Inteligencia Artificial II (ejercicio 1) **(11–Junio–2007)**

Apellidos: ....

Nombre: ....

Ejercicio 1 : 2 Contestar las siguientes cuestiones **de manera clara y concisa**, usando para ello el espacio en blanco que aparece a continuación de cada una de ellas:

(a) ¿Por qué el algoritmo FOIL devuelve un conjunto de reglas que cubre a todos los ejemplos positivos y a ninguno de los negativos?

(b) Supongamos un conjunto de atributos A_1, A_2,..., A_n cuyos valores determinan un valor de clasificación de un conjunto V y sea D un conjunto de entrenamiento. Si tenemos una instancia nueva ¿Qué valor de V le asigna un clasificador *naive* Bayes? ¿En qué sentido es ése el valor más probable? ¿Por qué el clasificador se denomina *naive*? ¿Cómo se estiman las probabilidades necesarias para el cálculo del valor de clasificación? ¿Qué propiedad tienen esas estimaciones?

(c) Describir, en pseudocódigo, el algoritmo de ponderación por verosimilitud.

(d) ¿Cuál es el objetivo del algoritmo de retropropagación? ¿Cómo actualiza en cada iteración los pesos de la red?

Inteligencia Artificial II (ejercicio 2) (11–Junio–2007)

Apellidos: ....

Nombre: ....

Ejercicio 2 : 1.5 La siguiente tabla muestra información sobre setas, indicando si son comestibles o no en función de algunas características: el `Color`, el `Tamaño` del pie, la `Forma` del sombrero, el `Entorno` en el que se presenta y la forma en que se `Agrupa` con otras setas:

Ej.	Color	Tamaño	Forma	Entorno	Agrupación	Comestible
Ej_1	rojo	mediano	plana	pinar	aislada	si
Ej_2	blanco	grande	plana	pinar	racimo	si
Ej_3	rojo	pequeño	plana	pradera	racimo	si
Ej_4	blanco	pequeño	cóncava	cueva	grupo	si
Ej_5	marrón	grande	convexa	pinar	grupo	si
Ej_6	rojo	grande	cóncava	cueva	grupo	no
Ej_7	marrón	mediano	convexa	pradera	aislada	no
Ej_8	marrón	mediano	plana	pinar	racimo	no
Ej_9	blanco	mediano	convexa	pinar	aislada	no
Ej_{10}	rojo	pequeño	convexa	cueva	grupo	no

Se pide:

1. Aplicar (detallando cada uno de los pasos realizados) el **algoritmo de cobertura** para encontrar, a partir de este conjunto de entrenamiento, un conjunto de reglas que nos permita clasificar nuevas instancias. Determinar las reglas para clasificar instancias tanto positivas como negativas. Según lo aprendido ¿hay algún atributo irrelevante para realizar esta clasificación?

Solución:

Aprendemos reglas para clasificar setas comestibles

- **Consideramos la siguiente regla genérica:**

  R1: Si ?
  Entonces **`Comestible`**$= +$

  Alternativas para ? **y frecuencia relativa de la regla resultante:**

`Color=rojo`	2/4
`Color=blanco`	2/3
`Color=marrón`	1/3
`Tamaño=pequeño`	2/3
`Tamaño=mediano`	1/4
`Tamaño=grande`	2/3
`Forma=convexa`	1/4
`Forma=plana`	3/4
`Forma=cóncava`	1/2
`Entorno=pinar`	3/5
`Entorno=cueva`	1/3
`Entorno=pradera`	1/2
`Agrupación=racimo`	2/3
`Agrupación=grupo`	2/4
`Agrupación=aislada`	1/3

  La mejor de todas las opciones es `Forma=plana` con frecuencia relativa 3/4. La regla parcialmente aprendida es:

R1: Si `Forma`$=$`plana` $\wedge$?
Entonces `Comestible`$= +$

Alternativas para ? y frecuencia relativa de la regla resultante:

`Color`=`rojo`	$2/2$
`Color`=`blanco`	$1/1$
`Color`=`marrón`	$0/1$
`Tamaño`=`pequeño`	$1/1$
`Tamaño`=`mediano`	$1/2$
`Tamaño`=`grande`	$1/1$
`Entorno`=`pinar`	$2/3$
`Entorno`=`cueva`	0
`Entorno`=`pradera`	$1/1$
`Agrupación`=`racimo`	$2/3$
`Agrupación`=`grupo`	0
`Agrupación`=`aislada`	$1/1$

La mejor de todas las opciones es `Color`=`rojo` con frecuencia relativa $2/2$, pues clasifica mayor número de ejemplos positivos frente a otras opciones con la misma frecuencia relativa. Llegado a este punto la regla clasifica correctamente todos los ejemplos que cubre. La regla finalmente aprendida es:

R1: Si `Forma`$=$`plana` $\wedge$ `Color`$=$`rojo`
Entonces `Comestible`$= +$

- La regla anterior no cubre todos los ejemplos positivos, así que tendremos que generar más reglas. Consideramos ahora el conjunto de ejemplos $\{Ej_2, Ej_4, Ej_5, Ej_6, Ej_7, Ej_8, Ej_9, Ej_{10}\}$ y la regla:

R2: Si ?
Entonces `Comestible`$= +$

Alternativas para ? y frecuencia relativa de la regla resultante:

`Color`=`rojo`	$0/2$
`Color`=`blanco`	$2/3$
`Color`=`marrón`	$1/3$
`Tamaño`=`pequeño`	$1/2$
`Tamaño`=`mediano`	$0/3$
`Tamaño`=`grande`	$2/3$
`Forma`=`convexa`	$1/4$
`Forma`=`plana`	$1/2$
`Forma`=`cóncava`	$1/2$
`Entorno`=`pinar`	$2/4$
`Entorno`=`cueva`	$1/3$
`Entorno`=`pradera`	$0/1$
`Agrupación`=`racimo`	$1/2$
`Agrupación`=`grupo`	$2/4$
`Agrupación`=`aislada`	$0/2$

Una de las mejores opciones es `Color`=`blanco` con frecuencia relativa $2/3$. La regla parcialmente aprendida es:

R2: Si `Color`$=$`blanco` $\wedge$?
Entonces `Comestible`$= +$

Alternativas para ? y frecuencia relativa de la regla resultante:

`Tamaño`=`pequeño`	$1/1$
`Tamaño`=`mediano`	$0/1$
`Tamaño`=`grande`	$1/1$
`Forma`=`convexa`	$0/1$
`Forma`=`plana`	$1/1$

`Forma`=`cóncava`	$1/1$
`Entorno`=`pinar`	$1/2$
`Entorno`=`cueva`	$1/1$
`Entorno`=`pradera`	0
`Agrupación`=`racimo`	$1/1$
`Agrupación`=`grupo`	$1/1$
`Agrupación`=`aislada`	$0/1$

Una de las mejores opciones es `Tamaño`=`pequeño` con frecuencia relativa $1/1$. Llegado a este punto la regla clasifica correctamente todos los ejemplos que cubre. La regla finalmente aprendida es:

R2: Si `Color`=`blanco` $\wedge$ `Tamaño`=`pequeño`
Entonces `Comestible`$= +$

- Las reglas anteriores no cubren todos los ejemplos positivos, así que tendremos que generar más reglas. Consideramos ahora el conjunto de ejemplos $\{Ej_2, Ej_5, Ej_6, Ej_7, Ej_8, Ej_9, Ej_{10}\}$ y la regla:

R3: Si ?
Entonces `Comestible`$= +$

Alternativas para ? y frecuencia relativa de la regla resultante:

`Color`=`rojo`	$0/2$
`Color`=`blanco`	$1/2$
`Color`=`marrón`	$1/3$
`Tamaño`=`pequeño`	$0/1$
`Tamaño`=`mediano`	$0/3$
`Tamaño`=`grande`	$2/3$
`Forma`=`convexa`	$1/4$
`Forma`=`plana`	$1/2$
`Forma`=`cóncava`	$0/1$
`Entorno`=`pinar`	$2/4$
`Entorno`=`cueva`	$0/2$
`Entorno`=`pradera`	$0/1$
`Agrupación`=`racimo`	$1/2$
`Agrupación`=`grupo`	$1/3$
`Agrupación`=`aislada`	$0/2$

La mejor de todas las opciones es `Tamaño`=`grande` con frecuencia relativa $2/3$. La regla parcialmente aprendida es:

R3: Si `Tamaño`=`grande` $\wedge$?
Entonces `Comestible`$= +$

Alternativas para ? y frecuencia relativa de la regla resultante:

`Color`=`rojo`	$0/1$
`Color`=`blanco`	$1/1$
`Color`=`marrón`	$1/1$
`Forma`=`convexa`	$1/1$
`Forma`=`plana`	$1/1$
`Forma`=`cóncava`	$0/1$
`Entorno`=`pinar`	$2/2$
`Entorno`=`cueva`	$0/1$
`Entorno`=`pradera`	0
`Agrupación`=`racimo`	$1/1$
`Agrupación`=`grupo`	$1/2$
`Agrupación`=`aislada`	0

La mejor de todas las opciones es `Entorno`=`pinar` con frecuencia relativa $2/2$, pues clasifica mayor número de ejemplos positivos frente a otras opciones con la misma frecuencia relativa. Llegado a este punto la regla clasifica correctamente todos los ejemplos que cubre. La regla finalmente aprendida es:

R3: Si **Tamaño**=**grande** $\wedge$ **Entorno**=**pinar**
Entonces **Comestible**$= +$

Las reglas aprendidas para clasificar setas comestibles son las siguientes:

R1: Si **Forma**=**plana** $\wedge$ **Color**=**rojo**
Entonces **Comestible**$= +$
R2: Si **Color**=**blanco** $\wedge$ **Tamaño**=**pequeño**
Entonces **Comestible**$= +$
R3: Si **Tamaño**=**grande** $\wedge$ **Entorno**=**pinar**
Entonces **Comestible**$= +$

Aprendemos reglas para clasificar setas no comestibles

- Consideramos la siguiente regla genérica:

R4: Si ?
Entonces **Comestible**$= -$

Alternativas para ? y frecuencia relativa de la regla resultante:

Color=**rojo**	$2/4$
Color=**blanco**	$1/3$
Color=**marrón**	$2/3$
Tamaño=**pequeño**	$1/3$
Tamaño=**mediano**	$3/4$
Tamaño=**grande**	$1/3$
Forma=**convexa**	$3/4$
Forma=**plana**	$1/4$
Forma=**cóncava**	$1/2$
Entorno=**pinar**	$2/5$
Entorno=**cueva**	$2/3$
Entorno=**pradera**	$1/2$
Agrupación=**racimo**	$1/3$
Agrupación=**grupo**	$2/4$
Agrupación=**aislada**	$2/3$

Una de las mejores opciones es **Tamaño**=**mediano** con frecuencia relativa $3/4$. La regla parcialmente aprendida es:

R4: Si **Tamaño**=**mediano** $\wedge$?
Entonces **Comestible**$= -$

Alternativas para ? y frecuencia relativa de la regla resultante:

Color=**rojo**	$0/1$
Color=**blanco**	$1/1$
Color=**marrón**	$2/2$
Forma=**convexa**	$2/2$
Forma=**plana**	$1/2$
Forma=**cóncava**	0
Entorno=**pinar**	$2/3$
Entorno=**cueva**	0
Entorno=**pradera**	$1/1$
Agrupación=**racimo**	$1/1$
Agrupación=**grupo**	0
Agrupación=**aislada**	$2/3$

Una de las mejores opciones es **Color**=**marrón** con frecuencia relativa $2/2$, pues clasifica mayor número de ejemplos positivos frente a otras opciones con la misma frecuencia relativa. Llegado a este punto la regla clasifica correctamente todos los ejemplos que cubre. La regla finalmente aprendida es:

R4: Si `Tamaño`=`mediano` $\wedge$ `Color`=`marrón`
Entonces `Comestible`$= -$

- **La regla anterior no cubre todos los ejemplos negativos, así que tendremos que generar más reglas. Consideramos ahora el conjunto de ejemplos $\{Ej_1, Ej_2, Ej_3, Ej_4, Ej_5, Ej_6, Ej_9, Ej_{10}\}$ y la regla:**

R5: Si $?$
Entonces `Comestible`$= -$

Alternativas para $?$ y frecuencia relativa de la regla resultante:

`Color`=`rojo`	$2/4$
`Color`=`blanco`	$1/3$
`Color`=`marrón`	$0/1$
`Tamaño`=`pequeño`	$1/3$
`Tamaño`=`mediano`	$1/2$
`Tamaño`=`grande`	$1/3$
`Forma`=`convexa`	$2/3$
`Forma`=`plana`	$0/3$
`Forma`=`cóncava`	$1/2$
`Entorno`=`pinar`	$1/4$
`Entorno`=`cueva`	$2/3$
`Entorno`=`pradera`	$0/1$
`Agrupación`=`racimo`	$0/2$
`Agrupación`=`grupo`	$2/4$
`Agrupación`=`aislada`	$1/2$

Una de las mejores opciones es `Entorno`=`cueva` con frecuencia relativa $2/3$. La regla parcialmente aprendida es:

R5: Si `Entorno`=`cueva` $\wedge$ $?$
Entonces `Comestible`$= -$

Alternativas para $?$ y frecuencia relativa de la regla resultante:

`Color`=`rojo`	$2/2$
`Color`=`blanco`	$0/1$
`Color`=`marrón`	0
`Tamaño`=`pequeño`	$1/2$
`Tamaño`=`mediano`	0
`Tamaño`=`grande`	$1/1$
`Forma`=`convexa`	$1/1$
`Forma`=`plana`	0
`Forma`=`cóncava`	$1/2$
`Agrupación`=`racimo`	0
`Agrupación`=`grupo`	$2/3$
`Agrupación`=`aislada`	0

La mejor de todas las opciones es `Color`=`rojo` con frecuencia relativa $2/2$, pues clasifica mayor número de ejemplos positivos frente a otras opciones con la misma frecuencia relativa. Llegado a este punto la regla clasifica correctamente todos los ejemplos que cubre. La regla finalmente aprendida es:

R5: Si `Entorno`=`cueva` $\wedge$ `Color`=`rojo`
Entonces `Comestible`$= -$

- **Las reglas anteriores no cubren todos los ejemplos positivos, así que tendremos que generar más reglas. Consideramos ahora el conjunto de ejemplos $\{Ej_1, Ej_2, Ej_3, Ej_4, Ej_5, Ej_9\}$ y la regla:**

R6: Si $?$
Entonces `Comestible`$= -$

Alternativas para ? y frecuencia relativa de la regla resultante:

`Color`=`rojo`	$0/2$
`Color`=`blanco`	$1/3$
`Color`=`marrón`	$0/1$
`Tamaño`=`pequeño`	$0/2$
`Tamaño`=`mediano`	$1/2$
`Tamaño`=`grande`	$0/2$
`Forma`=`convexa`	$1/2$
`Forma`=`plana`	$0/3$
`Forma`=`cóncava`	$0/1$
`Entorno`=`pinar`	$1/4$
`Entorno`=`cueva`	$0/1$
`Entorno`=`pradera`	$0/1$
`Agrupación`=`racimo`	$0/2$
`Agrupación`=`grupo`	$0/2$
`Agrupación`=`aislada`	$1/2$

Una de las mejores opciones es `Tamaño`=`mediano` con frecuencia relativa $1/2$. La regla parcialmente aprendida es:

R6: Si `Tamaño`=`mediano` $\wedge$?
Entonces `Comestible`= $-$

Alternativas para ? y frecuencia relativa de la regla resultante:

`Color`=`rojo`	$0/1$
`Color`=`blanco`	$1/1$
`Color`=`marrón`	0
`Forma`=`convexa`	$1/1$
`Forma`=`plana`	$0/1$
`Forma`=`cóncava`	0
`Entorno`=`pinar`	$1/2$
`Entorno`=`cueva`	0
`Entorno`=`pradera`	0
`Agrupación`=`racimo`	0
`Agrupación`=`grupo`	0
`Agrupación`=`aislada`	$1/2$

Una de las mejores opciones es `Color`=`blanco` con frecuencia relativa $1/1$. Llegado a este punto la regla clasifica correctamente todos los ejemplos que cubre. La regla finalmente aprendida es:

R6: Si `Tamaño`=`mediano` $\wedge$ `Color`=`blanco`
Entonces `Comestible`= $-$

Las reglas aprendidas para clasificar setas no comestibles son las siguientes:

R4: Si `Tamaño`=`mediano` $\wedge$ `Color`=`marrón`
Entonces `Comestible`= $-$
R5: Si `Entorno`=`cueva` $\wedge$ `Color`=`rojo`
Entonces `Comestible`= $-$
R6: Si `Tamaño`=`mediano` $\wedge$ `Color`=`blanco`
Entonces `Comestible`= $-$

Según lo aprendido el atributo `Agrupación` no es necesario para realizar la clasificación

2. Clasificar las siguientes instancias utilizando el conjunto de reglas aprendido:

Ej.	Color	Tamaño	Forma	Entorno	Agrupación
I_1	blanco	pequeño	cóncava	pradera	racimo
I_2	rojo	mediano	plana	cueva	aislada
I_3	marrón	mediano	plana	pinar	racimo
I_4	rojo	mediano	cóncava	pinar	grupo

Solución:

- **La instancia I_1 es clasificada positivamente por la regla R_2.**
- **La instancia I_2 es clasificada positivamente por la regla R_1.**
- **La instancia I_3 es clasificada negativamente por la regla R_4.**
- **La instancia I_4 no es clasificada por ninguna regla.**

3. Usando todos los ejemplos de la tabla como conjunto de entrenamiento, utilizar un clasificador *naive bayes* para clasificar las instancias anteriores.

Solución:

Según el clasificador Naive Bayes:

$$\textsc{Comestible}_{NB} = argmax_{v \in \{+,-\}} P(\textsc{Comestible} = v) \\ P(\textsc{Color} = v1|\textsc{Comestible} = v) \\ P(\textsc{Tamaño} = v2|\textsc{Comestible} = v) \\ P(\textsc{Forma} = v3|\textsc{Comestible} = v) \\ P(\textsc{Entorno} = v4|\textsc{Comestible} = v) \\ P(\textsc{Agrupación} = v5|\textsc{Comestible} = v)$$

- **Para la instancia I_1:**
  $P(\textsc{Comestible} = +) = \frac{5}{10}$
  $P(\textsc{Color} = \textsc{blanco}|\textsc{Comestible} = +) = \frac{2}{5}$
  $P(\textsc{Tamaño} = \textsc{pequeño}|\textsc{Comestible} = +) = \frac{2}{5}$
  $P(\textsc{Forma} = \textsc{cóncava}|\textsc{Comestible} = +) = \frac{1}{5}$
  $P(\textsc{Entorno} = \textsc{pradera}|\textsc{Comestible} = +) = \frac{1}{5}$
  $P(\textsc{Agrupación} = \textsc{racimo}|\textsc{Comestible} = +) = \frac{2}{5}$
  Producto $= \frac{5}{10}\frac{2}{5}\frac{2}{5}\frac{1}{5}\frac{1}{5}\frac{2}{5}$
  $P(\textsc{Comestible} = -) = \frac{5}{10}$
  $P(\textsc{Color} = \textsc{blanco}|\textsc{Comestible} = -) = \frac{1}{5}$
  $P(\textsc{Tamaño} = \textsc{pequeño}|\textsc{Comestible} = -) = \frac{1}{5}$
  $P(\textsc{Forma} = \textsc{cóncava}|\textsc{Comestible} = -) = \frac{1}{5}$
  $P(\textsc{Entorno} = \textsc{pradera}|\textsc{Comestible} = -) = \frac{1}{5}$
  $P(\textsc{Agrupación} = \textsc{racimo}|\textsc{Comestible} = -) = \frac{1}{5}$
  Producto $= \frac{5}{10}\frac{1}{5}\frac{1}{5}\frac{1}{5}\frac{1}{5}\frac{1}{5}$
  Por tanto, según el clasificador Naive Bayes la seta es comestible.
- **Para la instancia I_2:**
  $P(\textsc{Comestible} = +) = \frac{5}{10}$
  $P(\textsc{Color} = \textsc{rojo}|\textsc{Comestible} = +) = \frac{2}{5}$
  $P(\textsc{Tamaño} = \textsc{mediano}|\textsc{Comestible} = +) = \frac{1}{5}$
  $P(\textsc{Forma} = \textsc{plana}|\textsc{Comestible} = +) = \frac{3}{5}$
  $P(\textsc{Entorno} = \textsc{cueva}|\textsc{Comestible} = +) = \frac{1}{5}$
  $P(\textsc{Agrupación} = \textsc{aislada}|\textsc{Comestible} = +) = \frac{1}{5}$
  Producto $= \frac{5}{10}\frac{2}{5}\frac{1}{5}\frac{3}{5}\frac{1}{5}\frac{1}{5}$
  $P(\textsc{Comestible} = -) = \frac{5}{10}$
  $P(\textsc{Color} = \textsc{rojo}|\textsc{Comestible} = -) = \frac{2}{5}$
  $P(\textsc{Tamaño} = \textsc{mediano}|\textsc{Comestible} = -) = \frac{3}{5}$
  $P(\textsc{Forma} = \textsc{plana}|\textsc{Comestible} = -) = \frac{1}{5}$

$P(\textsc{Entorno} = \textsc{cueva}|\textsc{Comestible} = -) = \frac{2}{5}$
$P(\textsc{Agrupación} = \textsc{aislada}|\textsc{Comestible} = -) = \frac{2}{5}$
Producto $= \frac{5}{10}\frac{2}{5}\frac{3}{5}\frac{1}{5}\frac{2}{5}\frac{2}{5}$
Por tanto, según el clasificador Naive Bayes la seta no es comestible.

- **Para la instancia I_3:**
  $P(\textsc{Comestible} = +) = \frac{5}{10}$
  $P(\textsc{Color} = \textsc{marrón}|\textsc{Comestible} = +) = \frac{1}{5}$
  $P(\textsc{Tamaño} = \textsc{mediano}|\textsc{Comestible} = +) = \frac{1}{5}$
  $P(\textsc{Forma} = \textsc{plana}|\textsc{Comestible} = +) = \frac{3}{5}$
  $P(\textsc{Entorno} = \textsc{pinar}|\textsc{Comestible} = +) = \frac{3}{5}$
  $P(\textsc{Agrupación} = \textsc{racimo}|\textsc{Comestible} = +) = \frac{2}{5}$
  Producto $= \frac{5}{10}\frac{1}{5}\frac{1}{5}\frac{3}{5}\frac{3}{5}\frac{2}{5}$
  $P(\textsc{Comestible} = -) = \frac{5}{10}$
  $P(\textsc{Color} = \textsc{marrón}|\textsc{Comestible} = -) = \frac{2}{5}$
  $P(\textsc{Tamaño} = \textsc{mediano}|\textsc{Comestible} = -) = \frac{3}{5}$
  $P(\textsc{Forma} = \textsc{plana}|\textsc{Comestible} = -) = \frac{1}{5}$
  $P(\textsc{Entorno} = \textsc{pinar}|\textsc{Comestible} = -) = \frac{2}{5}$
  $P(\textsc{Agrupación} = \textsc{racimo}|\textsc{Comestible} = -) = \frac{1}{5}$
  Producto $= \frac{5}{10}\frac{2}{5}\frac{3}{5}\frac{1}{5}\frac{2}{5}\frac{1}{5}$
  Por tanto, según el clasificador Naive Bayes la seta es comestible.

- **Para la instancia I_4:**
  $P(\textsc{Comestible} = +) = \frac{5}{10}$
  $P(\textsc{Color} = \textsc{rojo}|\textsc{Comestible} = +) = \frac{2}{5}$
  $P(\textsc{Tamaño} = \textsc{mediano}|\textsc{Comestible} = +) = \frac{1}{5}$
  $P(\textsc{Forma} = \textsc{cóncava}|\textsc{Comestible} = +) = \frac{1}{5}$
  $P(\textsc{Entorno} = \textsc{pinar}|\textsc{Comestible} = +) = \frac{3}{5}$
  $P(\textsc{Agrupación} = \textsc{grupo}|\textsc{Comestible} = +) = \frac{2}{5}$
  Producto $= \frac{5}{10}\frac{2}{5}\frac{1}{5}\frac{1}{5}\frac{3}{5}\frac{2}{5}$
  $P(\textsc{Comestible} = -) = \frac{5}{10}$
  $P(\textsc{Color} = \textsc{rojo}|\textsc{Comestible} = -) = \frac{2}{5}$
  $P(\textsc{Tamaño} = \textsc{mediano}|\textsc{Comestible} = -) = \frac{3}{5}$
  $P(\textsc{Forma} = \textsc{cóncava}|\textsc{Comestible} = -) = \frac{1}{5}$
  $P(\textsc{Entorno} = \textsc{pinar}|\textsc{Comestible} = -) = \frac{2}{5}$
  $P(\textsc{Agrupación} = \textsc{grupo}|\textsc{Comestible} = -) = \frac{2}{5}$
  Producto $= \frac{5}{10}\frac{2}{5}\frac{3}{5}\frac{1}{5}\frac{2}{5}\frac{2}{5}$
  Por tanto, según el clasificador Naive Bayes la seta no es comestible.

Inteligencia Artificial II (ejercicio 3) (11–Junio–2007)

Apellidos: ....

Nombre: ....

Ejercicio 3 : 1.5 Consideremos el problema de caracterizar el perfil del comprador de un nuevo modelo de tienda de campaña en función de 5 atributos: `Altura`, nivel de `Estudios`, `Ingresos`, grado de `Ocupación` y `Utilidad`, que pueden tomar 3 valores cada uno: `bajo`, `medio` y `alto`.

Consideremos las hipótesis expresadas de la forma $\langle A, E, I, O, U\rangle$, donde A, E, I, O y U son, respectivamente, subconjuntos (pueden ser vacíos) del conjunto de valores que pueden tomar los atributos `Altura`, `Estudios`, `Ingresos`, `Ocupación` y `Utilidad`. Cada una de estas hipótesis representa un concepto en el que las instancias positivas son aquellas en las que los atributos toman, respectivamente, un valor de los incluidos en los conjuntos A, E, I, O y U. Así, una instancia positiva en el concepto representado por $\langle A, E, I, O, U\rangle$ es cualquier instancia $\langle a, e, i, o, u\rangle$ con $a \in A$, $e \in E$, $i \in I$, $o \in O$ y $u \in U$.

Se pide:

- Consideremos la hipótesis $h = \langle\{bajo, alto\}, \{bajo\}, \{medio\}, \{alto\}, \{bajo, medio\}\rangle$. Enumera todas las instancias positivas en el concepto representado por h. Da un ejemplo de hipótesis más específica que h, un ejemplo de hipótesis más general que h y un ejemplo de hipótesis que no sea comparable con h (ni más general, ni más específica).

Solución:

Instancias positivas en el concepto representado por h:

- $\langle bajo, bajo, medio, alto, bajo\rangle$
- $\langle bajo, bajo, medio, alto, medio\rangle$
- $\langle alto, bajo, medio, alto, bajo\rangle$
- $\langle alto, bajo, medio, alto, medio\rangle$

Hipótesis más específica que h: $\langle\{bajo\}, \{bajo\}, \{medio\}, \{alto\}, \{bajo\}\rangle$.
Hipótesis más general que h: $\langle\{bajo, alto\}, \{bajo, alto\}, \{medio\}, \{alto\}, \{bajo, medio\}\rangle$.
Hipótesis no comparable con h: $\langle\{medio\}, \{bajo\}, \{medio\}, \{alto\}, \{bajo, medio\}\rangle$.

- ¿Cuántos conceptos hay?, ¿y cuántas hipótesis? Si usamos este espacio de hipótesis en el algoritmo de eliminación de candidatos, ¿existe sesgo en el lenguaje?

Solución:

Número de conceptos $N_C = 2^{(3^5)} = 1413 \cdot 10^{70}$
Número de hipótesis $N_H = (2^3)^5 = 32768$
Sí existe sesgo en el lenguaje pues $N_H < N_C$

- Consideremos los siguientes ejemplos:

Ej	Altura	Estudios	Ingresos	Ocupación	Utilidad	Comprador
Ej_1	medio	medio	alto	medio	medio	no
Ej_2	alto	medio	bajo	alto	medio	si
Ej_3	alto	alto	alto	alto	bajo	no
Ej_4	medio	bajo	alto	bajo	alto	si

Calcula la cota general y la cota específica del espacio de versiones consistente con estos ejemplos. Para ello se ha de aplicar el **algoritmo de eliminación de candidatos** considerando las instancias en el orden en que se presentan en la tabla.

Solución:

- **Paso 0:**
  $S_0 = \{\langle\emptyset, \emptyset, \emptyset, \emptyset, \emptyset\rangle\}$
  $G_0 = \{\langle\{b,m,a\},\{b,m,a\},\{b,m,a\},\{b,m,a\},\{b,m,a\}\rangle\}$
- **Paso 1:**
  Ejemplo negativo $\langle m,m,a,m,m\rangle$
  Nada que eliminar de S_0**:** $S_1 = S_0$
  Especificación minimal de G_0
  $$G_1 = \{\langle\{b,a\},\{b,m,a\},\{b,m,a\},\{b,m,a\},\{b,m,a\}\rangle,$$
  $$\langle\{b,m,a\},\{b,a\},\{b,m,a\},\{b,m,a\},\{b,m,a\}\rangle,$$
  $$\langle\{b,m,a\},\{b,m,a\},\{b,m\},\{b,m,a\},\{b,m,a\}\rangle,$$
  $$\langle\{b,m,a\},\{b,m,a\},\{b,m,a\},\{b,a\},\{b,m,a\}\rangle,$$
  $$\langle\{b,m,a\},\{b,m,a\},\{b,m,a\},\{b,m,a\},\{b,a\}\rangle\}$$
- **Paso 2:**
  Ejemplo positivo $\langle a,m,b,a,m\rangle$
  Eliminamos de G_1 **las hipótesis 2ªy 5ª:**
  $$G_2 = \{\langle\{b,a\},\{b,m,a\},\{b,m,a\},\{b,m,a\},\{b,m,a\}\rangle,$$
  $$\langle\{b,m,a\},\{b,m,a\},\{b,m\},\{b,m,a\},\{b,m,a\}\rangle,$$
  $$\langle\{b,m,a\},\{b,m,a\},\{b,m,a\},\{b,a\},\{b,m,a\}\rangle\}$$
  Generalización minimal de S_1
  $S_2 = \{\langle\{a\},\{m\},\{b\},\{a\},\{m\}\rangle\}$
- **Paso 3:**
  Ejemplo negativo $\langle a,a,a,a,b\rangle$
  Nada que eliminar de S_2**:** $S_3 = S_2$
  Especificación minimal de G_2
  $$G_3 = \{\langle\{b,a\},\{b,m\},\{b,m,a\},\{b,m,a\},\{b,m,a\}\rangle,$$
  $$\langle\{b,a\},\{b,m,a\},\{b,m\},\{b,m,a\},\{b,m,a\}\rangle,$$
  $$\langle\{b,a\},\{b,m,a\},\{b,m,a\},\{b,m,a\},\{m,a\}\rangle,$$
  $$\langle\{b,m,a\},\{b,m,a\},\{b,m\},\{b,m,a\},\{b,m,a\}\rangle,$$
  $$\langle\{b,m,a\},\{b,m\},\{b,m,a\},\{b,a\},\{b,m,a\}\rangle,$$
  $$\langle\{b,m,a\},\{b,m,a\},\{b,m,a\},\{b,a\},\{m,a\}\rangle\}$$
- **Paso 4:**
  Ejemplo positivo $\langle m,b,a,b,a\rangle$
  Eliminamos de G_3 **las hipótesis 1ª, 2ª, 3ªy 5ª:**
  $$G_4 = \{\langle\{b,m,a\},\{b,m\},\{b,m,a\},\{b,a\},\{b,m,a\}\rangle,$$
  $$\langle\{b,m,a\},\{b,m,a\},\{b,m,a\},\{b,a\},\{m,a\}\rangle\}$$
  Generalización minimal de S_3
  $S_4 = \{\langle\{m,a\},\{b,m\},\{b,a\},\{b,a\},\{m,a\}\rangle\}$

Inteligencia Artificial II (ejercicio 4) **(11–Junio–2007)**

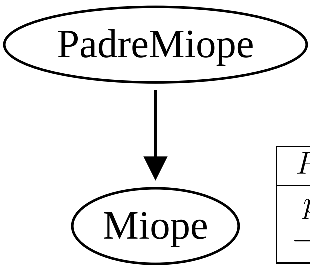
Apellidos: ....

Nombre: ....

Ejercicio 4 : 1.5 Se sabe que la miopía es un defecto visual que tiene cierta componente hereditaria, ya que tener un padre miope influye, de manera probabilística, en su aparición. Para realizar un modelo probabilístico de tal circunstancia, consideraremos dos variables aleatorias booleanas $Miope$ y $PadreMiope$ que representan, respectivamente, el tener miopía y el tener un progenitor con miopía. Se pide:

- Modelar la situación mediante una red bayesiana

Solución:



$P(pm)$
θ

PM	$P(m\|PM)$
pm	θ_1
$\neg pm$	θ_2

- Supongamos que realizamos un cuestionario a 1000 personas, obteniendo la siguiente información: 400 tenían un progenitor miope, de los cuales 300 eran a su vez miopes; otros 100 eran miopes sin tener progenitor miope. Con estos datos, obtener una estimación de máxima verosimilitud de los parámetros de la red bayesiana del apartado anterior ¿Qué significa exactamente que dichas estimaciones sean de máxima verosimilitud? Demostrar (con el desarrollo matemático correspondiente) que lo son.

Solución:

Una estimación de máxima verosimilitud es aquella que maximiza la probabilidad de los datos de los ejemplos de entrenamiento: $P(d|h)$**. De forma general, la estimación de máxima verosimilitud para un parámetro** $P(X = v|Y_1 = a_1, \ldots, Y_k = a_k)$ **se obtiene calculando la fracción de datos observados con valor** $X = v$ **de entre los que cumplen** $Y_1 = a_1, \ldots, Y_k = a_k$**. En este caso se tienen las siguientes estimaciones de** θ**,** θ_1 **y** θ_2**:**

$$\theta = \frac{400}{1000} = 0.4 \qquad \theta_1 = \frac{300}{400} = 0.75 \qquad \theta_2 = \frac{100}{600} = \frac{1}{6}$$

En este caso, el conjunto de hipótesis es $\{\langle\theta, \theta_1, \theta_2\rangle : \theta, \theta_1, \theta_2 \in [0, 1]\}$

La verosimilitud de los datos es:

$$P(d|h_{\theta,\theta_1,\theta_2}) = \theta^{400} \cdot (1-\theta)^{600} \cdot \theta_1^{300} \cdot (1-\theta_1)^{100} \cdot \theta_2^{100} \cdot (1-\theta_2)^{500}$$

La log-verosimilitud de los datos es:

$$L(d|h_{\theta,\theta_1,\theta_2}) = 400 \cdot log(\theta) + 600 \cdot log(1-\theta) + 300 \cdot log(\theta_1) +$$

$$+ 100 \cdot log(1-\theta_1) + 100 \cdot log(\theta_2) + 500 \cdot log(1-\theta_2)$$

Realizamos las derivadas parciales con respecto a θ**,** θ_1 **y** θ_2**, igualamos a 0 y calculamos el valor del parámetro:**

$$\frac{dL(d|h_{\theta,\theta_1,\theta_2})}{d\theta} = \frac{400}{\theta} - \frac{600}{1-\theta} = 0 \Rightarrow \frac{400}{\theta} = \frac{600}{1-\theta} \Rightarrow 400 \cdot (1-\theta) = 600 \cdot \theta \Rightarrow$$

$$\Rightarrow 400 - 400 \cdot \theta = 600 \cdot \theta \Rightarrow 400 = 1000 \cdot \theta \Rightarrow \theta = \frac{400}{1000} = 0.4$$

$$\frac{dL(d|h_{\theta,\theta_1,\theta_2})}{d\theta_1} = \frac{300}{\theta_1} - \frac{100}{1-\theta_1} = 0 \Rightarrow \frac{300}{\theta_1} = \frac{100}{1-\theta_1} \Rightarrow 300 \cdot (1-\theta_1) = 100 \cdot \theta_1 \Rightarrow$$

$$\Rightarrow 300 - 300 \cdot \theta_1 = 100 \cdot \theta_1 \Rightarrow 300 = 400 \cdot \theta_1 \Rightarrow \theta_1 = \frac{300}{400} = 0.75$$

$$\frac{dL(d|h_{\theta,\theta_1,\theta_2})}{d\theta_2} = \frac{100}{\theta_2} - \frac{500}{1-\theta_2} = 0 \Rightarrow \frac{100}{\theta_2} = \frac{500}{1-\theta_2} \Rightarrow 100 \cdot (1-\theta_2) = 500 \cdot \theta_2 \Rightarrow$$

$$\Rightarrow 100 - 100 \cdot \theta_2 = 500 \cdot \theta_2 \Rightarrow 100 = 600 \cdot \theta_2 \Rightarrow \theta_2 = \frac{100}{600} = \frac{1}{6}$$

- Supongamos ahora que en la encuesta realizada no hubiéramos preguntado por los progenitores, y simplemente tuviéramos la información de que tenemos 400 miopes y 600 no miopes. ¿Por qué necesitamos recurrir a un algoritmo EM para encontrar una estimación ML de los parámetros de la red? ¿Hacia qué converge el algoritmo EM? Aplicar **dos iteraciones** del algoritmo, tomando inicialmente todos los parámetros igual a 0.5.

Solución:

Tendremos que recurrir al algoritmo EM para estimar el valor de los parámetros asociados a las variables ocultas (en este caso $PadreMiope$). Este algoritmo converge hacia un óptimo local que maximiza la log-verosimilitud.

Aplicamos dos iteraciones del algoritmo EM para el caso del ejercicio, con los siguientes valores iniciales de los parámetros: $\theta^{(0)} = 0.5, \theta_1^{(0)} = 0.5, \theta_2^{(0)} = 0.5$**. En cada iteración:**

$$\theta^{(i+1)} = \frac{\hat{N}(pm)}{1000} \qquad \theta_1^{(i+1)} = \frac{\hat{N}(m,pm)}{\hat{N}(pm)} \qquad \theta_2^{(i+1)} = \frac{\hat{N}(m,\neg pm)}{\hat{N}(\neg pm)}$$

$$\hat{N}(pm) = \sum_{M}^{1000} P(pm|M) = 400 \cdot P(pm|m) + 600 \cdot P(pm|\neg m)$$

$$\hat{N}(\neg pm) = \sum_{M}^{1000} P(\neg pm|M) = 400 \cdot P(\neg pm|m) + 600 \cdot P(\neg pm|\neg m)$$

$$\hat{N}(m,pm) = \sum_{m}^{400} P(pm|m) = 400 \cdot P(pm|m)$$

$$\hat{N}(m,\neg pm) = \sum_{m}^{400} P(\neg pm|m) = 400 \cdot P(\neg pm|m)$$

$$P(pm|m) = \frac{P(m|pm) \cdot P(pm)}{\sum_{PM} P(m|PM) \cdot P(PM)} = \frac{\theta_1^{(i)} \cdot \theta^{(i)}}{\theta_1^{(i)} \cdot \theta^{(i)} + \theta_2^{(i)} \cdot (1-\theta^{(i)})}$$

$$P(pm|\neg m) = \frac{P(\neg m|pm) \cdot P(pm)}{\sum_{PM} P(\neg m|PM) \cdot P(PM)} = \frac{(1-\theta_1^{(i)}) \cdot \theta^{(i)}}{(1-\theta_1^{(i)}) \cdot \theta^{(i)} + (1-\theta_2^{(i)}) \cdot (1-\theta^{(i)})}$$

$$P(\neg pm|m) = 1 - P(pm|m) \qquad P(\neg pm|\neg m) = 1 - P(pm|\neg m)$$

En la primera iteración:

$$P(pm|m) = \frac{\theta_1^{(0)} \cdot \theta^{(0)}}{\theta_1^{(0)} \cdot \theta^{(0)} + \theta_2^{(0)} \cdot (1-\theta^{(0)})} = \frac{0.5 \cdot 0.5}{0.5 \cdot 0.5 + 0.5 \cdot (1-0.5)} = 0.5$$

$$P(pm|\neg m) = \frac{(1-\theta_1^{(0)}) \cdot \theta^{(0)}}{(1-\theta_1^{(0)}) \cdot \theta^{(0)} + (1-\theta_2^{(0)}) \cdot (1-\theta^{(0)})} = \frac{(1-0.5) \cdot 0.5}{(1-0.5) \cdot 0.5 + (1-0.5) \cdot (1-0.5)} = 0.5$$

$$P(\neg pm|m) = 1 - P(pm|m) = 1 - 0.5 = 0.5 \qquad P(\neg pm|\neg m) = 1 - P(pm|\neg m) = 1 - 0.5 = 0.5$$

$$\hat{N}(pm) = 400 \cdot P(pm|m) + 600 \cdot P(pm|\neg m) = 400 \cdot 0.5 + 600 \cdot 0.5 = 500$$

$$\hat{N}(\neg pm) = 400 \cdot P(\neg pm|m) + 600 \cdot P(\neg pm|\neg m) = 400 \cdot 0.5 + 600 \cdot 0.5 = 500$$

$$\hat{N}(m, pm) = 400 \cdot P(pm|m) = 400 \cdot 0.5 = 200$$

$$\hat{N}(m, \neg pm) = 400 \cdot P(\neg pm|m) = 400 \cdot 0.5 = 200$$

$$\theta^{(1)} = \frac{\hat{N}(pm)}{1000} = \frac{500}{1000} = 0.5 \qquad \theta_1^{(1)} = \frac{\hat{N}(m, pm)}{\hat{N}(pm)} = \frac{200}{500} = 0.4 \qquad \theta_2^{(1)} = \frac{\hat{N}(m, \neg pm)}{\hat{N}(\neg pm)} = \frac{200}{500} = 0.4$$

En la segunda iteración:

$$P(pm|m) = \frac{\theta_1^{(1)} \cdot \theta^{(1)}}{\theta_1^{(1)} \cdot \theta^{(1)} + \theta_2^{(1)} \cdot (1 - \theta^{(1)})} = \frac{0.4 \cdot 0.5}{0.4 \cdot 0.5 + 0.4 \cdot (1 - 0.5)} = 0.5$$

$$P(pm|\neg m) = \frac{(1 - \theta_1^{(1)}) \cdot \theta^{(1)}}{(1 - \theta_1^{(1)}) \cdot \theta^{(1)} + (1 - \theta_2^{(1)}) \cdot (1 - \theta^{(1)})} = \frac{(1 - 0.4) \cdot 0.5}{(1 - 0.4) \cdot 0.5 + (1 - 0.4) \cdot (1 - 0.5)} = 0.5$$

$$P(\neg pm|m) = 1 - P(pm|m) = 1 - 0.5 = 0.5 \qquad P(\neg pm|\neg m) = 1 - P(pm|\neg m) = 1 - 0.5 = 0.5$$

$$\hat{N}(pm) = 400 \cdot P(pm|m) + 600 \cdot P(pm|\neg m) = 400 \cdot 0.5 + 600 \cdot 0.5 = 500$$

$$\hat{N}(\neg pm) = 400 \cdot P(\neg pm|m) + 600 \cdot P(\neg pm|\neg m) = 400 \cdot 0.5 + 600 \cdot 0.5 = 500$$

$$\hat{N}(m, pm) = 400 \cdot P(pm|m) = 400 \cdot 0.5 = 200$$

$$\hat{N}(m, \neg pm) = 400 \cdot P(\neg pm|m) = 400 \cdot 0.5 = 200$$

$$\theta^{(2)} = \frac{\hat{N}(pm)}{1000} = \frac{500}{1000} = 0.5 \qquad \theta_1^{(2)} = \frac{\hat{N}(m, pm)}{\hat{N}(pm)} = \frac{200}{500} = 0.4 \qquad \theta_2^{(2)} = \frac{\hat{N}(m, \neg pm)}{\hat{N}(\neg pm)} = \frac{200}{500} = 0.4$$
