
- 3-NN

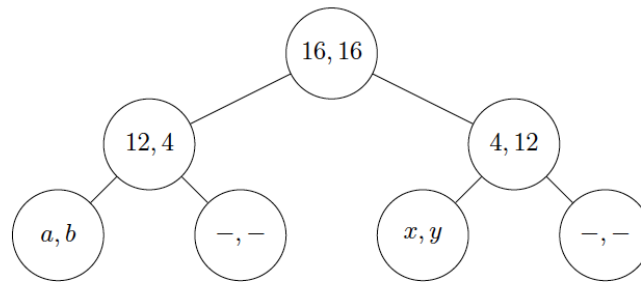


**9.- [0.5 puntos]** Explica las similitudes y diferencias entre la InfoGain (IG) y GainRatio (GR)

**10.-** La siguiente figura muestra un árbol de decisión. Los números de cada nodo representan el número de ejemplos en cada clase de ese nodo. Por ejemplo, en el nodo raíz, los números **16**; **16** representan 16 ejemplos de la clase 0 y 16 ejemplos de la clase 1. Para ahorrar espacio,

- a) **[0.5 puntos]** ¿Cuál es la ganancia de información en el nodo raíz?
- b) **[0.5 puntos]** ¿Cómo tienen que ser los valores de a y b para que la ganancia de información sea la más pequeña posible?





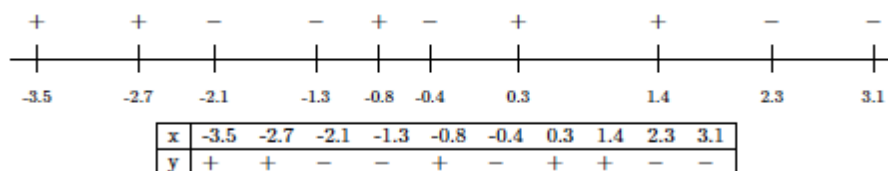
2020-06-26

8.- [0.5 puntos] Explica qué es un vector soporte, cuál es su posición relativa con respecto al hiperplano de separación y por qué son importantes en SVM.

9. [0.5 puntos] Dado el siguiente conjunto de entrenamiento, responde razonadamente si mediante el método C4.5 es posible construir un árbol de decisión que obtenga un porcentaje de acierto del 100% para ese conjunto. Si la respuesta es verdadera dibuja el árbol. Si es falsa explica la razón por la que es falsa.

A	B	C	Y
0	1	0	Yes
1	0	1	Yes
0	0	0	No
1	0	1	No
0	1	1	No
1	1	0	Yes

10.- [1.25 puntos] Considera los siguientes ejemplos con un solo atributo real clasificados en dos clases (+,-) y el clasificador KNN, con K=3.



Suponiendo que entrenamos este clasificador con Validación cruzada de 5 cajas, **da una posible partición (solo una)** y calcula el error cometido en **UNA** iteración de la validación cruzada.

11.- [1.25 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.

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

2020-01-14

**6.- [0.75 puntos]** Se utiliza un SVM con margen blando (considerando  $C=1.5$ ) para predecir si se le concede un préstamo o no a un cliente a partir de sus ingresos anuales, su edad y el importe del préstamo. Resolviendo el problema de optimización se obtienen 3 vectores soporte ( $sv_1$ ,  $sv_2$ ,  $sv_3$ ),  $sv_1$  y  $sv_2$  asociados a la clase *concede préstamo* (+1) y  $sv_3$  asociado a la clase *no concede préstamo* (-1). **Responde razonadamente** si los coeficientes  $\alpha_1, \alpha_2, \alpha_3$  pueden tomar los valores indicados en a) o b).

- a)  $\alpha_1=1, \alpha_2=1, \alpha_3=2$
- b)  $\alpha_1=0.5, \alpha_2=1, \alpha_3=0.5$

**7.- [0.5 puntos]** Considera un problema de clasificación binaria. Para resolverlo disponemos de un conjunto de entrenamiento con 30 ejemplos de la clase positiva y 10 de la clase negativa. Si utilizamos como conjunto de test el conjunto de entrenamiento, **responde razonadamente** cuál será el porcentaje de acierto que obtenemos al utilizar 40-NN.

**8.- [0.75 puntos]** Si tenemos un conjunto de entrenamiento con 150 ejemplos y realizamos validación cruzada con 10 cajas para entrenar un método de aprendizaje,

- a) ¿Cuántos entrenamientos se realizan?
- b) ¿Con cuántos ejemplos entrenamos el modelo en cada entrenamiento?
- c) ¿Cuántos ejemplos utilizamos para validar en cada entrenamiento de la validación cruzada?
- d) ¿Cuántas veces utilizamos cada ejemplo para entrenar en todo el proceso de validación cruzada?

**9.- [0.75 puntos]** Supongamos que tenemos que construir un árbol de decisión para predecir una variable binaria C. Tenemos 100 ejemplos de entrenamiento, de los cuales 30 son de la clase positiva y 70 de la negativa. Se selecciona el atributo A como raíz, se ramifica el conjunto según los valores de A en dos ramas, dando lugar a un hijo con 18 ejemplos positivos y 22 negativos y otro con 12 ejemplos positivos y 48 negativos.

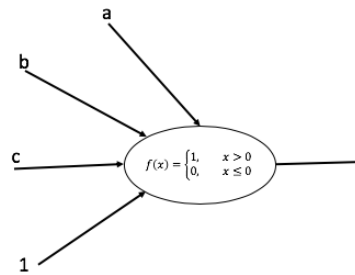
- a) Calcula  $H(C/A)$ .
- b) Calcula el error obtenido tras ramificar utilizando el atributo A.

**10.- [0.75 puntos]** Para qué sirve la **tasa de aprendizaje**  $\eta$  en una red neuronal?

Escribe una combinación de pesos que permita aprender la función **(a AND b) OR d** mediante una red neuronal con las siguientes características.

Nota1: No existe una solución única.

Nota2: Solo se piden los pesos finales, no que se entrene la red



2019-12-12

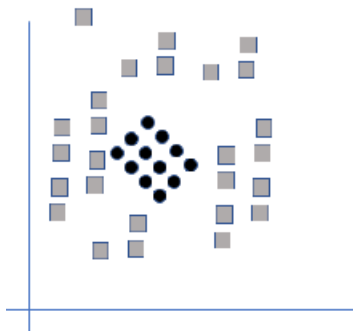
1 - Tenemos un conjunto de datos que vamos a utilizar para predecir si un individuo padece una enfermedad. La variable a predecir toma dos valores (SI, NO). En nuestro conjunto de entrenamiento tenemos 25 individuos que padecen la enfermedad y 275 que no la padecen. Para resolver este problema utilizamos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

2. En el eje x de la curva ROC se representa \_\_\_\_ que se define como \_\_\_\_\_. En el eje y se representa \_\_\_\_ que se define como \_\_\_\_\_.

3. Dada una variable, ¿qué significa que su entropía sea alta?

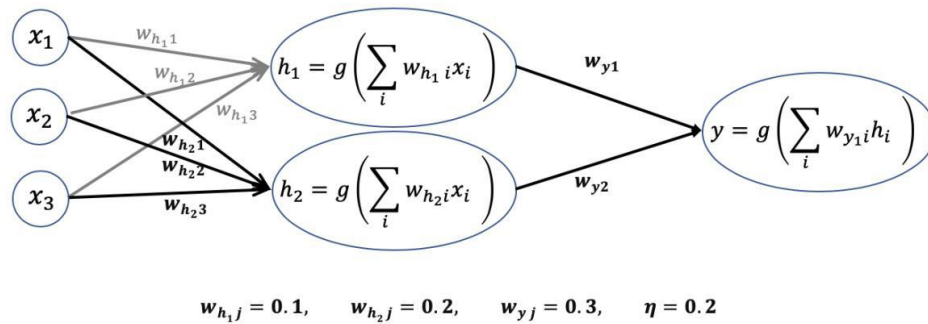
4. Un árbol de decisión \_\_\_\_ profundo es más proclive al sobreajuste que un árbol \_\_\_\_ profundo. Completa poniendo las palabras más o menos.

5. Dado el siguiente conjunto de datos, indica cual de los siguientes métodos es capaz de identificar las dos clases representadas.



Método
C4.5
KNN
Red Neuronal con función de activación salto, sin capas ocultas
Red Neuronal con función de activación lineal, sin capas ocultas
Red Neuronal con función de activación lineal, con una capa oculta
Red Neuronal con función de activación sigmoidea, sin capas ocultas
Red Neuronal con función de activación sigmoidea, con al menos una capa oculta
SVM con margen duro
SVM con margen blando
SVM con Kernel polinomial de grado 2
SVM con kernel radial de ancho 0.5

7. Considera la siguiente red neuronal, con  $g(x)=11-e^{-x}$ . 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  $w_{h1}$  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.



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

2019-06-26

5.- [0,5 puntos] Explica en que consiste la validación cruzada y sus ventajas y/o desventajas.

6.- [1 puntos] Dado el siguiente conjunto de datos donde Play es la variable a predecir, indica razonadamente que variable se seleccionaría como nodo raíz de un árbol de decisión usando el algoritmo ID3.

Outlook	Temp.	Humidity	Wind	Play
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

7.- [1 punto] Considera una red neuronal con función de activación lineal. Tiene dos entradas (a, b), una neurona oculta (c) y una unidad de salida (d). Habrá un total de cinco pesos ( $w_{ca}$ ,  $w_{cb}$ ,  $w_{c0}$ ,  $w_{dc}$ ,  $w_{d0}$ ) que vamos a inicializar todos a 0.1 y con una tasa de aprendizaje  $\eta = 0.3$ .

Calcula el valor de los pesos utilizando el algoritmo de propagación hacia atrás para la siguiente instancia:

a	b	d
1	1	0

Recuerda que en este caso:  $\delta_k = (y_k - o_k)$  y  $\delta_h = \sum_k \delta_k w_{kh}$

8.- [1 punto] Dado el siguiente conjunto de entrenamiento  $\{(1,+), (2,+), (4,-), (5,-), (6,+)\}$  al que se le ha aplicado SVM con kernel  $K(x, y) = (xy + 1)^2$ , obteniendo  $\alpha_1 = 0$ ,  $\alpha_2 = 2.499$ ,  $\alpha_3 = 0$ ,  $\alpha_4 = 7.331$ ,  $\alpha_5 = 4.832$ . Calcula la predicción de este modelo para 6.5.

2019-05-27

5.- [0,25 puntos] Supón un problema de clasificación para predecir si se da una enfermedad que suele aparecer en 5 casos de cada 1000 personas. Indica que medida de evaluación de modelos sería más adecuada en este caso y porqué.

6.- [0,75 puntos] Dada la siguiente tabla que muestra un conjunto de datos donde *Play* es la variable a predecir:

Temp	40	48	60	72	80	90
Play	No	No	Yes	Yes	Yes	No

Calcula razonadamente el valor de la ganancia de información (IG) sobre *Play* con *Temp*, que es numérica.

7.- [0,75 puntos] Dado el siguiente conjunto de ejemplos con un atributo real y una variable objetivo que toma dos valores (+,-)

+	+	-	-	+	-	+	+	-	-
-3.5	-2.7	-2.1	-1.3	-0.8	-0.4	0.3	1.4	2.3	3.1

x	-3.5	-2.7	-2.1	-1.3	-0.8	-0.4	0.3	1.4	2.3	3.1
y	+	+	-	-	+	-	+	+	-	-

Utilizando como conjunto de validación los ejemplos 1, 5 y 6  $\{-3.5, +\}$ ,  $\{-0.8, +\}$  y  $\{-0.4, -\}$  y como conjunto de entrenamiento el resto, calcula razonadamente el porcentaje de error que se obtiene con el paradigma de vecinos más cercanos con  $K=3$  y distancia euclídea.

8.- [1 punto] Describe en pseudo-código el algoritmo de propagación hacia atrás o backpropagation para aprender una red neuronal artificial a partir de un conjunto de entrenamiento con  $D$  casos de la forma  $\langle x, y \rangle$ , una tasa de aprendizaje  $\eta$ , y suponiendo una sola capa de neuronas ocultas con  $nh$  neuronas.

9.- [0,75 puntos] Dado el siguiente conjunto de entrenamiento  $\{(1,+), (2,+), (4,-), (5,-), (6,+)\}$  al que se le ha aplicado SVM con kernel  $K(x, y) = (xy + 1)^2$ , obteniendo  $\alpha_1 = 0$ ,  $\alpha_2 = 2.499$ ,  $\alpha_3 = 0$ ,  $\alpha_4 = 7.331$ ,  $\alpha_5 = 4.832$ . Sabiendo que  $b = 9$ , calcula razonadamente la predicción de este modelo para 3.5 y 5.5. NOTA: usa al menos un decimal significativo en los cálculos.

2019-01-11

5.- [0,25 puntos] Explica en que consiste la validación cruzada y sus ventajas y/o desventajas.

6.- [0,75 puntos] Describe el concepto de sobreajuste y con qué factores de los datos de entrenamiento está relacionado. Explica cómo se puede mitigar en árboles de decisión y SVM.

7.- [0,5 puntos] Explica la relación entre las métricas InfoGain (IG) y GainRatio (GR). ¿Hay algún motivo para preferir una sobre la otra?

8.- [1 punto] Considera una red neuronal con **función de activación lineal**. Tiene dos entradas (a, b), una neurona oculta (c) y una unidad de salida (d). Habrá un total de cinco pesos ( $w_{ca}, w_{cb}, w_{c0}, w_{dc}, w_{d0}$ ) que vamos a inicializar todos a 0.1 y con una tasa de aprendizaje  $\eta=0.3$ . Calcula el valor de los pesos utilizando el algoritmo de propagación hacia atrás para la siguiente instancia:

a	b	d
1	1	0

Recuerda que en este caso:  $\delta_k = (y_k - o_k)$  y  $\delta_h = \sum_k \delta_k w_{kh}$

9.- [1 punto] Considera el siguiente conjunto de entrenamiento, al que se le ha aplicado SVM con kernel lineal para predecir la variable Y. Sabiendo que  $b=10$ , calcula la predicción de este modelo para la instancia  $X1=3, X2=8$ .

X1	4	5	9	7	2	4	9	2
X2	5	6	4	9	1	4	8	0
Y	1	-1	-1	-1	1	1	-1	1
$\alpha$	1	1	0	0	0	0	0	0

2018-01-11

4.- [3,25 puntos]

4.1. [0,75 puntos] Describe el método K-NN (K-vecinos). Además de los pasos fundamentales del algoritmo, debes explicar si es o no necesario normalizar el conjunto de datos y por qué.

4.2. [0,75 puntos] Dado el siguiente conjunto de entrenamiento  $\{(1,+), (2,+), (4,-), (5,-), (6,+)\}$  al que se le ha aplicado SVM usando como kernel  $K(x,y)=(xy+1)^2$ , obteniendo  $a1=1, a2=3, a3=0, a4=4, a5=0$ . ¿Cuántos vectores soporte tiene cada clase? , ¿cuáles son los vectores soporte? ¿Es un clasificador lineal? ¿Por qué?

4.3. [0,75 puntos] Describe el fenómeno del sobreajuste en un árbol de decisión y explica como detectarlo.

4.4. [1 punto] Utilizando el siguiente conjunto de datos para construir un árbol de decisión que ayude a predecir si una seta es comestible o no en función de su forma, color y olor, resuelve las siguientes cuestiones:

Forma	Color	Olor	Comestible
C	Negro	A	SI
D	Negro	A	SI
D	Blanco	A	SI
D	Blanco	B	SI
C	Negro	B	SI
D	Negro	B	NO
D	Verde	B	NO
C	Marrón	B	NO
C	Negro	C	NO
C	Blanco	C	NO
D	Blanco	C	NO

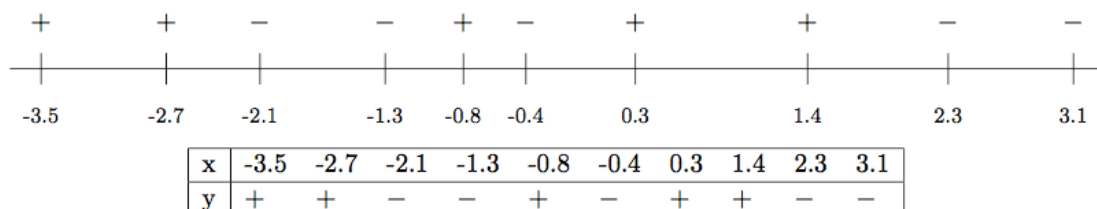
- ¿Qué atributo selecciona ID3 como raíz? ¿Y C4.5? ¿Es el mismo atributo? Explica la razón de que ID3 y C4.5 seleccionen ( o no) el mismo atributo como raíz como si no lo seleccionan.
- Construye el árbol de decisión que se obtiene utilizando ID3. **NOTA: En algunos casos no es necesario realizar todos los cálculos (y en otros no es necesario realizar ningún cálculo). En estos casos es suficiente que expliques por qué no hay que realizar ningún cálculo.**
- ¿Cuál es el error en entrenamiento?
- Si consideramos el siguiente conjunto de test, calcula el error y la medida F. ¿Es un buen clasificador? ¿Por qué?

Forma	Color	Olor	Comestible
C	Negro	B	NO
D	Negro	B	NO
C	Blanco	B	SI
D	Negro	B	NO
D	Verde	B	NO
C	Marrón	B	SI

2017-05-22

#### 4.- [3 puntos]

- Dado el siguiente conjunto de ejemplos con un atributo real y una variable objetivo que toma dos valores (+,-)



Utilizando como conjunto de test los ejemplos (-3.5, +), (-0.8, +) y (-0.4,-) y como conjunto de entrenamiento el resto de ejemplos, calcula razonadamente el porcentaje de error que se obtiene con K-NN, K=3.

**b)** Dado un conjunto de datos con 7 variables binarias, Rain, Wind, Summer, Winter, Day y Night que toman el valor “yes” o “no” y Flight\_Delay que toma los valores “Delayed” y “not Delayed”, considera la siguiente tabla de datos, que explica cómo se comporta la variable Flight\_Delay con respecto a las otras 6.

Feature	Value = yes	Value = no
Rain	Delayed - 30, not Delayed - 10	Delayed - 10, not Delayed - 30
Wind	Delayed - 25, not Delayed - 15	Delayed - 15, not Delayed - 25
Summer	Delayed - 5, not Delayed - 35	Delayed - 35, not Delayed - 5
Winter	Delayed - 20, not Delayed - 10	Delayed - 20, not Delayed - 30
Day	Delayed - 20, not Delayed - 20	Delayed - 20, not Delayed - 20
Night	Delayed - 15, not Delayed - 10	Delayed - 25, not Delayed - 30

Si construimos un árbol de decisión utilizando c4.5, responde razonadamente a la pregunta ¿cuál es la variable raíz?

**c)** Supongamos que tenemos 50 libros de minería de datos de un total de 200 libros en una biblioteca. Si un clasificador predice que 10 libros son de minería de datos, pero solo 5 de ellos lo son realmente, ¿Cuál es la precisión y a cobertura?

2016-12-19

4.- [3 puntos]

**a) [1.25 punto]** Dado un conjunto de entrenamiento con 150 ejemplos y tres variables, la variable Color que puede tomar los valores Rojo, Azul o Amarillo, la variable Tamaño que toma los valores Grande o Pequeño y la variable a predecir, COMPRA; que toma los valores SÍ o NO. Suponiendo que :

- Hay 50 ejemplos donde Color=Rojo, 50 para los que Color=Azul y 50 para los que Color=Amarillo.
- Hay 20 ejemplos para los que Tamaño=Grande y 130 para los que Tamaño=Pequeño.
- Hay 100 ejemplos de la clase COMPRA= SI y 50 de la clase COMPRA=NO
- $IG(\text{Color})=IG(\text{Tamaño})$

Contesta razonadamente: ¿Qué variable selecciona como raíz el método C4.5? ¿Es el 70% un porcentaje de acierto aceptable para el método C4.5?

**b) [0.5 puntos]** Considera las variables Altura y Tipo, donde Tipo es la variable a predecir y Altura es una variable continua utilizada para predecir el Tipo. Dados los siguientes 6 ejemplos {Ej1, Ej2, ..., Ej6}

	Ej1	Ej2	Ej3	Ej4	Ej5	Ej6
Altura	20	25	35	50	80	150
Tipo	Árbol	Árbol	Arbusto	Arbusto	Árbol	Árbol

¿Cuántos Split Points puedes obtener para la variable Altura? Calcula dichos Split Points.

**c) [0.75]** Proporciona una matriz de confusión para el problema de predecir si la inversión de una determinada empresa debe ser “alta”, “baja” o “media” de modo que el Recall (Cobertura) de la clase “alta” sea del 75% y la el número de Falsos Positivos (FP) de la clase “baja” sea 50. NOTA: En la matriz de confusión no puede haber 0’s.



2016-07-01

**1.-a) [2 puntos]** La NASA quiere construir un sistema para discriminar entre Humanos (H) y Marcianos (M) basándose en las siguientes características: Green  $\in \{N, Y\}$ , Legs  $\in \{2, 3\}$ , Height  $\in \{S, T\}$ , Smelly  $\in \{N, Y\}$ . El conjunto de entrenamiento es el siguiente:

<i>Ejemplo</i>	<i>Especie</i>	<i>Green</i>	<i>Legs</i>	<i>Height</i>	<i>Smelly</i>
1	M	N	3	S	Y
2	M	Y	2	T	N
3	M	Y	3	T	N
4	M	N	2	S	Y
5	M	Y	3	T	N
6	H	N	2	T	Y
7	H	N	2	S	N
8	H	N	2	T	N
9	H	Y	2	S	N
10	H	N	2	T	Y

Construye un árbol de decisión para este problema siguiendo los pasos del método ID3.

**b) [0,5 puntos]** Pon un ejemplo de una matriz de confusión para un problema de clasificación en TRES clases {A,B,C} de modo que la precisión para la clase A sea 1 y el recall para la clase B sea 0.5.

**c) [0,5 puntos]** Considera el siguiente conjunto de datos:

<i>Ejemplo</i>	<i>x</i>	<i>y</i>	<i>Clase</i>
1	-1	1	-
2	0	1	+
3	0	2	-
4	1	-1	-
5	1	0	+
6	1	2	+
7	2	2	-
8	2	3	+

- ¿Cuál es la predicción para el punto (1,1) con 3-NN y 5-NN?
- ¿Y cuando ponderas por el inverso de la distancia?

2016-05-23

**1.- [3 puntos]** Aprendizaje

**a) [1 punto]** Considera el siguiente conjunto de ejemplos utilizado para aprender a identificar si una seta es venenosa o no

<i>Ejemplo</i>	<i>EsDura</i>	<i>EsOlorosa</i>	<i>TieneManchas</i>	<i>EsRegular</i>	<i>EsVenenosa</i>
A	0	0	0	0	0
B	0	0	1	0	0
C	1	1	0	1	0
D	1	0	0	1	1
E	0	1	1	0	1
F	0	0	1	1	1
G	0	0	0	1	1
H	1	1	0	0	1

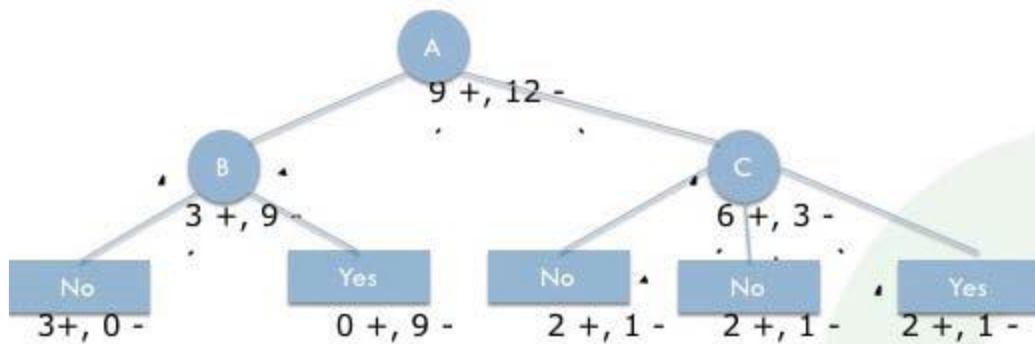
Considerando sólo los ejemplos de la A a la H, selecciona la raíz del árbol de decisión cuando se utiliza como métrica el índice de Gini.

**b) [0,5 puntos]** Dado un conjunto de 1500 ejemplos distribuidos en dos clases de modo que 1300 son de la clase "APRUEBA" y "200" de la clase "SUSPENDE", proporciona RAZONADAMENTE una cota inferior del porcentaje de error que debería tener cualquier clasificador.

**c) [0,75 punto]** Dada la siguiente matriz de confusión, calcula la precisión y el recall del clasificador utilizado para obtener dicha matriz.

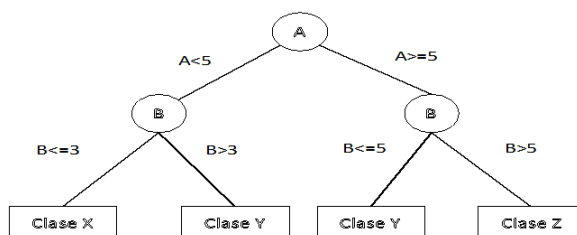
	<i>A – predicho</i>	<i>B – predicho</i>	<i>C – predicho</i>
<i>A – real</i>	23	70	1
<i>B – real</i>	4	5	0
<i>C – real</i>	6	2	100

**d) [0,75 puntos]** Dado el siguiente árbol de decisión donde se especifica en cada nodo cuantos ejemplos hay de la clase "+" y cuantos de la clase "-", si aplicáramos un mecanismo de poda, ¿podríamos eliminar alguna rama del árbol obteniendo al menos el mismo porcentaje de acierto? ¿Por qué?



2016-01-11

**1.- 1.5 puntos]** Dado el siguiente árbol de decisión y el conjunto de test de la tabla adjunta:



Ejemplo	A	B	CLASE
EJ1	3	3	Y
EJ2	2	2	X
EJ3	5	2	X
EJ4	7	6	Y
EJ5	7	7	Y
EJ6	8	6	Z
EJ7	2	4	Y
EJ8	8	9	Z
EJ9	5	5	Z
EJ10	1	4	X

- Calcula la matriz de confusión asociada al modelo descrito por el árbol de decisión.
- Calcula los valores de TP y TN para cada una de las clases.
- ¿Qué clase tiene una precisión mayor? Justifica tu respuesta.

2.- [1.5 puntos] Dado el conjunto de ejemplos  $S=\{EJ1, EJ2, \dots, EJ9\}$  descrito a continuación

Examples	ATT1	ATT2	ATT3	Class
EJ1	0	250	36	A
EJ2	10	150	34	B
EJ3	4	20	1	B
EJ4	6	78	8	B
EJ5	2	90	10	A
EJ6	1	170	70	A
EJ7	6	200	45	A
EJ8	8	160	41	B
EJ9	10	180	38	A

a) Completa la siguiente tabla de entropías de las particiones asociadas a los atributos ATT1, ATT2, ATT3:

Entropía	YES	NO
$ATT1 \leq 5$	0.811	0.971
$ATT2 \leq 160$	0.722	
$ATT3 \leq 40$		0.918

b) Sabiendo que  $E(S)=0.991$ , y utilizando los datos de la tabla del apartado a), calcula el atributo más informativo según la GANANCIA DE INFORMACIÓN.

c) Utilizando el conjunto  $S=\{EJ1, EJ2, \dots, EJ8\}$  como entrenamiento, clasifica el ejemplo EJ9 utilizando 3NN con la distancia euclídea. ¿Es un acierto? Se debe calcular la distancia euclídea del ejemplo EJ9 al resto.

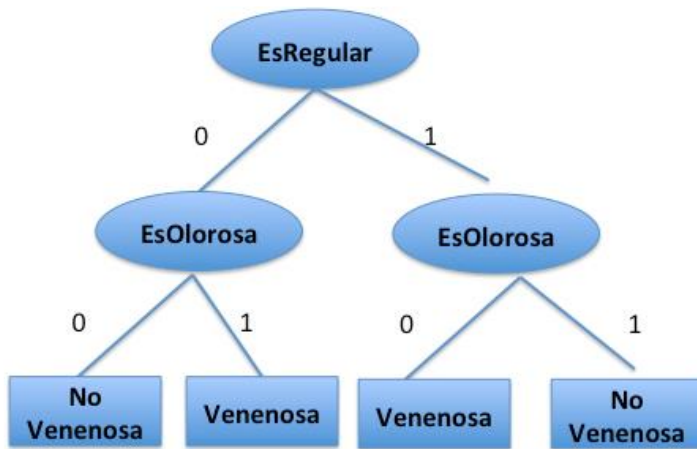
2015-06-11

4.- Considera el siguiente conjunto de ejemplos utilizado para aprender a identificar si una seta es venenosa o no

Ejemplo	EsDura	EsOlorosa	TieneManchas	EsRegular	EsVenenosa
A	0	0	0	0	0
B	0	0	1	0	0
C	1	1	0	1	0
D	1	0	0	1	1
E	0	1	1	0	1
F	0	0	1	1	1
G	0	0	0	1	1
H	1	1	0	0	1
U	1	1	1	1	1
V	0	1	0	1	0
W	1	1	0	0	1
X	1	0	1	1	0
Y	0	0	1	1	1
Z	1	1	0	1	1

a) [1 punto] Considerando sólo los ejemplos de la A a la H, ¿Cuál es el atributo que proporciona mayor Ganancia de Información? (Cálculalo detalladamente).

**b) [1 punto]** Suponiendo que el árbol siguiente fuera una solución proporcionada por algún algoritmo y considerando los ejemplos U a Z como conjunto de test, ¿Cuál sería la precisión obtenida con dicho algoritmo?



<i>Ejemplo</i>	<i>EsOlorosa</i>	<i>EsRegular</i>	<i>EsVenenosa</i>	<i>Predicho</i>	
U	1	1	1	0	FN
V	1	1	0	0	TN
W	1	0	1	1	TP
X	0	1	0	1	FP
Y	0	1	1	1	TP
Z	1	1	1	0	FN

**c) [1 punto]** Considera ahora la tabla resultante de eliminar la columna correspondiente al atributo *EsRegular* en la tabla anterior. Aplicando el algoritmo IB3 sobre el conjunto de datos de la A a la H de la nueva tabla, y teniendo en cuenta que la clase objetivo es “EsVenenosa”, ¿A qué clase pertenece el ejemplo V?

2015-01-21

### 3.- [3 puntos. Aprendizaje]

**a)** Sea un algoritmo de aprendizaje automático cualquiera. Supongamos que un clasificador que utiliza a este algoritmo se usa para identificar correo spam (Problema A) y para clasificar pacientes como diabéticos o no diabéticos (Problema B). Para probar la bondad del clasificador sobre dichos conjuntos se utiliza:

Problema A: 1000 ejemplos de test, 950 de la clase spam y 50 de la clase no spam.

Problema B: 650 ejemplos de test, 300 ejemplos de la clase diabético, 350 de la clase No\_diabético.

**a1) [0.5 puntos]** Si el porcentaje de acierto obtenido con el algoritmo para el Problema A es del 92%, ¿puede decirse que el rendimiento del algoritmo para el Problema A es bueno? ¿por qué? Y si para el Problema B el porcentaje de acierto es de 70%, puede considerarse que el algoritmo de aprendizaje obtiene buenos resultados para el problema B? ¿por qué?

**a2) [0.25 puntos]** ¿Cómo debería ser la matriz de confusión para que la precisión fuese del 50% para el Problema A?

**b)** Considera el siguiente conjunto de ejemplos utilizado para aprender a identificar si una seta es venenosa o no

<i>Ejemplo</i>	<i>EsDura</i>	<i>EsOlorosa</i>	<i>TieneManchas</i>	<i>EsRegular</i>	<i>EsVenenosa</i>
A	0	0	0	0	0
B	0	0	1	0	0
C	1	1	0	1	0
D	1	0	0	1	1
E	0	1	1	0	1
F	0	0	1	1	1
G	0	0	0	1	1
H	1	1	0	0	1
U	1	1	1	1	¿
V	0	1	0	1	¿
W	1	1	0	0	¿

Considerando sólo los ejemplos de la A a la H, responde a las siguientes cuestiones

**b1) [0.25 puntos]** ¿Qué atributo debería seleccionarse como la raíz del árbol de decisión que genera C4.5? ¿Por qué?

**b2) [1 punto]** Construye RAZONADAMENTE un árbol del decisión utilizando como métrica Gain Ratio (la que utiliza C4.5). Profundiza en el árbol hasta que en cada nodo todos los ejemplos sean de la misma clase O haya como máximo 2 ejemplos. Para ello puedes utilizar la información contenida en la siguiente tabla, donde cada valor representa  $\text{GainRatio}(Y/X)$  (Por ejemplo el 0.2368 que aparece en la fila 1 y columna 2 es  $\text{GainRatio}(\text{EsOlorosa}/\text{EsDura}=0)$ )

<i>X</i>	<i>Y</i>			
	<i>EsDura</i>	<i>EsOlorosa</i>	<i>TieneManchas</i>	<i>EsRegular</i>
<i>EsDura=0</i>		0.2368	0.0206	0.4325
<i>EsOlorosa=0</i>	0.2328		0.0206	1
<i>TieneManchas=0</i>	0.0206	0.0206		0.0206
<i>EsRegular=0</i>	0.384	1	0	
<i>EsDura=1</i>		0.274	0	0.274
<i>EsOlorosa=1</i>	0.274		0.274	1
<i>TieneManchas=1</i>	0	0.274		0.274
<i>EsRegular=1</i>	0.311	1	0.151	

**b3) [0.5 puntos]** Utilizando el árbol de decisión que has obtenido, clasifica los ejemplos U y W.

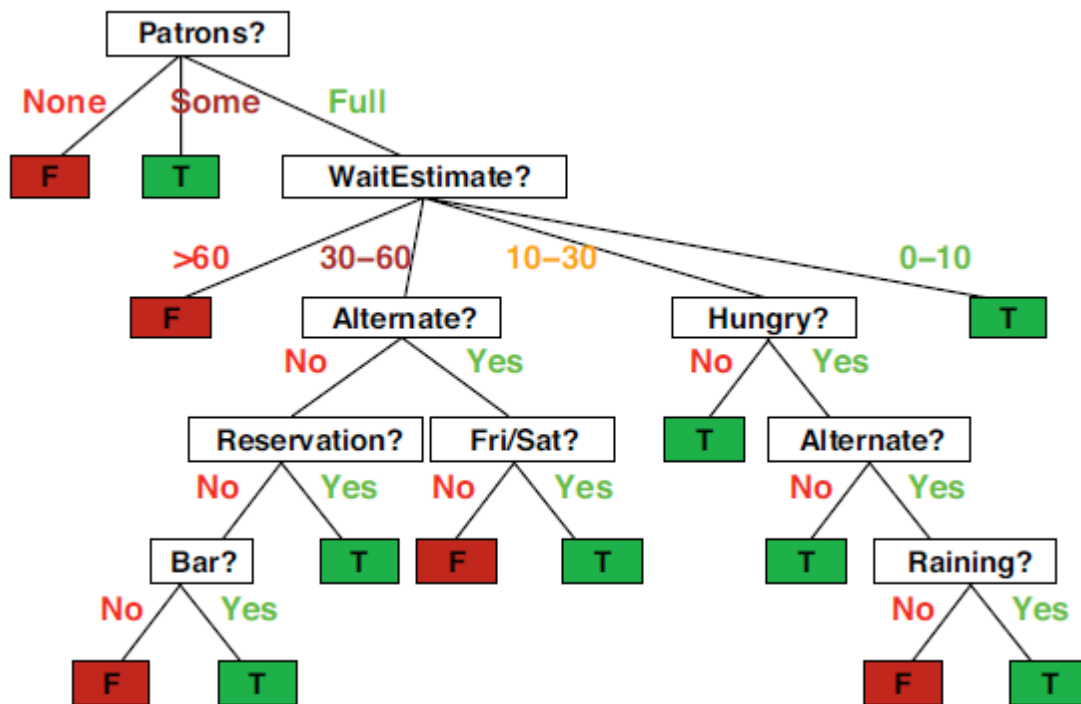
**b4) [0.5 puntos]** Aplicando el algoritmo IB3, ¿A qué clase pertenece el ejemplo V?

2014-06-14

**4.- [3 puntos]** Dado el siguiente conjunto de ejemplos para el problema del restaurante

Example	Attributes											Target
	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	Wait	
X <sub>1</sub>	T	F	F	T	Some	\$\$\$	F	T	French	0-10	T	
X <sub>2</sub>	T	F	F	T	Full	\$	F	F	Thai	30-60	F	
X <sub>3</sub>	F	T	F	F	Some	\$	F	F	Burger	0-10	T	
X <sub>4</sub>	T	F	T	T	Full	\$	F	F	Thai	10-30	T	
X <sub>5</sub>	T	F	T	F	Full	\$\$\$	F	T	French	>60	F	
X <sub>6</sub>	F	T	F	T	Some	\$\$	T	T	Italian	0-10	T	
X <sub>7</sub>	F	T	F	F	None	\$	T	F	Burger	0-10	F	
X <sub>8</sub>	F	F	F	T	Some	\$\$	T	T	Thai	0-10	T	
X <sub>9</sub>	F	T	T	F	Full	\$	T	F	Burger	>60	F	
X <sub>10</sub>	T	T	T	T	Full	\$\$\$	F	T	Italian	10-30	F	
X <sub>11</sub>	F	F	F	F	None	\$	F	F	Thai	0-10	F	
X <sub>12</sub>	T	T	T	T	Full	\$	F	F	Burger	30-60	T	

Un determinado algoritmo de aprendizaje ha calculado el siguiente árbol de decisión:



A la vista del resultado, nos planteamos dos cuestiones:

- ¿Es posible que el “determinado algoritmo” sea el conocido constructor de árboles de decisión que vimos en clase, pero utilizando un heurístico diferente al de máxima ganancia? Para responder a esta pregunta hay que identificar las partes del árbol anterior que nunca serían generadas por el algoritmo de construcción de árboles de decisión.
- ¿Cómo podemos reducir el tamaño de este árbol de 6 niveles de forma que se quede en 3 niveles, aunque se pierda calidad de clasificación? Para responder a esta pregunta lo que se pide es definir una estrategia de poda razonable que utilice solamente la información del árbol (y no la de los ejemplos) y aplicarla para reducir el árbol anterior.

Luego se debe indicar el árbol resultante y el porcentaje de aciertos y fallos en la clasificación del conjunto de entrenamiento anterior.