15F1

PREGUNTA 2. Valoración: 3

Aplique el algoritmo de eliminación de candidatos (**Espacio de Versiones**) a los datos reflejados en la tabla adjunta, suponiendo que éstos son ejemplos positivos (+) y negativos (-) del concepto buscado ("Cuándo comprar un paquete de viajes") y que se presentan uno a uno en el orden indicado en la propia tabla. Especificar en cada paso los valores de los conjuntos S y G. Justifique si se llega o no a aprender el concepto buscado.

	Destino	Transporte	Transbordo	Alojamiento	Precio	Clase
1	Europa	Tren	Sí	Pensión	Barato	Comprar(+)
2	Europa	Avión	No	Hotel	Caro	Descartar(-)
3	Europa	Avión	No	Pensión	Barato	Comprar(+)
4	Asia	Tren	No	Hotel	Barato	Descartar(-)
5	Europa	Tren	Sí	Albergue	Barato	Comprar(+)

G_0 = { (Destino, Transporte, Transbordo, Alojamiento, Precio)}
S_0 = { Ø }

---

e_1 = (europa, tren, sí, pensión, barato, +)

G_1 =G_0 = { (Destino, Transporte, Transbordo, Alojamiento, Precio) }, no se elimina nada, no hay inconsistencia entre G_0 y e_1 (la hipótesis G_0 describe el ejemplo positivo)
S_1 = { (**europa**, **tren**, **sí**, **pensión**, **barato**) } generalización del conjunto vacío con los atributos del ejemplo.

---

e_2 = (europa, avión, no, hotel, caro, –)

S_2 = S_1 = { (europa, tren, sí, pensión, barato) }, no se elimina nada, no hay inconsistencia S_1 y e_2 (la hipótesis S_1 no describe el ejemplo negativo)
G_2 = { ~~(Destino, Transporte, Transbordo, Alojamiento, Precio)~~, hay inconsistencia en esta hipótesis de G_1 respecto a e_2 (describe el ejemplo negativo) por ello se elimina y se añaden las hipótesis mínimamente especializadas de la eliminada:

~~(asia, Transporte, Transbordo, Alojamiento, Precio)~~, inconsistencia con S_2 , no la describe.
(Destino, tren, Transbordo, Alojamiento, Precio),
(Destino, Transporte, sí, Alojamiento, Precio),
(Destino, Transporte, Transbordo, pensión, Precio),
~~(Destino, Transporte, Transbordo, albergue, Precio)~~, inconsistencia con S_2 , no la describe.
(Destino, Transporte, Transbordo, Alojamiento, barato)}

---

e_3 = (europa, avión, no, pensión, barato, +)

G_3 = { ~~(Destino, tren, Transbordo, Alojamiento, Precio)~~, inconsistencia con e_3 , no lo describe.
~~(Destino, Transporte, sí, Alojamiento, Precio)~~, inconsistencia con e_3 , no lo describe.
(Destino, Transporte, Transbordo, pensión, Precio),
(Destino, Transporte, Transbordo, Alojamiento, barato)}
S_3 = { (europa, **Transporte**, **Transbordo**, pensión, barato) }

---

e_4 = (asia, tren, no, hotel, barato, –)

$S_4 = S_3$ = { (europa, Transporte, Transbordo, pensión, barato) }
G_4 = { (Destino, Transporte, Transbordo, pensión, Precio),
~~(Destino, Transporte, Transbordo, Alojamiento, barato)~~, hay inconsistencia en esta hipótesis de G_3 respecto a e_4 (describe el ejemplo negativo) por ello se elimina y se añaden las hipótesis mínimamente especializadas de la hipótesis eliminada:

(europa, Transporte, Transbordo, Alojamiento, barato),
~~(Destino, avión, Transbordo, Alojamiento, barato)~~, inconsistencia con S_4 , no la describe.
~~(Destino, Transporte, sí, Alojamiento, barato)~~, inconsistencia con S_4 , no la describe.
~~(Destino, Transporte, Transbordo, pensión, barato)~~, existe hipótesis más general.
~~(Destino, Transporte, Transbordo, albergue, barato)~~ } inconsistencia con S_4 , no la describe.

---

e_5 = (europa, tren, sí, albergue, barato, +)

G_5 = { ~~(Destino, Transporte, Transbordo, pensión, Precio)~~, inconsistencia con e_5 , no lo describe.
(europa, Transporte, Transbordo, Alojamiento, barato)}
S_5 = { (europa, Transporte, Transbordo, **Alojamiento**, barato) }

15F2

PREGUNTA 3. Valoración: 3
Aplique el algoritmo de eliminación de candidatos (**Espacio de Versiones**) a los datos reflejados en la tabla adjunta, suponiendo que éstos son ejemplos positivos (+) y negativos (-) del concepto buscado ("Cuándo viajar") y que se presentan uno a uno en el orden indicado en la propia tabla. Especificar en cada paso los valores de los conjuntos S y G. Justifique si se llega o no a aprender el concepto buscado.

	Destino	**Transporte**	**Transbordo**	**Alojamiento**	**Precio**	**Clase**
1	Europa	Tren	Sí	Albergue	Barato	Viajar (+)
2	Asia	Tren	No	Hotel	Barato	No Viajar (-)
3	Europa	Avión	No	Pensión	Barato	Viajar (+)
4	Europa	Avión	No	Hotel	Caro	No Viajar (-)
5	Europa	Tren	Sí	Pensión	Barato	Viajar (+)

G_0 = { (Destino, Transporte, Transbordo, Alojamiento, Precio)}
S_0 = { Ø }

---

e_1 = (europa, tren, sí, albergue, barato, +)

$G_1 = G_0$ = { (Destino, Transporte, Transbordo, Alojamiento, Precio) }
S_1 = {(europa, tren, sí, albergue, barato) }

---

e_2 = (asia, tren, no, hotel, barato, –)

$S_2 = S_1$ = { (europa, tren, sí, albergue, barato) }
G_2 = { ~~(Destino, Transporte, Transbordo, Alojamiento, Precio)~~, hay inconsistencia en esta hipótesis de G_1 respecto a e_2 (describe el ejemplo negativo) por ello se elimina y se añaden las hipótesis míimamente especializadas de la eliminada:

(europa, Transporte, Transbordo, Alojamiento, Precio),
~~(Destino, avión, Transbordo, Alojamiento, Precio)~~, inconsistencia con S_2 , no la describe.
(Destino, Transporte, sí, Alojamiento, Precio),
(Destino, Transporte, Transbordo, albergue, Precio),
~~(Destino, Transporte, Transbordo, pensión, Precio)~~, inconsistencia con S_2 , no la describe.
~~(Destino, Transporte, Transbordo, Alojamiento, caro)~~} inconsistencia con S_2 , no la describe.

---

e_3 = (europa, avión, no, pensión, barato, +)

G_3 = { (europa, Transporte, Transbordo, Alojamiento, Precio),
~~(Destino, Transporte, sí, Alojamiento, Precio)~~, inconsistencia con e_3 , no lo describe.
~~(Destino, Transporte, Transbordo, albergue, Precio)~~ } inconsistencia con e_3 , no lo describe.
S_3 = { ~~(europa, Transporte, Transbordo, albergue, barato)~~, inconsistencia con e_3, no lo describe, se elimina...
(europa, Transporte, Transbordo, **Alojamiento**, barato) } ...y se generaliza para que describa a e_3.

---

e_4 = (europa, avión, no, hotel, caro, –)

$S_4 = S_3$ = {(europa, Transporte, Transbordo, Alojamiento, barato) } se mantiene, no describe al e_3 negativo
G_4 = { ~~(europa, Transporte, Transbordo, Alojamiento, Precio)~~, hay inconsistencia en esta hipótesis de G_3 respecto a e_4 (describe el ejemplo negativo) por ello se elimina y se añaden las hipótesis mínimamente especializadas de la eliminada:

~~(europa, tren, Transbordo, Alojamiento, Precio)~~, inconsistencia con S_4, no la describe.
~~(europa, Transporte, sí, Alojamiento, Precio)~~, inconsistencia con S_4, no la describe.
~~(europa, Transporte, Transbordo, pensión, Precio)~~, inconsistencia con S_4, no la describe.
~~(europa, Transporte, Transbordo, albergue, Precio)~~, inconsistencia con S_4, no la describe.
(europa, Transporte, Transbordo, Alojamiento, barato) }

---

e_5 = (europa, tren, sí, pensión, barato, +)

$G_5 = G_4$ = { (europa, Transporte, Transbordo, Alojamiento, barato) }
$S_5 = S_4$ = { (europa, Transporte, Transbordo, Alojamiento, barato) }
G_5 y S_5 son iguales, se ha aprendido el concepto (todos los paquetes de viaje a Europa se compran).

15SO

PREGUNTA 3. Valoración 2.5

La tabla 1 adjunta muestra ejemplos positivos y negativos del concepto: ¿Cuándo jugar a tenis? Se pide:

1. Aplicar el algoritmo de **eliminación de candidatos (espacio de versiones)** a los 3 primeros ejemplos de dicha tabla y en el orden indicado ¿Se llega a aprender el concepto? Desde el punto de vista del espacio de versiones ¿qué se puede decir acerca del concepto aprendido? Es muy importante que, al procesar cada ejemplo, especifique y justifique los valores de los conjuntos S y G.

2. Supóngase que, continuando con lo realizado en el apartado anterior, se procesa ahora los ejemplos número 4 y 5. Obtenga los valores de los conjuntos S y G ¿Se aprendió ahora el concepto?

3. Si a continuación se procesa el ejemplo 6, ¿cuál es el valor de los conjuntos G y S? Justifique cuál es la causa del resultado obtenido al procesar este ejemplo.

	Cielo	Temperatura	Humedad	Viento	¿Jugar a tenis?
1	soleado	alta	baja	si	+
2	soleado	alta	baja	no	+
3	soleado	baja	baja	no	-
4	lluvioso	alta	alta	si	+
5	lluvioso	baja	alta	si	-
6	soleado	alta	Alta	no	-

Tabla 1

G_0 = { (Cielo, Temperatura, Humedad, Viento)}
S_0 = { Ø }

---

e_1 = (soleado, alta, baja, sí, +)

G_1 = { (Cielo, Temperatura, Humedad, Viento) }, no se elimina nada, no hay inconsistencia G_0 y e_1
S_1 = { (**soleado**, **alta**, **baja**, **sí**) }

---

e_2 = (soleado, alta, baja, no, +)

$G_2 = G_1$ ={ (Cielo, Temperatura, Humedad, Viento)}, no se elimina nada, no hay inconsistencia G_1 y e_2
S_2 = { ~~(soleado, alta, baja, sí)~~, eliminación de la única hipótesis de S_1 porque hay inconsistencia entre S_1 y e_2
(soleado, alta, baja, **Viento**) } y se añade generalización mínimamente generalizadas de la hipótesis eliminada para que sea consistente e_2.

---

e_3 = (soleado, baja, baja, no, –)

$S_3 = S_2$ = { (soleado, alta, baja, **Viento**) }, no se elimina nada, no hay inconsistencia en S_2 respecto a e_3
G_3 = { ~~(Cielo, Temperatura, Humedad, Viento)~~, hay inconsistencia en esta hipótesis de G_2 respecto a e_3 (describe el ejemplo negativo) por ello se elimina y se añaden las hipótesis mínimamente especializadas de la eliminada:
~~(lluvioso, Temperatura, Humedad, Viento)~~, inconsistencia con S_3 , no describe la hipótesis

(Cielo, alta, Humedad, Viento),
~~(Cielo, Temperatura, alta, Viento)~~, inconsistencia con S_3 , no describe la hipótesis
~~(Cielo, Temperatura, Humedad, sí)~~ } inconsistencia con S_3 , no describe la hipótesis

G_3 y S_3 no coinciden, por lo que no se ha aprendido el concepto de cuándo se juega al tenis.

---

e_4 = (lluvioso, alta, alta, sí, +)

$G_4 = G_3$ = { (Cielo, alta, Humedad, Viento), no se elimina nada, no hay inconsistencia en G_3 respecto a e_4

S_4 = { ~~(soleado, alta, baja, Viento)~~, hay inconsistencia en S_3 respecto a e_4 (no describe el ejemplo positivo) por ello se añade las hipótesis máximamente específica de la hipótesis eliminada incluyendo a e_4
(**Cielo**, alta, Humedad, Viento) }, se generaliza el atributo no común entre S_3 y e_4.

G_4 y S_4 coinciden, por lo que se ha aprendido el concepto de que con temperatura alta siempre se juega al tenis.

---

e_5 = (lluvioso, baja, alta, sí, –)

$S_5 = S_4$ = { (Cielo, alta, Humedad, Viento) }, no se elimina nada, no hay inconsistencia entre S_4 y e_5
$G_5 = G_4$ = { (Cielo, alta, Humedad, Viento) } no se elimina nada, no hay inconsistencia entre S_4 y e_5

G_5 y S_5 siguen coincidiendo.

---

e_6 = (soleado, alta, alta, no, –)

S_6 = { ~~(Cielo, alta, Humedad, Viento)~~ } = { Ø }, inconsistencia entre S_5 y e_6
G_6 = { ~~(Cielo, alta, Humedad, Viento)~~, hay inconsistencia en esta hipótesis de G_5 respecto a e_6 (describe el ejemplo negativo) por ello se elimina y se añaden las hipótesis mínimamente especializadas de la eliminada:
(lluvioso, alta, Humedad, Viento),
(Cielo, alta, baja, Viento),
(Cielo, alta , Humedad, sí) }

G_6 y S_6 no coinciden, habiéndose aprendido antes el concepto, lo que indica que hay ruido en los datos de la muestra.

15SR

PREGUNTA 1. Valoración: 2.5

Describir la estrategia del **Espacio de Versiones** indicando:

a) Bias.
b) Cómo le afecta el ruido: en entradas y en estructura.
c) Su complejidad espacial y temporal.
d) Control de la tarea aprendida (crítica/valoración/utilización) y dependencia del conocimiento del dominio.

a) Bias.

No está sujeto a restricciones o bias de preferencia en su espacio de búsqueda, salvo la aportada por el propio lenguaje de descripción del concepto, que debe ser lo más restringido posible pero abarcando el concepto.

b) Cómo le afecta el ruido: en entradas y en estructura.

En entradas no se trata con ejemplos con ruido, y en estructura, la salida del EV no es tolerante a fallos.

c) Su complejidad espacial y temporal.

El espacio de almacenamiento necesario es del orden de O(s+g), donde **s** es el tamaño máximo del conjunto **S** y **g** es el tamaño máximo alcanzado por **G**.
El tiempo crece polinómicamente con el número de instancias de entrenamientos (**p** + **n**, instacias positivas más negativas), exactamente es de $O(sg(p+n) + s^2p + g^2n)$.
En defeinitiva, la eficiciencia dependerá de los tamaños de **G** y **S**.

d) Control de la tarea aprendida (crítica/valoración/utilización) y dependencia del conocimiento del dominio.

El Espacio de Versiones no utiliza ningún procedimiento para el control de la tarea aprendida, pero permite identificar los ejemplos clasificados perfectamente (ya sean positivos o negativos) como los ejemplos clasificados parcialmente.
La dependencia del conocimiento del dominio se refleja en la definición de los atributos y en sus posibles valores. También el proceso dependerá del lenguaje de descripción de hipótesis.

16F1 – Preguntas 2.b – 2.e

PREGUNTA 2. Valoración: 5

El **algoritmo** ***ID3*** y el **algoritmo** ***Eliminación de Candidatos*** son dos claros ejemplos de métodos de aprendizaje inductivos. En este contexto, conteste a las siguientes cuestiones:

(a) Obtenga el árbol de decisión que se puede aprender mediante el algoritmo *ID3* a partir de los ejemplos dados en la tabla 1.

Ejemplo	Hiper-tensión	Taqui-cardia	Gota	Hipotiroidismo	Diabetes	Urea en sangre	Administrar Tratamiento T0
1	Sí	No	Sí	No	No	Sí	SI
2	Sí	No	No	Sí	No	Sí	SI
3	Sí	No	No	Sí	Sí	No	SI
4	No	No	No	Sí	No	No	NO

Tabla 1. Ejemplos positivos y negativos del concepto objetivo: "Administrar tratamiento T0"

(b) Aplique el algoritmo de Eliminación de Candidatos a los ejemplos de la tabla 1. ¿Se pudo encontrar el concepto aprendido? Si la respuesta es negativa, indique el conjunto de hipótesis que forman el espacio de versiones. Justifique sus respuestas.

(c) Cuál es la relación entre el *EV* y el árbol de decisión aprendidos en los dos apartados anteriores. Ambos obtienen hipótesis consistentes con los ejemplos de entrenamiento pero ¿cuál es la diferencia?

(d) Describa las *bias* que utiliza cada algoritmo: Eliminación de candidatos e ID3.

(e) Añadir los dos siguientes ejemplos y calcular el nuevo árbol de decisión y el nuevo espacio de versiones. En cada caso, justifique los distintos pasos.

5	Sí	Sí	Sí	No	No	Sí	NO
6	Sí	Sí	No	Sí	No	Sí	NO

2.b)

G_0 = { (HT, T, G, H, D, US)}
S_0 = { Ø }

---

e_1 = (sí, no, sí, no, no, sí, +)

G_1 = G_0 = { (HT, T, G, H, D, US)}
S_0 = {**(sí**, **no**, **sí**, **no**, **no**, **sí)**} añadir generalizaciones minimales consistentes con e_1, con alguna hipótesis de G_{31} más general (o igual) que S_1.

---

e_2 = (sí, no, no, sí, no, sí, +)

G_2 =G_1 = { (HT, T, G, H, D, US)}
S_1 = {~~(sí, no, sí, no, no, sí)~~,
(sí, no, **G**, **H**, no, sí)}

---

e_3 = (sí, no, no, sí, si, no, +)

G_3 =G_2 = { (HT, T, G, H, D, US)}
S_3 = {~~(sí, no, G, H, no, sí)~~,
(sí, no, G, H, **D**, **US**)}

---

e_4 = (no, no, no, sí, no, no, **–)**

S_4 = S_3 = {(sí, no, G, H, D, US)}
G_4 = { ~~(HT, T, G, H, D, US)~~, hay inconsistencia en esta hipótesis de G_3 respecto a e_4 (describe el ejemplo negativo) por ello se elimina y se añaden las hipótesis mínimamente especializadas de la eliminada:
(sí, T, G, H, D, US),
~~(HT, sí, G, H, D, US)~~, inconsistencia con S_3 , no la describe.
~~(HT, T, sí, H, D, US)~~, inconsistencia con S_3 , no la describe.
~~(HT, T, G, no, D, US)~~, inconsistencia con S_3 , no la describe.
~~(HT, T, G, H, sí, US)~~, inconsistencia con S_3 , no la describe.
~~(HT, T, G, H, D, sí)~~ } inconsistencia con S_3 , no la describe.

G_4 y S_4 no coinciden, por lo que no se ha aprendido un concepto respecto a cuándo administrar el tratamiento T0.

El conjunto de hipótesis que forman el EV sería el siguiente:

G_4 = { (sí, T, G, H, D, US) }
↓
S_4 = { (sí, no, G, H, D, US) }

2.c)

La principal diferencia entre ambos resultados, es que los ejemplos son clasificados inequívocamente en base al atributo *Hipertensión*, no dejando espacio a la duda. Sin embargo, con el resultado de EV, la *Hipertensión* no clasifica perfectamente a los ejemplos. Un ejemplo con *Hipertensión* = Sí y *Taquicardia* = Sí según el conjunto S no sería candidato a administrar el tratamiento aunque sí está dentro de las hipótesis de G, por lo tanto, no sabríamos que hacer.
Podemos añadir que con ID3, con el resultado obtenido, no somos conscientes de si el árbol obtenido está completo y clasifica perfectamente a los ejemplos, sin embargo, con el EV sí sabemos que necesitamos más ejemplos para obtener un modelo definitivo.

2.d)

Espacio de Vectores: no está sujeto a restricciones o bias de preferencia en su espacio de búsqueda, salvo la aportada por el propio lenguaje de descripción del concepto, que debe ser lo más restringido posible pero abarcando el concepto.

ID3 (árboles de decisión):
La bias viene marcada por la entropía/ganancia de información, en la que se prefieren árboles que tengan atributos con más información más cerca de la raíz y que sean más cortos. También favorece por su heurística, atributos con muchos valores.

2.e)

e_5 = (si, si, si, no, no, si, –)

$S_5 = S_4$ = { (sí, no, G, H, D, US) }
G_4 = { ~~(sí, T, G, H, D, US)~~, hay inconsistencia en esta hipótesis de G_4 respecto a e_5 (describe el ejemplo negativo) por ello se elimina y se añaden las hipótesis mínimamente especializadas de la eliminada:

(sí,no, G, H, D, US),
~~(si, T, no, H, D, US)~~, inconsistencia con S_5 , no la describe.
~~(si, T, G, si, D, US)~~, inconsistencia con S_5 , no la describe.
~~(si, T, G, H, si, US)~~, inconsistencia con S_5 , no la describe.
~~(si, T, G, H, D, no)~~ } inconsistencia con S_5 , no la describe.

G_5 y S_5 no coinciden, por lo que se ha aprendido el concepto: "Cuando se tiene hipertensión y se tiene taquicardia, administrar el tratamiento T0".

---

e_6 = (si, si, no, si, no, si, –)

$S_6 = S_5$ = { (sí, no, G, H, D, US) }
$G_6 = G_5$ = { (sí, no, G, H, D, US) }

16SO

PREGUNTA 2. Valoración: 3

Aplique el algoritmo del **Espacio de Versiones** a los datos reflejados en la tabla adjunta, suponiendo que éstos son ejemplos positivos y negativos del concepto buscado ("Cuándo comprar un paquete de viajes"), y que se presentan, uno a uno, en el orden indicado en la propia tabla. Especificar, en cada paso, los valores de los conjuntos S y G, justificando las decisiones tomadas. Indique el espacio de versiones resultante y si se aprendió o no el concepto.

	Destino	Transporte	Transbordo	Alojamiento	Precio	Clase
1	Europa	Tren	Sí	Pensión	Barato	Comprar(+)
2	Europa	Avión	No	Pensión	Barato	Comprar(+)
3	Europa	Tren	Sí	Albergue	Barato	Comprar(+)
4	Europa	Avión	No	Hotel	Caro	Descartar(-)
5	Asia	Tren	No	Hotel	Barato	Descartar(-)

Si la secuencia de patrones fuese idéntica a la del caso anterior, pero no procesara el ejemplo número 5, indique cuál sería el espacio de versiones resultante. En este caso, clasifique, positiva o negativamente, los siguientes ejemplos:

- (Europa, Avión, Sí, Pensión, Caro)
- (Europa, Avión, Sí, Albergue, Barato)
- (Europa, Avión, Sí, Albergue, Barato).

G_0 = { (Destino, Transporte, Transbordo, Alojamiento, Precio)}
S_0 = { Ø }

---

e_1 = (europa, tren, sí, pensión, barato, +)

$G_1 = G_0$ = { (Destino, Transporte, Transbordo, Alojamiento, Precio) }
S_1 = { (**europa**, **tren**, **sí**, **pensión**, **barato**) }

---

e_2 = (europa, avión, no, pensión, barato, +)

$G_2 = G_1$ = { (Destino, Transporte, Transbordo, Alojamiento, Precio) }
S_1 = {~~(europa, tren, sí, pensión, barato)~~, (inconsistente con e_2, no describe ejemplo +)
(europa, **Transporte**, **Transbordo**, pensión, barato) }

---

e_3 = (europa, tren, sí, albergue, barato, +)

$G_3 = G_2$ = { (Destino, Transporte, Transbordo, Alojamiento, Precio)}
S_3 = {~~(europa, Transporte, Transbordo, pensión, barato)~~, (inconsistente con e_3, no describe ejemplo +)
(europa, Transporte, Transbordo, **Alojamiento**, barato) }

---

e_4 = (europa, avión, no, hotel, caro, –)

$S_4 = S_3$ = { (europa, Transporte, Transbordo, Alojamiento, barato) }
G_4 = { ~~(Destino, Transporte, Transbordo, Alojamiento, Precio)~~, hay inconsistencia en esta hipótesis de G_3 respecto a e_4 (describe el ejemplo negativo) por ello se elimina y se añaden las hipótesis mínimamente especializadas de la eliminada:

~~(asia, Transporte, Transbordo, Alojamiento, Precio)~~, inconsistente con S_4, no la describe
~~(Destino, tren, Transbordo, Alojamiento, Precio)~~, inconsistente con S_4, no la describe
~~(Destino, Transporte, sí, Alojamiento, Precio)~~, inconsistente con S_4, no la describe
~~(Destino, Transporte, Transbordo, albergue, Precio)~~, inconsistente con S_4, no la describe
~~(Destino, Transporte, Transbordo, pensión, Precio)~~, inconsistente con S_4, no la describe
(Destino, Transporte, Transbordo, Alojamiento, barato) }

---

e_5 = (asia, tren, no, hotel, barato, –)

$S_5 = S_4$ = { (europa, Transporte, Transbordo, Alojamiento, barato) }
G_5 = { ~~(Destino, Transporte, Transbordo, Alojamiento, barato)~~, hay inconsistencia en esta hipótesis de G_4 respecto a e_5 (describe el ejemplo negativo) por ello se elimina y se añaden sus hipótesis mínimamente especializadas:

(europa, Transporte, Transbordo, Alojamiento, barato),
~~(Destino, avión, Transbordo, Alojamiento, barato)~~, inconsistencia con S_5 , no la describe.
~~(Destino, Transporte, sí, Alojamiento, barato)~~, inconsistencia con S_5 , no la describe.
~~(Destino, Transporte, Transbordo, pensión, barato)~~, (existe hipótesis más general)
~~(Destino, Transporte, Transbordo, albergue, barato)~~ (existe hipótesis más general) }

G_5 = { (Destino, Transporte, Transbordo, Alojamiento, barato) }
↓
S_5 = { (europa, Transporte, Transbordo, Alojamiento, barato) }

(europa, avión, sí, pensión, caro) : verdadero
(europa, avión, sí, albergue, barato) : verdadero

17F1

PREGUNTA 2. Valoración: 3
DADOS:
(a) Concepto a aprender: Condiciones ideales para "volar en globo".
(b) Ejemplos: ver Tabla 1

	Cielo	Temperatura	Humedad	Viento	Volar en globo
1	soleado	alta	baja	Sí	+
2	soleado	alta	baja	No	+
3	soleado	baja	baja	No	-
4	lluvioso	alta	alta	Sí	+
5	lluvioso	baja	alta	Sí	-
6	soleado	alta	alta	No	+

Tabla 1

SE PIDE:

1. Aplicar el **algoritmo de eliminación de candidatos** (espacio de versiones) a los 3 primeros ejemplos de dicha tabla y en el orden indicado ¿Se llega a aprender el concepto? Indique el conjunto de hipótesis que componen el espacio de versiones obtenido hasta el momento.

2. Teniendo en cuenta el espacio de versiones obtenido en el paso anterior, establezca sí es posible o no clasificar los siguientes ejemplos de clasificación desconocida. En caso afirmativo, asigne la clasificación correspondiente.

   - (soleado, baja, baja, sí)
   - (soleado, alta, baja, no)
   - (soleado, alta, alta, no)

3. Continúe aplicando el algoritmo a los tres ejemplos restantes (y en el orden indicado) y muestre el espacio de versiones resultante.

   Nota: Al procesar cada ejemplo deberá mostrar y justificar los conjuntos G y S obtenidos.

G_0 = { (Cielo, Temperatura, Humedad, Viento) }
S_0 = { Ø }

---

e_1 = (soleado, alta, baja, sí, +)

$G_1 = G_0$ = { (Cielo, Temperatura, Humedad, Viento) }
S_1 = { (soleado, alta, baja, sí) }

---

e_2 = (soleado, alta, baja, no, +)

$G_2 = G_1$ = { (Cielo, Temperatura, Humedad, Viento) }
S_2 = { (soleado, alta, baja, **Viento**) }

---

e_3 = (soleado, baja, baja, no, –)

$S_3 = S_2$ = { (soleado, alta, baja, **Viento**) }
G_3 = { ~~(Cielo, Temperatura, Humedad, Viento)~~, hay inconsistencia en esta hipótesis de G_2 respecto a e_3 (describe el ejemplo negativo) por ello se elimina y se añaden las hipótesis mínimamente especializadas de la eliminada:

~~(lluvioso, Temperatura, Humedad, Viento)~~, inconsistencia con S_3 , no la describe.
(Cielo, alta, Humedad, Viento),
~~(Cielo, Temperatura, alta, Viento)~~, inconsistencia con S_3 , no la describe.
~~(Cielo, Temperatura, Humedad, sí)~~, inconsistencia con S_3 , no la describe. }

1) S_3 y G_3 son distintos, por lo que no se llega a aprender el concepto. El conjunto de hipótesis obtenido hasta el momento es:

S_3 = { (soleado, alta, baja, Viento) }
G_3 = { (Cielo, alta, Humedad, Viento) }

EV = { S_3 ,G_3 } = { (soleado, alta, baja, Viento), (Cielo, alta, Humedad, Viento) }

2)

Si S y G fueran iguales, cualquier ejemplo que fuera equiparable con (cubierto por) S o G, sería positivo, y el que no, sería negativo (éste no sería nuestro caso).

Si S y G son diferentes, la certeza que podemos tener para clasificar ejemplos es :

1. que cualquier ejemplo que sea equiparable con todos los miembros de S, realmente se clasificaría como positivo.
2. que cualquier ejemplo que no se equipare con ninguno de los miembros de G realmente, se clasificaría como negativo.

(soleado, baja, baja, sí): Este ejemplo no se equipara con S_3 en el segundo atributo ('baja' frente a 'alta') por lo que no tenemos la certeza de que sea positivo. El ejemplo no se equipara con ningún (en este caso el único) miembro de G, por lo que este ejemplo sería **negativo**.
(soleado, alta, baja, no): Este ejemplo se equipara con todos los miembros (en este caso solo uno) de S_3, por lo que tenemos la certeza de que es **positivo**.
(soleado, alta, alta, no): Este ejemplo no se equipara con S_3 en el tercer atributo ('alta' frente a 'baja') por lo que no tenemos la certeza de que sea positivo. Por otro lado, el ejemplo se equipara con G, por lo que no tenemos la certeza de que el ejemplo sea negativo. Por tanto, no es posible clasificar este ejemplo con los datos que tenemos.

3)

e_4 = (lluvioso, alta, alta, sí, +)

$G_4 = G_3$ = { (Cielo, alta, Humedad, Viento) }
S_4 = { (**Cielo**, alta, **Humedad**, Viento) }

---

e_5 = (lluvioso, baja, alta, sí, –)

$S_5 = S_4$ = { (Cielo, alta, Humedad, Viento) }

$G_5 = G_4 = \{$ (Cielo, alta, Humedad, Viento) $\}$

$e_6 =$ (soleado, alta, alta, no, +)

$G_6 = G_5 = \{$ (Cielo, alta, Humedad, Viento) $\}$
$S_6 = S_5 = \{$ (Cielo, alta, Humedad, Viento) $\}$

EV = $\{$ (Cielo, alta, Humedad, Viento), (Cielo, alta, Humedad, Viento) $\}$

18F1

PREGUNTA 2. Valoración: 2

La tabla 1 muestra diferentes ejemplos positivos (+) y negativos (-) asociados al concepto: “Cuándo conceder un préstamo”). Cada uno de los ejemplos depende del valor de un conjunto de indicadores $\mathbf{I_i}$, con i=1,…,5. Aplique el **algoritmo de eliminación de candidatos (Espacio de Versiones)** a los datos almacenados en dicha tabla, suponiendo que éstos se presentan al algoritmo en el orden indicado. Especificar en cada paso los valores de los conjuntos S y G. Justifique si se llega o no a aprender el concepto buscado.

Ejemplos	I1	I2	I3	I4	I5	Concepto
1	Bajo	Hombre	No	Bajo	No	Conceder(-)
2	Alto	Mujer	No	Bajo	Sí	Conceder(-)
3	Alto	Mujer	No	Medio	No	Conceder(+)
4	Alto	Hombre	Sí	Medio	No	Conceder(+)
5	Alto	Hombre	Sí	Alto	No	Conceder(+)

Tabla 1

G_0 = { (I_1, I_2, I_3 ,I_4 ,I_5) }
S_0 = { Ø }

---

e_1 = (bajo, hombre, no, bajo, no, –)

S_1 = { Ø }
G_1 = { ~~(I_1, I_2, I_3 ,I_4 ,I_5)~~, hay inconsistencia en esta hipótesis de G_0 respecto a e_1 (describe el ejemplo negativo) por ello se elimina y se añaden las hipótesis mínimamente especializadas de la eliminada:

(alto, I_2, I_3, I_4, I_5),
(I_1, mujer, I_3, I_4, I_5),
(I_1, I_2, sí, I_4, I_5),
(I_1, I_2, I_3, medio, I_5),
(I_1, I_2, I_3, alto,I_5),
(I_1, I_2, I_3, I_4, sí), }

---

e_2 = (alto, mujer, no, bajo, sí, –)

S_2 = { Ø }
G_2 = { ~~(alto, I_2, I_3, I_4, I_5)~~, hay inconsistencia en esta hipótesis de G_1 respecto a e_2 (describe el ejemplo negativo) por ello se elimina y se añaden las hipótesis mínimamente especializadas de la eliminada:

(alto, hombre, I_3, I_4, I_5),
~~(alto, I_2, sí, I_4, I_5)~~, (existe hipótesis más general)
~~(alto, I_2, I_3, medio, I_5)~~, (existe hipótesis más general)
~~(alto, I_2, I_3, alto, I_5)~~, (existe hipótesis más general)
(alto, I_2, I_3, I_4, no),

~~(I_1, mujer, I_3, I_4, I_5)~~, hay inconsistencia en esta hipótesis de G_1 respecto a e_2 (describe el ejemplo negativo) por ello se elimina y se añaden las hipótesis mínimamente especializadas de la eliminada:

(bajo, mujer, I_3, I_4, I_5),
~~(I_1, mujer, sí, I_4, I_5)~~, (existe hipótesis más general)
~~(I_1, mujer, I_3, medio, I_5)~~, (existe hipótesis más general)
~~(I_1, mujer, I_3, alto, I_5)~~, (existe hipótesis más general)
(I_1, mujer, I_3, I_4, no),

(I_1, I_2, sí, I_4, I_5),
(I_1, I_2, I_3, medio, I_5),
(I_1, I_2, I_3, alto,I_5),
~~(I_1, I_2, I_3, I_4, sí)~~, hay inconsistencia en esta hipótesis de G_1 respecto a e_2 (describe el ejemplo negativo) por ello se elimina y se añaden las hipótesis mínimamente especializadas de la eliminada:

(bajo, I_2, I_3, I_4, sí),
(I_1, hombre, I_3, I_4, sí),
~~(I_1, I_2, sí, I_4, sí)~~, (existe hipótesis más general)
~~(I_1, I_2, I_3, medio, sí)~~, (existe hipótesis más general)
~~(I_1, I_2, I_3, alto, sí)~~ } (existe hipótesis más general)

---

e_3 = (alto, mujer, no, medio, no, +)

G_3 = { ~~(alto, hombre, I_3, I_4, I_5)~~, inconsistencia con e_3 , no lo describe.
(alto, I_2, I_3, I_4, no),
~~(bajo, mujer, I_3, I_4, I_5)~~, inconsistencia con e_3 , no lo describe.
(I_1, mujer, I_3, I_4, no),
~~(I_1, I_2, sí, I_4, I_5)~~, inconsistencia con e_3 , no lo describe.
(I_1, I_2, I_3, medio, I_5),
~~(I_1, I_2, I_3, alto,I_5)~~, inconsistencia con e_3 , no lo describe.
~~(bajo, I_2, I_3, I_4, sí)~~, inconsistencia con e_3 , no lo describe.
~~(I_1, hombre, I_3, I_4, sí)~~ } inconsistencia con e_3 , no lo describe.
S_3 = { (**alto**, **mujer**, **no**, **medio**, **no**) }, añadir las hipótesis mínimamente generalizada consistentes con e_3.

---

e_4 = (alto, hombre, sí, medio, no, +)

G_4 = { (alto, I_2, I_3, I_4, no),
~~(I_1, mujer, I_3, I_4, no)~~, inconsistencia con e_4 , no lo describe.
(I_1, I_2, I_3, medio, I_5) }
S_4 = { ~~(alto, mujer, no, medio, no)~~, inconsistencia con e_4 , no lo describe, por lo que se elimina la hipótesis y...
(alto, $\mathbf{I_2}$, $\mathbf{I_3}$, medio, no) } se añade la hipótesis mínimamente generalizada que cubre el ejemplo.

---

e_5 = (alto, hombre, sí, alto, no, +)

G_5 = { (alto, I_2, I_3, I_4, no),
~~(I_1, I_2, I_3, medio, I_5)~~ } inconsistencia con e_5 , no lo describe.
S_5 = { ~~(alto, I_2, I_3, medio, no)~~, inconsistencia con e_5 , no lo describe, se elimina la hipótesis y...
(alto, $\mathbf{I_2}$, $\mathbf{I_3}$, $\mathbf{I_4}$, no) } ...se añade la hipótesis minimamente generalizada para que cubra el ejemplo.

G_5 y S_5 coinciden, por lo que se logra aprender el concepto buscado.

18F2 – Pregunta 2a

PREGUNTA 1. Valoración: 2

Responda a las siguientes cuestiones:

a) Tras ejecutar el **algoritmo de eliminación de candidatos**, el valor del conjunto G={(x_1,A_{21},x_3)} y del conjunto S={(A_{11},A_{21},A_{31})}, donde A_{ij} corresponde al valor *j*-ésimo del atributo *i*-ésimo. Indique el conjunto de hipótesis que conforma el espacio de versiones (EV).

b) Justifique si en una **Red Neuronal Artificial (RNA),** que usa el algoritmo de retropropagación del error para aprender los pesos, se podría utilizar la función escalón como función de activación.

c) Muestre la función que expresa la salida, s_j, de una neurona perteneciente a una capa oculta de una **RNA**. Identifique todos los elementos expresados en dicha función.

d) En las **redes neuronales auto-organizadas de Kohonen**, los pesos de la célula ganadora, C_j, se actualizan de acuerdo a la siguiente regla hebbiana:

$$\frac{d\mu_{ij}}{dt} = \alpha(t)\tau_j(t)\left[\epsilon_i(t) - \mu_{ij}(t)\right]$$

Indique una función válida para actualizar el valor de la tasa de aprendizaje.

2.a)

$$G = \{ (x_1, A_{21}, x_3) \}$$
$$\downarrow$$
$$\{(x_1, A_{21}, A_{31})\ (A_{11}, A_{21}, x_3) \}$$
$$\downarrow$$
$$S = \{ (A_{11}, A_{21}, A_{31}) \}$$

18SO

PREGUNTA 2. Valoración: 3

La tabla 1 muestra diferentes ejemplos positivos (**+**) y negativos (**-**) asociados al concepto "Cuándo conceder un préstamo", dependiendo, cada uno de ellos, del valor de un conjunto de indicadores $\mathbf{I}_i$, con i=1,…,5. Conteste a las siguientes preguntas en el contexto del **algoritmo de eliminación de candidatos (Espacio de Versiones)**:

(i) Tras procesar el primer ejemplo, los conjuntos S y G quedan de la siguiente forma:

S_1=∅

G_1={(A,x2,x3,x4,x5), (x1,M,x3,x4,x5), (x1,x2,S,x4,x5), (x1,x2,x3,M,x5), (x1,x2,x3,A,x5), (x1,x2,x3,x4,S)}

Se pide obtener y justificar, de acuerdo a los diferentes pasos seguidos por el algoritmo, los valores de S y G al procesar el segundo ejemplo.

(ii) Tras procesar el último ejemplo, los conjuntos S y G quedan de la siguiente forma:

S_4={(A,x2,x3,M,N)}

G_4={(A,x2,x3,x4,N), (x1,x2,x3,M,x5)}

Indique y justifique cuáles son las hipótesis pertenecientes al espacio de versiones aprendido.

(iii) Desde un punto de vista genérico, existen tres tipos de *bias* para un algoritmo de aprendizaje: *bias* de representación, *bias* de preferencia o búsqueda y *bias* mixtas. Describa y justifique la ausencia o presencia de cada una de estas *bias* para el algoritmo indicado.

Ejemplos	I1	I2	I3	I4	I5	Concepto
1	Bajo	Hombre	No	Bajo	No	Conceder(**-**)
2	Alto	Mujer	No	Bajo	Sí	Conceder(**-**)
3	Alto	Mujer	No	Medio	No	Conceder(**+**)
4	Alto	Hombre	Sí	Medio	No	Conceder(**+**)

Tabla 1

i)

S_1 = Ø

G_1 = { (A, x2, x3, x4, x5), (x1, M, x3, x4, x5), (x1, x2, S, x4, x5), (x1, x2, x3, M, x5), (x1, x2, x3, A, x5), (x1, x2, x3,x4, S) }

---

e_2 = (A, M, N, B, S, –)

S_2 = S_1 = Ø

G_2 = { ~~(A, x2, x3, x4, x5)~~, hay inconsistencia en esta hipótesis de G_1 respecto a e_2 (describe el ejemplo negativo) por ello se elimina y se añaden las hipótesis mínimamente especializadas de la eliminada:

(A, H, x3, x4, x5),

~~(A, x2, S, x4, x5)~~, (existe hipótesis más general)

~~(A, x2, x3, M, x5)~~, (existe hipótesis más general)

(A, x2, x3, x4, N),

~~(x1, M, x3, x4, x5)~~, hay inconsistencia en esta hipótesis de G_1 respecto a e_2 (describe el ejemplo negativo) por ello se elimina y se añaden las hipótesis mínimamente especializadas de la eliminada:

(B, M, x3, x4, x5),
~~(x1, M, S, x4, x5)~~, (existe hipótesis más general)
~~(x1, M, x3, M, x5)~~, (existe hipótesis más general)
(x1, M, x3, x4, N),
(x1, x2, S, x4, x5),
(x1, x2, x3, M, x5),
(x1, x2, x3, A, x5),
~~(x1, x2, x3,x4, S)~~, hay inconsistencia en esta hipótesis de G_1 respecto a e_2 (describe el ejemplo negativo) por ello se elimina y se añaden las hipótesis mínimamente especializadas de la eliminada:
(B, x2, x3,x4, S),
(x1, H, x3,x4, S),
~~(x1, x2, S, x4, S)~~, (existe hipótesis más general)
~~(x1, x2, x3, M, S)~~ } (existe hipótesis más general)

---

e_2 ...
e_3 ...

---

ii)

e_4 = (A, H, M, S, S, –)

S_4 = { (A, x2, x3, M, N) }
G_4 = { (A, x2, x3, x4, N), (A, x2, x3, M, x5) }

G_4 = { (A, x2, x3, x4, N), (A, x2, x3, M, x5) }
↓
{ (A, x2, x3, B, N), (A, x2, x3, M, N), (A, x2, x3, M, S) }
↓
S_4 = { (A, x2, x3, M, N) }

iii)

Bias de preferencia o búsqueda: El EV no está sujeto a bias de preferencia.
Bias de representación: vendría especificado por el lenguaje de descripción de hipótesis.
Bias mixta:

19F1

PREGUNTA 2. Valoración: 3

La tabla 1 muestra diferentes ejemplos positivos (+) y negativos (-) asociados al concepto: "Detección de patología". Cada uno de los ejemplos depende del valor de un conjunto de síntomas $\mathbf{S_i}$, con i=1,…,5. Aplique el **algoritmo de eliminación de candidatos (Espacio de Versiones)** a los datos almacenados en dicha tabla, suponiendo que éstos se presentan al algoritmo en el orden indicado. Especificar en cada paso los valores de los conjuntos S y G. Justifique si se llega o no a aprender el concepto buscado.

Ejemplos	S1	S2	S3	S4	S5	Concepto
1	Agudo	Bajo	No	Bajo	No	patología(-)
2	Leve	Alto	No	Medio	No	patología (+)
3	Leve	Alto	No	Bajo	Sí	patología (-)
4	Agudo	Alto	Sí	Medio	No	patología (+)
5	Agudo	Alto	Sí	Alto	No	patología (+)

Tabla 1

$S_0 = \emptyset$
$G_0 = \{ (S1, S2, S3, S4, S5) \}$

---

e_1 = (agudo, bajo, no, bajo, no, –)

$S_1 = \emptyset$
G_1 = { ~~(S1, S2, S3, S4, S5)~~, hay inconsistencia en esta hipótesis de G_0 respecto a e_1 (describe el ejemplo negativo) por ello se elimina y se añaden las hipótesis mínimamente especializadas de la eliminada:
(leve, S2, S3, S4, S5),
(S1, alto, S3, S4, S5),
(S1, S2, sí, S4, S5),
(S1, S2, S3, medio, S5),
(S1, S2, S3, alto, S5),
(S1, S2, S3, S4, sí)}

---

e_2 = (leve, alto, no, medio, no, +)

G_2 = { (leve, S2, S3, S4, S5),
(S1, alto, S3, S4, S5),
~~(S1, S2, sí, S4, S5)~~, inconsistencia con e_2 , no lo describe.
(S1, S2, S3, medio, S5),
~~(S1, S2, S3, alto, S5)~~, inconsistencia con e_2 , no lo describe.
~~(S1, S2, S3, S4, sí)~~} inconsistencia con e_2 , no lo describe.
S_2 = { (**leve**, **alto**, **no**, **medio**, **no**) }

---

e_3 = (leve, alto, no, bajo, sí, –)

$S_3 = S_2$ = { (leve, alto, no, medio, no) }

G_3 = { ~~(leve, S2, S3, S4, S5)~~, hay inconsistencia en esta hipótesis de G_2 respecto a e_3 (describe el ejemplo negativo) por ello se elimina y se añaden las hipótesis mínimamente especializadas de la eliminada:

~~(leve, bajo, S3, S4, S5)~~, inconsistencia con S_3 , no lo describe.

~~(leve, S2, sí, S4, S5)~~, inconsistencia con S_3 , no lo describe.

~~(leve, S2, S3, medio, S5)~~, (existe hipótesis más general)

~~(leve, S2, S3, alto, S5)~~, inconsistencia con S_3 , no lo describe.

(leve, S2, S3, S4, no),

~~(S1, alto, S3, S4, S5)~~, hay inconsistencia en esta hipótesis de G_2 respecto a e_3 (describe el ejemplo negativo) por ello se elimina y se añaden las hipótesis mínimamente especializadas de la eliminada:

~~(agudo, alto, S3, S4, S5)~~, inconsistencia con S_3 , no lo describe.

~~(S1, alto, sí, S4, S5)~~, inconsistencia con S_3 , no lo describe.

~~(S1, alto, S3, medio, S5)~~, (existe hipótesis más general)

~~(S1, alto, S3, alto, S5)~~, inconsistencia con S_3 , no lo describe.

(S1, alto, S3, S4, no),

(S1, S2, S3, medio, S5)}

---

e_4 = (agudo, alto, sí, medio, no, +)

G_4 = { ~~(leve, S2, S3, S4, no)~~, inconsistencia con e_4 , no lo describe.

(S1, alto, S3, S4, no),

(S1, S2, S3, medio, S5) }

S_4 = { ~~(leve, alto, no, medio, no)~~, inconsistencia con e_4 , no lo describe, se elimina la hipótesis y...

(**S1**, alto, **S3**, medio, no) } se añade la hipótesis resultante de la generalización mínima de la hipótesis eliminada para que cubra el ejemplo.

---

e_5 = (agudo, alto, si, alto, no, +)

G_5 = { (S1, alto, S3, S4, no),

~~(S1, S2, S3, medio, S5)~~} inconsistencia con e_5, no lo describe.

S_5 = { ~~(S1, alto, S3, medio, no)~~, inconsistencia con e_5, no lo describe, se elimina la hipótesis y...

(S1, alto, S3, **S4**, no) } se añade la hipótesis resultante de la generalización mínima de la hipótesis eliminada para que cubra el ejemplo.

G_5 y S_5 son iguales, por lo que se consigue que se aprenda el concepto.

19F2

PREGUNTA 4. Valoración: 2

La tabla 2 muestra ejemplos positivos (+) y negativos (-) asociados al concepto: "Detección de patología". Cada uno de los ejemplos depende del valor de un conjunto de síntomas $\mathbf{S_i}$, con i=1,…,5. Aplique el **algoritmo de eliminación de candidatos (Espacio de Versiones)** sólo a los dos primeros ejemplos mostrados en dicha tabla, suponiendo que éstos se presentan al algoritmo en el orden indicado. Especificar en cada paso los valores de los conjuntos (S_1, G_1) y (S_2, G_2).

Ejemplos	S1	S2	S3	S4	S5	Concepto
1	Agudo	Bajo	No	Bajo	No	patología(-)
2	Leve	Alto	No	Bajo	Sí	patología (-)
3	Leve	Alto	No	Medio	No	patología (+)
4	Agudo	Alto	Sí	Medio	No	patología (+)
5	Agudo	Alto	Sí	Alto	No	patología (+)

Tabla 2

S_0 = Ø
G_0 = { (S1, S2, S3, S4, S5) }

---

e_1 = (agudo, bajo, no, bajo, no, –)

S_1 = Ø
G_1 = { ~~(S1, S2, S3, S4, S5)~~, hay inconsistencia en esta hipótesis de G_0 respecto a e_1 (describe el ejemplo negativo) por ello se elimina y se añaden las hipótesis mínimamente especializadas de la eliminada:

(leve, S2, S3, S4, S5),
(S1, alto, S3, S4, S5),
(S1, S2, sí, S4, S5),
(S1, S2, S3, medio, S5),
(S1, S2, S3, alto, S5),
(S1, S2, S3, S4, sí)}

---

e_2 = (leve, alto, no, bajo, sí, –)

S_2 = S_1 = Ø
G_2 = { ~~(leve, S2, S3, S4, S5)~~, hay inconsistencia en esta hipótesis de G_1 respecto a e_2 (describe el ejemplo negativo) por ello se elimina y se añaden las hipótesis mínimamente especializadas de la eliminada:

(leve, bajo, S3, S4, S5),
~~(leve, S2, sí, S4, S5)~~, (existe hipótesis más general)
~~(leve, S2, S3, medio, S5)~~, (existe hipótesis más general)
~~(leve, S2, S3, alto, S5)~~, (existe hipótesis más general)
(leve, S2, S3, S4, no),

~~(S1, alto, S3, S4, S5)~~, hay inconsistencia en esta hipótesis de G_1 respecto a e_2 (describe el ejemplo negativo) por ello se elimina y se añaden las hipótesis mínimamente especializadas de la eliminada:

(agudo, alto, S3, S4, S5),

~~(S1, alto, sí, S4, S5)~~, (existe hipótesis más general)
~~(S1, alto, S3, medio, S5)~~, (existe hipótesis más general)
~~(S1, alto, S3, alto, S5)~~, (existe hipótesis más general)
(S1, alto, S3, S4, no),
(S1, S2, sí, S4, S5),
(S1, S2, S3, medio, S5),
(S1, S2, S3, alto, S5),
~~(S1, S2, S3, S4, sí)~~, hay inconsistencia en esta hipótesis de G_1 respecto a e_2 (describe el ejemplo negativo) por ello se elimina y se añaden las hipótesis mínimamente especializadas de la eliminada:
(agudo, S2, S3, S4, sí),
(S1, bajo, S3, S4, sí),
~~(S1, S2, sí, S4, sí)~~, (existe hipótesis más general)
~~(S1, S2, S3, medio, sí)~~, (existe hipótesis más general)
~~(S1, S2, S3, alto, sí)~~ } (existe hipótesis más general)

19SO – Pregunta 1

PREGUNTA 1. Valoración: 2

Describa las dos siguientes características del **algoritmo espacio de versiones**:

a) *Bías.*

b) Complejidad: espacial y temporal.

a) Bias.

No está sujeto a restricciones o bias de preferencia en su espacio de búsqueda, salvo la aportada por el propio lenguaje de descripción del concepto, que debe ser lo más restringido posible pero abarcando el concepto.

b) Su complejidad espacial y temporal.

El espacio de almacenamiento necesario es del orden de $O(s+g)$, donde **s** es el tamaño máximo del conjunto **S** y **g** es el tamaño máximo alcanzado por **G**.

El tiempo crece polinómicamente con el número de instancias de entrenamientos (**p** + **n**, instacias positivas más negativas), exactamente es de $O(sg(p+n) + s^2p + g^2n)$.

En definitiva, la eficiciencia dependerá de los tamaños de **G** y **S**.

14SO

PREGUNTA 1. Valoración: 3

Aprendizaje de reglas de clasificación: A^Q.

(a) Defina los conceptos de selector, complejo, recubrimiento y función LEF. Ponga ejemplos de cada uno de ellos. Qué ocurriría si al aplicar la función LEF a un conjunto de complejos no la verifica ninguno de ellos o si, por el contrario, hay más de un complejo que la verifica.

(b) Describa la tarea que realizan los sistemas A^Q como un proceso que realiza dos tipos de búsqueda, Para ello, especifique el conjunto de estados, el conjunto de operadores, el estado inicial, la meta y la heurística utilizados en cada una de ellas.

(c) Describa el proceso de búsqueda externa en pseudocódigo.

a)

Selector: concepto que permite realizar preguntas sobre valores de los atributos de la entrada
Complejo: es la conjunción de varios selectores.
Recubrimiento: es una disyunción de complejos que se usa para cubrir los ejemplos positivos
Función LEF: función suministrado por el usuario para la selección, en base a unos criterios, de las reglas candidatas que describen un ejemplo.

b)

- Búsqueda interna:

Conjunto de estados: conjunto de complejos
Conjunto de Operadores: único, el de 'incorporar un selector a algún complejo del estado anterior
Estado inicial: conjunto de complejos con un solo elemento, el complejo vacío
Estado Meta: un conjunto de complejos que describen a todos los ejemplos positivos pero a ninguno negativo.
Heurística: evaluar los complejos de acuerdo a la función LEF

- Búsqueda externa:

Conjunto de estados: cada estado será un recubrimiento que cubre parcial o totalmente los ejemplos
Conjunto de Operadores: único, 'incorporar un complejo al recubrimiento'
Estado inicial: recubrimiento vacío
Estado Meta: recubrimiento que describa correctamente las instancias de entrada
Heurística: seleccionar el complejo que maximice la heurísitica proporcionada por el usuario a través de la LEF

c)

Comenzamos con un recubrimiento formado por el conjunto vacío.
Mientras haya elementos positivos en la muestra, elegimos el primero de ellos, calculamos su estrella (formada por los complejos que pueden describir el ejemplo positivo pero ninguno negativo) y elegimos uno de ellos de acuerdo a las expecificaciones de la LEF. Antes de volver al comienzo del bucle, introducimos el complejo elegido en el recubrimiento y sacamos de la muestra todos los ejemplos cubiertos por dicho complejo.

15F2

PREGUNTA 2. Valoración: 2

Haciendo uso del **algoritmo A^Q**, se desea aprender el concepto de paciente patológico a partir de los valores de una serie de síntomas. Los ejemplos de entrenamiento aparecen en la tabla inferior. Realice sólo el proceso que implica obtener aquella estrella que describe el primer ejemplo positivo de la tabla y no describa ningún ejemplo negativo. Para ello, parta de una estrella, *E*, vacía, de una lista, *L*, que contiene el complejo vacío y de un conjunto de selectores, *S*, que contiene los valores de los atributos de la semilla. Muestre los sucesivos pasos empleados por el algoritmo y la actualización de *E*, *S* y *L*. Utilice la siguiente función LEF:

LEF = {(cobertura, 1), (número de premisas, 3)}

Paciente	Síntoma$_1$	Síntoma$_2$	Síntoma$_3$	Síntoma$_4$	**Patología**
1	0	0	1	0	Sí
2	0	0	0	1	No
3	0	1	2	0	Sí
4	1	1	0	0	Sí
5	0	2	0	0	No
6	2	1	0	0	No
7	1	1	1	0	Sí

Síntoma$_i$=0 ⇒ Síntoma$_i$ ausente.
Síntoma$_i$=1 ⇒ Síntoma$_i$ presente.
Síntoma$_i$=2 ⇒ Síntoma$_i$ actualmente ausente pero apareció en el último mes.

E = {C: cualquier hipótesis presenta patología}
L: [()]

e1: (0,0,1,0)

semilla S: {(S1=0), (S2=0), (S3=1), (S4=0)}

C1: (S1=0), cubre ejemplos negativos, hay que especializarlo.
C2: (S2=0), cubre ejemplos negativos, hay que especializarlo.
C3: (S3=1), ok, pasa al conjunto E estrella.
C4: (S4=1), cubre ejemplos negativos, hay que especializarlo.

E = { C3 }

L' : { C1, C2, C4 }

L = L'

Especialización de elementos de L
C12: (S1=0) ^ (S2=0), cubre ejemplos negativos, hay que especializarlo.
C13: (S1=0) ^ (S3=1), descartado por incluir a C3
C14: (S1=0) ^ (S4=0), cubre ejemplos negativos, hay que especializarlo.
C24: (S2=0) ^ (S4=0), ok, pasa a la estrella

C21, C41 y C42, no se consideran al ser casos ya estudiados (C12, C14,y C24)
C11, C22 y C44, no se consideran por ser los mismos elementos que se quiere especializar.
C23 y C43 tampoco se consideran ya que incluye C3 ya en la estrella

E = { C3, C24 }

L' = { C12, C14 }
L = L'

> Especialización de L:
> C121 y C122 no se consideran, por ser los mismos elementos que se quieren especializar.
> C123 no se considera ya que incluye a C3, ya en la estrella.
> C124 no se considera porque incluye a C24, ya en la estrella.
> C141 y C144 no se consideran, por ser los mismos elementos que se quieren especializar.
> C142 no se considera porque incluye a C24, ya en la estrella.
> C143 no se considera ya que incluye a C3, ya en la estrella.

No se añade ningún complejo por estar todos formados por otros complejos ya contenidos en E. No se sigue iterando porque: los nuevos complejos estarían formados por 4 selectores y todos incumplirían LEF.

La estrella resultante sería, por tanto, E = { C1, C24 }.

Se añadiría C1 al recubrimiento ya que cubre 2 ejemplos positivos frente a C24, que solo cubre 1.

16F2

<u>PREGUNTA 1</u>. Valoración: 1.5

Conteste las siguientes cuestiones relacionadas con el algoritmo A^Q:

(a) *Bias* del algoritmo.

(b) Cómo se comporta ante el ruido en entradas y en estructura.

(c) Cuál es la complejidad espacial y temporal (tanto para el mejor como peor caso).

a)

Bias del algoritmo: la función LEF, proporcionada por el usuario, que representa los criterios de selección del complejo apropiado para pasar a formar parte del recubrimiento.

b)

Los sistemas Aq no tratan ejemplos con ruido.

La salida sí es tolerante a fallos porque la eliminación de algún complejo del recubrimiento permite todavía que los otros complejos restantes puedan seguir clasificando ejemplos.

c)

La complejidad espacial viene marcada por el conjunto de recubrimiento: en el mejor de los casos este conjunto estará vacío (todos los ejemplos positivos) y en el peor de los casos tendrá un complejo por cada ejemplo (espacio proporcional, por tanto, a |A| * |P|, con A siendo el número de atributos y P es el conjunto de atributos positivos).

La complejidad temporal depende a = |A|, número de atributos, y de n = |N|, número de ejemplos negativos.

1º) Generación de los selectores (proporcional a 'a')

2º) Añadir selector a cada posible complejo, en el peor de los casos de 'a' selectores por complejo. Como esto conlleva una busqueda con un factor de ramificación de 'a', y en cada paso hay que realizar un recorrido por todos los complejos para ver si alguno cubre a algún ejemplo negativo (n*a), será proporcional a $O(\Sigma\ a^i * n * a)$. Haciendo esto para cada ejemplo positivo tendríamos $O(p * \Sigma\ a^i * n * a)$. En el mejor de los casos todos los complejos con un sólo selector que se pueden formar con el primer ejemplo positivo serviría para cubrir todos los ejemplos positivos y ninguno negativo, por lo que el cálculo de la estrella acabaría en el primer ciclo y tendríamos un coste temporal de $O(\Sigma\ a * a * n)$

17F2

PREGUNTA 2. Valoración: 2.5

Se desea usar el **algoritmo A^Q** para aprender el concepto de "máquina averiada" a partir del valor proporcionado por cuatro leds indicadores. Los ejemplos de entrenamiento aparecen en la tabla 1. Mostrar cómo evoluciona la estrella y el recubrimiento en cada paso del algoritmo. Finalmente, muestre el valor del recubrimiento que permite realizar el proceso de clasificación. Utilizar la siguiente función LEF:

LEF = {(cobertura≥1), (número de premisas≤2)}

Paciente	Led_1	Led_2	Led_3	Led_4	Avería
1	0	1	1	1	Sí
2	1	0	0	0	No
3	0	0	1	2	Sí
4	0	0	0	1	Sí
5	0	1	1	0	Sí
6	0	0	2	0	No
7	0	2	1	0	No
Led_i=0 ⇒ El Led_i está apagado. Led_i=1 ⇒ El Led_i está encendido. Led_i=2 ⇒ El Led_i está intermitente.					

Tabla 1

E = {C: cualquier hipótesis presenta avería}
L: [()]

e1: (0,1,1,1)

semilla S: {(Led1=0), (Led2=1), (Led3=1), (Led4=1)}

C1: (Led1 = 0), cubre ejemplos negativos, hay que especializarlo
C2: (Led2 = 1), ok, pasa a la estrella
C3: (Led3 = 1), cubre ejemplos negativos, hay que especializarlo
C4: (Led4 = 1), ok, pasa a la estrella

E = { C2, C4 }
L' = {C1, C3 }
L = L'

Especializaciones de L

C12: se descarta, incluye C2 que está en la estrella
C13: cubre ejemplos negativos, hay que especializarlo
C14: se descarta, incluye C4 que está en la estrella

C31: ya estudiado como C13
C32: se descarta, incluye C2 que está en la estrella
C34: se descarta, incluye C4 que está en la estrella

E = { C2, C4 }
L' = { C13 }
L = L'

Especializaciones de L

La función LEF no permite más de 2 premisas, por lo que se descarta cualquier especialización.

L = { vacío }

Se aplica la función LEF a los complejos de la estrella para elegir aquel que pasará a formar parte del recubrimiento.

C2	(Led2 = 1)	cobertura: 2	nº premisas: 1
C4	(Led4 = 1)	cobertura: 2	nº premisas: 1

Ambas cumplen la función LEF con las mismas características. Elijo aleatoriamente (por ejemplo, la primera) a C2.
R = { **(Led2 = 1)** }, que cubre los ejemplos positivos 1 y 5. Los descarto y analizo el siguiente elemento positivo.

Se pasa a estudiar el ejemplo 3:

e3: (0,0,1,2)

semilla S: {(Led1=0), (Led2=0), (Led3=1), (Led4=2)}

C1: (Led1 = 0), cubre ejemplos negativos, hay que especializarlo
C2: (Led2 = 0), cubre ejemplos negativos, hay que especializarlo
C3: (Led3 = 1), cubre ejemplos negativos, hay que especializarlo
C4: (Led4 = 2), ok, pasa a la estrella

E = { C4 }
L' = {C1, C2, C3 }
L = L'

Especializaciones de L

C11: se descarta, es el mismo elemento que se quiere especializar
C12: (Led1 = 0) ^ (Led2 = 0), cubre ejemplos negativos (ejemplo 6), hay que especializarlo
C13: (Led1 = 0) ^ (Led3 = 1), cubre ejemplos negativos (ejemplo 7), hay que especializarlo
C14: se descarta, incluye C4 que está en la estrella

C21: se descarta, ya estudiado como C12
C22: se descarta, es el mismo elemento que se quiere especializar
C23: ok, pasa a la estrella
C24: se descarta, incluye C4 que está en la estrella

C31: ya estudiado como C13
C32: se descarta, ya estudiado como C23
C33: se descarta, es el mismo elemento que se quiere especializar
C34: se descarta, incluye C4 que está en la estrella

E = { C4, C23 }
L' = { C12, C13 }

L = L'

Especialización de L:
la función LEF no permite más de 2 premisa, se descarta cualquier especialización.

L = { vacío }

Se aplica la función LEF a los complejos de la estrella para elegir aquel que pasará a formar parte del recubrimiento.

C4	(Led4 = 2)	cobertura: 1	nº premisas: 1
C23	(Led2 = 0) ^ (Led3 = 1)	cobertura: 1	nº premisas: 2

Las 2 cumplen la función LEF y pero C4 tiene mejores características, por lo que la elijo para formar parte del recubrimiento..
R = { (Led2 = 1), **(Led4 = 2)**}, que cubre los ejemplos positivos 1, **3** y 5. Los descarto y analizo el siguiente elemento positivo.

Se pasa a estudiar el ejemplo 4:

e4: (0,0,0,1)

semilla S: {(Led1=0), (Led2=0), (Led3=0), (Led4=1)}

C1: (Led1 = 0), cubre ejemplos negativos (ejemplos 6 y 7), hay que especializarlo
C2: (Led2 = 0), cubre ejemplos negativos (ejemplos 2 y 6), hay que especializarlo
C3: (Led3 = 0), cubre ejemplos negativos (ejemplo 2), hay que especializarlo
C4: (Led4 = 1), ok, pasa a la estrella

E = { C4 }
L' = {C1, C2, C3 }
L = L'

Especializaciones de L
C11: se descarta, es el mismo elemento que se quiere especializar
C12: (Led1 = 0) ^ (Led2 = 0), cubre ejemplos negativos (ejemplo 6), hay que especializarlo
C13: (Led1 = 0) ^ (Led3 = 0), ok, pasa a la estrella
C14: se descarta, inlcuye a C4 que está en la estrella

C21: se descarta, ya estudiado como C12
C22: se descarta, es el mismo elemento que se quiere especializar
C23: cubre ejemplos negativos (ejemplo 6), hay que especializarlo
C24: se descarta, incluye a C4 que está en la estrella

C31: ya estudiado como C13
C32: se descarta, ya estudiado como C23
C33: se descarta, es el mismo elemento que se quiere especializar
C34: se descarta, incluye C4 que está en la estrella

E = { C4, C13 }
L' = {C12, C23 }

L = L'

Especializaciones de L
la función LEF no permite más de 2 premisa, se descarta cualquier especialización.

L = { vacío }

Se aplica la función LEF a los complejos de la estrella para elegir aquel que pasará a formar parte del recubrimiento.

C4	(Led4 = 1)	cobertura: 2	nº premisas: 1
C13	(Led1 = 0) ^ (Led3 = 1)	cobertura: 1	nº premisas: 2

Las 2 cumplen la función LEF y pero elijo C4 que tiene mejores caraterísticas (menor número de premisas para igual cobertura) y la introduzco en el recubrimiento.

R = { (Led2 = 1), (Led4 = 2) , **(Led4 = 1)**}, que cubre todos los ejemplos positivos y nos da la solución.

18F2

PREGUNTA 3. Valoración: 3

Se desea usar el **algoritmo A^Q** para aprender el concepto de "producto defectuoso" a partir del valor proporcionado por cuatro sensores, S_i, con $i=1,\dots,4$, instalados en una cadena de montaje. Los ejemplos de entrenamiento aparecen en la tabla 1. Mostrar cómo evoluciona la estrella y el recubrimiento en cada paso del algoritmo. Finalmente, muestre el valor del recubrimiento que permite realizar el proceso de clasificación. Utilizar la siguiente función LEF:

$$LEF = \{(cobertura,1),\ (número\ de\ premisas,2)\}$$

Ejemplo	S_1	S_2	S_3	S_4	Producto Defectuoso
1	0	1	1	1	Sí
2	1	0	0	0	No
3	0	1	1	2	Sí
4	0	0	0	1	Sí
5	0	1	1	0	Sí
6	0	0	2	0	No
7	0	2	1	0	No

Tabla 1

E = {C: cualquier producto es defectuoso}
L: [()]

e1: (0,1,1,1)

semilla S: {(S1=0), (S2=1), (S3=1), (S4=1)}

C1: (S1 = 0), cubre ejemplos negativos, hay que especializarlo
C2: (S2 = 1), ok, pasa a la estrella
C3: (S3 = 1), cubre ejemplos negativos (ejemplo 7), hay que especializarlo
C4: (S4 = 1), ok, pasa a la estrella

E = { C2, C4 }
L' = {C1, C3 }
L = L'

Especializaciones de L

C11: se descarta, es el mismo ejemplo que se quiere especializar
C12: se descarta, incluye C2 que está en la estrella
C13: cubre ejemplos negativos (ejemplo 7), hay que especializarlo
C14: se descarta, incluye a C4, ya en la estrella

C31: ya estudiado como C13
C32: se descarta, incluye C2 que está en la estrella
C33: se descarta, es el mismo ejemplo que se quiere especializar
C34: se descarta, incluye C4 que está en la estrella

E = { C2, C4 }
L' = { C13 }
L = L'

Especialiación de L
La función LEF no permite más de 2 premisas, por lo que se descarta a este nivel cualquier especialización (darían reglas de, al menos, 3 premisas).

Aplicamos la función LEF a las dos reglas de la estrella y tenemos:

C2	(S2 = 1)	cobertura: 3	nº premisas: 1
C4	(S4 = 1)	cobertura: 2	nº premisas: 2

Ambas cumplen los requerimientos de la función LEF, pero elegimos a C2 para entrar a formar parte del recubrimiento por tener mejores características (mayor ejemplo cubiertos).

R = { **(S2 = 1)** }
Se eliminan los ejemplos positivos cubiertos y se pasa a analizar el siguiente ejemplo positivo disponible, el ejemplo 4

e3: (0, 0, 0, 1)

semilla S = { (S1=0, S2=0, S3=0, S4=1) }

C1: S1 = 0, describe ejemplos negativos, hay que especializarlo.
C2: S2 = 0, describe ejemplos negativos, hay que especializarlo.
C3: S3 = 0, describe ejemplos negativos, hay que especializarlo.
C4: S4 = 1, ok, pasa a la estrella

E = { C4 }
L' = { C1, C2, C3 }

L = L'

Especialización de L:
C11, no se trata al ser mismo elemento que se quiere especializar.
C12, describe ejemplo negativo (ejemplo 6), hay que especializarlo
C13, ok, pasa a la estrella
C14 no se trata por contener a C4, ya en la estrella

C21, se descarta por ser igual a C12, ya tratado
C22, no se trata al ser mismo elemento que se quiere especializar
C23, describe ejemplos negativo (ejemplo 2), hay que especializar
C24, no se trata por contener a C4, ya en la estrella

C31, se descarta por ser igual a C13, ya tratado
C32, se descarta por ser igual a C23, ya tratado
C33, no se trata por ser el mismo elemento que se quiere especializar.
C34, no se trata por contener a C4, ya en la estrella

E = { C4, C13 }
L' = { C12, C23 }

L = L'

Especialización de L

La LEF indica que las reglas no pueden tener más de 2 premisas, por lo que en este punto no se realiza ninguna especialización.

L = { vacío }

Aplicamos la función LEF a las dos reglas de la estrella y tenemos:

C4	(S4 = 1)	cobertura: 1	nº premisas: 1
C13	(S1 = 0) ^ (S3 = 0)	cobertura: 1	nº premisas: 2

Ambas cumplen los requerimientos de la función LEF, pero elegimos a C4 para entrar a formar parte del recubrimiento por tener mejores características (menores premisas).

R = { (S2 = 1), **(S4 = 1)** }

Se eliminan los ejemplos positivos cubiertos y se pasa a analizar el siguiente ejemplo positivo. Al no haberlo, el recubrimiento anterior es la solución.

19F1

PREGUNTA 4. Valoración: 3

Se desea usar el **algoritmo A^Q** para aprender el concepto de "producto defectuoso" a partir del valor proporcionado por cuatro sensores, S_i, con i=1,…,4, instalados en una cadena de montaje. Los ejemplos de entrenamiento aparecen en la tabla 3. Mostrar cómo evoluciona la estrella y el recubrimiento en cada paso del algoritmo. Finalmente, muestre el valor del recubrimiento que permite realizar el proceso de clasificación. Utilizar la siguiente función LEF:

LEF = {(cobertura,1), (número de premisas,2)}

Ejemplo	S_1	S_2	S_3	S_4	Producto Defectuoso
1	1	1	1	0	Sí
2	0	0	0	1	No
3	2	1	1	0	Sí
4	1	0	0	0	Sí
5	0	1	1	0	Sí
6	0	2	0	0	No
7	0	1	2	0	No

Tabla 3

E = {C: cualquier producto es defectuoso}
L: [()]

e1: (1,1,1,0)

semilla S: { (S1=1), (S2=1), (S3=1), (S4=0) }

C1: (S1 = 1), ok, pasa a la estrella
C2: (S2 = 1), cubre ejemplos negativos (ejemplo 7), hay que especializarlo
C3: (S3 = 1), ok, pasa a la estrella
C4: (S4 = 0), cubre ejemplos negativos (ejemplos 6 y 7), hay que especializarlo

E = { C1, C3 }
L' = {C2, C4 }
L = L'

Especializaciones de L
C21: se descarta, incluye C1 que está en la estrella
C22: se descarta, es el mismo ejemplo que se quiere especializar
C23: se descarta, incluye C3 que está en la estrella
C24: cubre ejemplos negativos (ejemplo 7), hay que especializarlo

C41: se descarta, incluye C1 que está en la estrella
C42: se descarta, ya tratado como C24
C43: se descarta, incluye C3 que está en la estrella
C44: se descarta, es el mismo ejemplo que se quiere especializar

E = { C1, C3 }
L' = { C24 }

L = L'

Especialización de L

La función LEF no permite reglas de más de 2 premisas, por lo que en este punto no se tratan más especializaciones.

L = { vacío }

Aplicamos la función LEF a las dos reglas de la estrella y tenemos:

C1	(S1 = 1)	cobertura: 2	nº premisas: 1
C3	(S3 = 1)	cobertura: 3	nº premisas: 1

Ambas cumplen los requerimientos de la función LEF, pero elegimos a C3 para entrar a formar parte del recubrimiento por tener mejores características (mayor número de ejemplos cubiertos).

R = { **(S3 = 1)** }

Eliminamos ejemplos positivos cubiertos por R y evaluamos siguiente ejemplo positivo disponible, en este caso el ejemplo 4
e4: (1,0,0,0)

semilla S: { (S1=1), (S2=0), (S3=0), (S4=0) }

C1: (S1 = 1), ok, pasa a la estrella
C2: (S2 = 0), cubre ejemplos negativos (ejemplo 2), hay que especializarlo
C3: (S3 = 0), cubre ejemplos negativos (ejemplos 2 y 6), hay que especializarlo
C4: (S4 = 0), cubre ejemplos negativos (ejemplos 6 y 7), hay que especializarlo

E = { C1 }
L' = {C2, C3, C4 }
L = L'

Especializaciones de L
C21: se descarta, incluye C1 que está en la estrella
C22: se descarta, es el mismo ejemplo que se quiere especializar
C23: cubre ejemplos negativos (ejemplo 2), hay que especializarlo
C24: ok, pasa a la estrella

C31: se descarta, incluye C1 que está en la estrella
C32: se descarta, ya fue estudiado en C23
C33: se descarta, es el mismo ejemplo que se quiere especializar
C34: cubre ejemplos negativos (ejemplo 6), hay que especializarlo

C41: se descarta, incluye C1 que está en la estrella
C42: se descarta, ya tratado como C24
C43: se descarta, ya tratado como C34
C44: se descarta, es el mismo ejemplo que se quiere especializar

E = { C1, C24 }
L' = { C23, C34 }

L = L'

Especialización de L

La función LEF no permite reglas de más de 2 premisas, por lo que en este punto no se tratan más especializaciones.

L = { vacío }

Aplicamos la función LEF a las dos reglas de la estrella y tenemos:

C1	(S1 = 1)	cobertura: 1	nº premisas: 1
C24	(S2 = 0) ^ (S4 = 0)	cobertura: 1	nº premisas: 2

Ambas cumplen los requerimientos de la función LEF, pero elegimos a C1 para entrar a formar parte del recubrimiento por tener mejores características (menor número de premisas).

R = { (S3 = 1), **(S1 = 1)**}

Eliminamos ejemplos positivos cubiertos por R y evaluamos siguiente ejemplo positivo disponible, como no hay ninguno más disponible, el recubrimiento anterior es la solución.

14F1

PREGUNTA 3. Valoración: 3

Dado el grafo de la figura en el que los arcos que conectan los nodos se describen mediante el predicado *conectado(A,B)*, utilizar el **algoritmo FOIL** para obtener la descripción, en forma de cláusulas de Horn, del concepto "Cuándo un nodo *D* es alcanzable desde otro *C*", es decir, cuándo hay un camino que los conecta. Dicho concepto será representado por el predicado *alcanzable(C,D)*. Para simplificar el problema, de todos los posibles literales, supóngase que los literales que producen las tres mayores ganancias son:

{*Conectado(A, B), Conectado(A, C), Conectado(C, B)*}

Se pide, para cada literal, obtener:

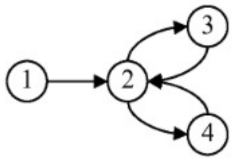
(a) Los conjuntos $\mathbf{E_0}^+$ y $\mathbf{E_0}^-$.

(b) Los conjuntos $\mathbf{E_0}^{++}$, $\mathbf{E_1}^+$ y $\mathbf{E_1}^-$ como resultado de añadir, en cada caso, cada uno de los literales indicados.

(c) La ganancia asociada al literal.

¿Cuál es la regla obtenida?

¿Es necesario realizar una nueva iteración del algoritmo FOIL para obtener una regla con un antecedente con dos literales? Justifique su respuesta.



Disponemos de la información del dominio de los conectados:

Conectado: [(1,2), (2,3), (2,4), (3,2), (4,2)]

Podemos obtener los conjuntos de ejemplos positivos (E^+) y negativos (E^-) del concepto "Alcanzable" a aprender:

E^+ = [(1,2), (1,3), (1,4), (2,2), (2,3), (2,4), (3,2), (3,3), (3,4), (4,2), (4,3), (4,4)]
Card(E^+) = 12

E^- = [(1,1), (2,1), (3,1), (4,1)]
Card(E^-) = 4

Por el enunciado suponemos que los 3 literales que producen mayor ganancia son:

{ Conectado(A,B), Conectado(A,C), Conectado(C,B) }

Si extendemos la regla utilizando el primer literal, L_1 = Conectado(A,B), es decir,

Alcanzable(C,D) :- Conectado(A,B)

Debido a que incumple la restricción del lenguaje de que al menos una de las variables entre el antecedente y el precedente debe ser común, esta regla no es válida, por lo que no se tiene en consideración.

Si extendemos la regla utilizando el segundo literal, L_2 = Conectado(A,C), es decir,

Alcanzable(C,D) :- Conectado(A,C)

Para calcular los ejemplos positivos de la nueva regla tenemos que ver una tupla donde A y C estén conectados y desde C se pueda alcanzar D. Hay que tener en cuenta que A, C y D deben ser todos distintos entre sí:

Conectado(A, C): [(1,2), (2,3), (2,4), (3,2), (4,2)]
Alcanzable(C, D): [(1,2), (1,3), (1,4), (2,2), (2,3), (2,4), (3,2), (3,3), (3,4), (4,2), (4,3), (4,4)]

E_L^+ tendrá la forma de una tupla (A,C,D)
E_L^+ = [(1,2,2), (1,2,3), (1,2,4),
(2,3,2), (2,3,3), (2,3,4),
(2,4,1), (2,4,2), (2,4,3),
(3,2,2), (3,2,3), (3,2,4),
(4,2,2), (4,2,3), (4,2,4)]   Card(E_L^+) = 15

Para calcular los ejemplos negativos de la nueva regla tenemos que ver una tupla donde A y C están conectados pero que C y D no son alcanzables.

Conectado(A, C): [(1,2), (2,3), (2,4), (3,2), (4,2)]
! Alcanzable(C, D): [(1,1), (2,1), (3,1), (4,1)]

E_L^- tendrá la forma de una tupla (A,C,D)
E_L^- = [(1,2,1),
(2,3,1),
(2,4,1),
(3,2,1),
(4,2,1)]
Card(E_L^-) = 5

$E_K = E^+$ => Card(E_K) = 12 (número de ejemplos positivos cubiertos por la regla)

$$G(L_2) = Card(E_k) \cdot [-\log_2(\frac{Card(E^+)}{Card(E^+) + Card(E^-)}) + \log_2(\frac{Card(E_L^+)}{Card(E_L^+) + Card(E_L^-)})]$$

$G(L_2)$ = 12 · [- $\log_2$(12 / 12 + 4) + $\log_2$(15 / 15 + 5)] = 12 · [0.4150 – 0.4150] = **0**

Si extendemos la regla utilizando el tercer literal, L_3 = Conectado(C,B), es decir,

Alcanzable(C,D) :- Conectado(C,B)

Para calcular los ejemplos positivos de la nueva regla tenemos que ver una tupla donde C y B estén conectados y C y D sean alcanzables. Hay que tener en cuenta que B, C y D deben ser todos distintos entre sí:

Conectado(C, B): [(1,2), (2,3), (2,4), (3,2), (4,2)]
Alcanzable(C, D): [(1,2), (1,3), (1,4), (2,2), (2,3), (2,4), (3,2), (3,3), (3,4), (4,2), (4,3), (4,4)]

E_L^+ tendrá la forma de una tupla (C,B,D)
E_L^+ = [(1,2,2), (1,2,3), (1,2,4),
(2,3,2), (2,3,3), (2,3,4),
(2,4,2), (2,4,3), (2,4,4),
(3,2,2), (3,2,3), (3,2,4),
(4,2,2), (4,2,3), (4,2,4)]
Card(E_L^+) = 15

Para calcular los ejemplos negativos de la nueva regla tenemos que ver una tupla donde A y C no están conectados o C y D no son alcanzables. Y también que A, C y D sean valores distintos.

Conectado(C, B): [(1,2), (2,3), (2,4), (3,2), (4,2)]
! Alcanzable(C,D): [(1,1), (2,1), (3,1), (4,1)]

E_L^- tendrá la forma de una tupla (C,B,D)
E_L^- = [(1,2,1),
(2,3,1),
(2,4,1),
(3,2,1),
(4,2,1)]   Card(E_L^-) = 5
$E_K = E^+$ => Card(E_K) = 12 (número de ejemplos positivos cubiertos por la regla)

G(L_3) = $Card(E_k) \cdot [-\log_2(\frac{Card(E^+)}{Card(E^+) + Card(E^-)}) + \log_2(\frac{Card(E_L^+)}{Card(E_L^+) + Card(E_L^-)})]$
G(L_3) = 12 · [- $\log_2$(12 / 12 + 4) + $\log_2$(15 / 15 + 5)] = 12 · [0.4150 – 0.4150] = **0**

La regla obtenida sería la que tenga el literal de mayor ganancia en el antecedente, pero en este caso, los 3 literales tienen la misma ganancia, así que elijo uno aleatoriamente. Por ejemplo:

Alcanzable(C,D) :- Conectado(C,B)

14F2 y parcialmente 16SR y 19SR (figura 1)

PREGUNTA 3. Valoración: 4

1. Dado el grafo de la figura 1 en el que los arcos que conectan los nodos se describen mediante el predicado *conectado(A,B)*, utilizar el **algoritmo FOIL** para obtener la descripción, en forma de cláusulas de Horn, del concepto "cuándo un nodo *Y* es alcanzable desde otro *X*", es decir, cuándo hay un camino que los conecta. Dicho concepto será representado por el predicado *alcanzable(X,Y)*. Para simplificar el problema, de todos los posibles literales, supóngase que los literales que producen las tres mayores ganancias son:

   {*Conectado(A, B), Conectado(A, C), Conectado(C, B)*}

   Se pide, para cada literal, obtener:
   (a) Los conjuntos $\mathbf{E_0}^+$ y $\mathbf{E_0}^-$.
   (b) Los conjuntos $\mathbf{E_0}^{++}$, $\mathbf{E_1}^+$ y $\mathbf{E_1}^-$ como resultado de añadir, en cada caso, cada uno de los literales indicados.
   (c) La ganancia asociada al literal.

   ¿Con el análisis realizado, se puede añadir ya alguna regla definitiva a la base de reglas? Si la respuesta a la anterior pregunta fue afirmativa, ¿la regla obtenida aprende el concepto buscado o sería necesario seguir iterando? Justifique su respuesta.

2. Repetir todo el proceso para el grafo de la figura 2.

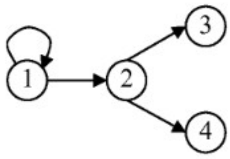


Figura 1

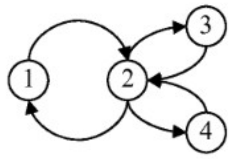


Figura 2

FIGURA 1

Predicado a aprender:
Alcanzable: [(1,1), (1,2), (1,3), (1,4), (2,3), (2,4)]

Dominio:
Conectado: [(1,1), (1,2), (2,3), (2,4)]

El conjunto de ejemplos positivos del concepto meta, Alcanzable, es:
E^+ = [(1,1), (1,2), (1,3), (1,4), (2,3), (2,4)] -> Card(E^+) = 6

El conjunto de ejemplos negativos del concepto meta, Alcanzable, es:
E^- = [(2,1), (2,2), (3,1), (3,2), (3,3), (3,4), (4,1), (4,2), (4,3), (4,4)] -> Card(E^-) = 10

Por el enunciado suponemos que los 3 literales que producen mayor ganancia son:

{ Conectado(A,B), Conectado(A,C), Conectado(C,B) }

Si extendemos la regla utilizando el primer literal, L_1 = Conectado(A,B), es decir,

Alcanzable(C,D) :- Conectado(A,B)

Debido a que incumple la restricción del lenguaje de que al menos una de las variables entre el antecedente y el precedente debe ser común, esta regla no es válida, por lo que no se tiene en consideración.

Si extendemos la regla utilizando el segundo literal, L_2 = Conectado(A,C), es decir,

Alcanzable(C,D) :- Conectado(A,C)

Para calcular los ejemplos positivos de la nueva regla tenemos que ver una tupla donde A y C estén conectados y desde C se pueda alcanzar D.

Conectado(A, C): [(1,1), (1,2), (2,3), (2,4)]
Alcanzable(C, D): [(1,1), (1,2), (1,3), (1,4), (2,3), (2,4)]

E_L^+ tendrá la forma de una tupla (A,C,D)
E_L^+ = [(1,1,1), (1,1,2), (1,1,3), (1,1,4),(1,2,3), (1,2,4)] -> Card(E_L^+) = 6
Para calcular los ejemplos negativos de la nueva regla tenemos que ver una tupla donde A y C están conectados pero que desde C no pueda alcanzar D.

Conectado(A, C): [(1,1), (1,2), (2,3), (2,4)]
! Alcanzable(C, D) : [(2,1), (2,2), (3,1), (3,2), (3,3), (3,4), (4,1), (4,2), (4,3), (4,4)]

E_L^- tendrá la forma de una tupla (A,C,D)
E_L^- = [(1,2,1), (1,2,2),
(2,3,1), (2,3,2), (2,3,3), (2,3,4),
(2,4,1), (2,4,2), (2,4,3), (2,4,4)]
Card(E_L^-) = 10

$E_K = E^+$ => Card(E_K) = 6 (número de ejemplos positivos cubiertos por la regla)

$$G(L_2) = Card(E_k) \cdot [-\log_2(\frac{Card(E^+)}{Card(E^+) + Card(E^-)}) + \log_2(\frac{Card(E_L^+)}{Card(E_L^+) + Card(E_L^-)})]$$

G(L_2) = 6 · [- $\log_2$(6 / 6 + 10) + $\log_2$(6 / 6 + 10)] = 6 · 0 = **0**

Si extendemos la regla utilizando el tercer literal, L_3 = Conectado(C,B), es decir,

Alcanzable(C,D) :- Conectado(C,B)

Para calcular los ejemplos positivos de la nueva regla tenemos que ver una tupla donde C y B estén conectados y desde C se pueda alcanzar D.

Conectado(C, B): [(1,1), (1,2), (2,3), (2,4)]
Alcanzable(C, D): [(1,1), (1,2), (1,3), (1,4), (2,3), (2,4)]

E_L^+ tendrá la forma de una tupla (C,B,D)
E_L^+ = [(1,1,1), (1,1,2), (1,1,3), (1,1,4),
(1,2,1), (1,2,2), (1,2,3), (1,2,4),
(2,3,3), (2,3,4),
(2,4,3), (2,4,4)]
Card(E_L^+) = 12

Para calcular los ejemplos negativos de la nueva regla tenemos que ver una tupla donde C y B están conectados pero que desde C no pueda alcanzar D.

Conectado(C, B): [(1,1), (1,2), (2,3), (2,4)]
! Alcanzable(C, D) : [(2,1), (2,2), (3,1), (3,2), (3,3), (3,4), (4,1), (4,2), (4,3), (4,4)]

E_L^- tendrá la forma de una tupla (C,B,D)
E_L^- = [(2,3,1), (2,3,2),
(2,4,1), (2,4,2)]
Card(E_L^-) = 4

$E_K = E^+$ => Card(E_K) = 12 (número de ejemplos positivos cubiertos por la regla)

$$G(L_3) = Card(E_k) \cdot [-\log_2(\frac{Card(E^+)}{Card(E^+) + Card(E^-)}) + \log_2(\frac{Card(E_L^+)}{Card(E_L^+) + Card(E_L^-)})]$$

$G(L_3)$ = 12 · [- $\log_2$(6 / 6 + 10) + $\log_2$(12 / 12 + 4)] = 12 · (1.4150 - 0.4150) = **12**

Para añadir una regla a la base de reglas el conjunto de elementos negativos cubiertos por la regla debería ser cero, y ninguno de los 3 literales lo cumplen, por lo que habría que seguir iterando para analizar otros literales que ayuden a no cubrir los ejemplos negativos del predicado a aprender.

FIGURA 2

Predicado a aprender:
Alcanzable: [(1,1), (1,2), (1,3), (1,4), (2,1), (2,2), (2,3), (2,4), (3,1), (3,2), (3,3), (3,4), (4,1),(4,2), (4,3), (4,4)]
Dominio:
Conectado: [(1,2), (2,1), (2,3), (2,4), (3,2), (4,2)]

El conjunto de ejemplos positivos del concepto meta, Alcanzable, es:
E^+ = [(1,1), (1,2), (1,3), (1,4), (2,1), (2,2), (2,3), (2,4), (3,1), (3,2), (3,3), (3,4), (4,1), (4,2), (4,3), (4,4)]
Card(E^+) = 16

El conjunto de ejemplos negativos del concepto meta, Alcanzable, es:
E^- = [Ø]
Card(E^-) = 0

Por el enunciado suponemos que los 3 literales que producen mayor ganancia son:

{ Conectado(A,B), Conectado(A,C), Conectado(C,B) }

Si extendemos la regla utilizando el primer literal, L_1 = Conectado(A,B), tendríamos:

Alcanzable(C,D) :- Conectado(A,B)

Debido a que incumple la restricción del lenguaje de que al menos una de las variables entre el antecedente y el precedente debe ser común, esta regla no es válida, por lo que no se tiene en consideración.

Si extendemos la regla utilizando el segundo literal, L_2 = Conectado(A,C), es decir,

Alcanzable(C,D) :- Conectado(A,C)

Para calcular los ejemplos positivos de la nueva regla tenemos que ver una tupla donde A y C estén conectados y desde C se pueda alcanzar D.

Conectado(A, C): [(1,1), (1,2), (2,3), (2,4)]
Alcanzable(C, D): [(1,1), (1,2), (1,3), (1,4), (2,3), (2,4)]

E_L^+ tendrá la forma de una tupla (A,C,D)
E_L^+ = [(1,1,1,1), (1,1,1,2), (1,1,1,3), (1,1,1,4),
(1,2,2,3), (1,2,2,4)]
Card(E_L^+) = 6

Para calcular los ejemplos negativos de la nueva regla tenemos que ver una tupla donde A y C están conectados pero que desde C no pueda alcanzar D.

Conectado(A, C): [(1,1), (1,2), (2,3), (2,4)]
! Alcanzable(C, D) : [ ]

E_L^- tendrá la forma de una tupla (A,C,D)
E_L^- = [ ]
Card(E_L^-) = 0

$E_K = E^+$ => Card(E_K) = 16 (número de ejemplos positivos cubiertos por la regla)

G(L_2) = $Card(E_k) \cdot [-\log_2(\frac{Card(E^+)}{Card(E^+) + Card(E^-)}) + \log_2(\frac{Card(E_L^+)}{Card(E_L^+) + Card(E_L^-)})]$
G(L_2) = 16 · [- $\log_2$(16 / 16 + 0) + $\log_2$(6 / 6 + 0)] = 16 · 0 = **0**

Si extendemos la regla utilizando el tercer literal, L_3 = Conectado(C,B), es decir,

Alcanzable(C,D) :- Conectado(C,B)

Para calcular los ejemplos positivos de la nueva regla tenemos que ver una tupla donde C y B estén conectados y desde C se pueda alcanzar D.

Conectado(C, B): [(1,1), (1,2), (2,3), (2,4)]
Alcanzable(C, D): [(1,1), (1,2), (1,3), (1,4), (2,3), (2,4)]

E_L^+ tendrá la forma de una tupla (C,B,D)
E_L^+ = [(1,1,1), (1,1,2), (1,1,3), (1,1,4),
(1,2,1), (1,2,2), (1,2,3), (1,2,4),
(2,3,3), (2,3,4),
(2,4,3), (2,4,4)]
Card(E_L^+) = 12

Para calcular los ejemplos negativos de la nueva regla tenemos que ver una tupla donde C y B están conectados pero que desde C no pueda alcanzar D.

Conectado(C, B): [(1,1), (1,2), (2,3), (2,4)]
! Alcanzable(C, D) : [ ]

E_L^- tendrá la forma de una tupla (C,B,D)
E_L^- = [ ]
Card(E_L^-) = 0

E_K = E^+ => Card(E_K) = 16 (número de ejemplos positivos cubiertos por la regla)

G(L_3) = $Card(E_k)\cdot[-\log_2(\frac{Card(E^+)}{Card(E^+)+Card(E^-)})+\log_2(\frac{Card(E_L^+)}{Card(E_L^+)+Card(E_L^-)})]$
G(L_3) = 16 · [- $\log_2$(16 / 16 + 0) + $\log_2$(12 / 12 + 0)] = 16 · 0 = **0**

Para añadir una regla a la base de reglas el conjunto de elementos negativos cubiertos por la regla debería ser cero, cosa que ocurre con cualquiera de los 3 literales, por lo que no sería necesario seguir iterando.
Se eligiría entonces el de máxima ganancia, pero ésta es la misma para todos los literales. Se eligiría una de ellas aleatoriamente o según algún criterio (la que mayor número de ejemplos positivos cubriera), en este caso con el primer literal y la regla obtenida sería:

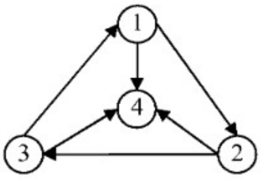
Alcanzable(C,D) :- Conectado(A,B)

Para aprender el concepto buscado debería cubrirse todos los ejemplos positivos con dicha regla, pero ninguna los cubren totalmente, por lo que habría que seguir buscando una o más reglas que terminen de cubrir todos los ejemplos positivos del predicado a aprender.

14SR, 15SR y parcialmente 16F2

PREGUNTA 3. Valoración: 5

Dado el grafo de la figura, en el que los arcos que conectan los nodos se describen mediante el predicado *conectado(A,B)*, se pretende utilizar el algoritmo FOIL para obtener la descripción, en forma de cláusulas de Horn, de cuándo un nodo *D* es alcanzable desde otro *C*, es decir, cuándo hay un camino que los conecta, y que será representado por el predicado *alcanzable(C,D)*.



Conectado = [(1,2), (2,3), (3,1), (1,4), (2,4), (3,4)]

Alcanzable = [(1,1), (1,2), (1,3), (1,4), (2,1), (2,2), (2,3), (2,4),
(3,1), (3,2), (3,3), (3,4)]

Contestar a las siguientes preguntas:

1. Si quisiéramos construir una regla con un único literal como antecedente (el consecuente sería justo el concepto a aprender: *Alcanzable(A,B)*, justifique cuáles serían todos los posibles literales a añadir a dicha regla.

2. Para reducir el tiempo de cálculo que el alumno debería emplear en calcular la ganancia de todos los posibles literales obtenidos en el apartado anterior, supóngase que los literales que producen las tres mayores ganancias son:

   {*Conectado(A, B), Conectado(A, C), Conectado(C, B)*}

   Se pide, para cada literal, obtener:
   (a) Los conjuntos $\mathbf{E_0^+}$ y $\mathbf{E_0^-}$.
   (b) Los conjuntos $\mathbf{E_0^{++}}$, $\mathbf{E_1^+}$ y $\mathbf{E_1^-}$ como resultado de añadir, en cada caso, cada uno de los literales indicados.
   (c) La ganancia asociada al literal.

3. En vista de los resultados obtenidos en el apartado anterior, ¿sería necesario seguir iterando el algoritmo para aprender el concepto? Justifique la respuesta. Si la respuesta dada fuese un *sí*, describa cualitativamente cuáles serían los pasos a seguir.

4. Describa las diferencias entre los algoritmos puramente inductivos y este tipo de algoritmos (FOIL) encuadrados dentro del paradigma de la Programación Lógica Inductiva. Además, respecto a este último, especifique cuáles son las *bias* y el comportamiento ante el ruido en las entradas y en la estructura aprendida.

1) Todos los posibles literales a añadir a dicha regla serían:

Se cumple un determinado literal, Conectado(x,y) o Alcanzado(x,y) donde al menos una de las variables ya existía:

Conectado(A,B), Conectado(A,C), Conectado(B,A), Conectado(C,A), Conectado(B,C), Conectado(C,B),

Alcanzable(A,C), Alcanzable(B,A), Alcanzable(C,A), Alcanzable(B,C), Alcanzable(C,B)

La negación de los anteriores (no se cumple un determinado literal):
!Conectado(A,B), !Conectado(A,C), !Conectado(B,A), !Conectado(C,A), !Conectado(B,C), !Conectado(C,B),
!Alcanzable(A,C), !Alcanzable(B,A), !Alcanzable(C,A), !Alcanzable(B,C), !Alcanzable(C,B)

Una variable es igual o diferente a otra:
$A = B$, $A \neq B$

Una varibable es mayor que o menor e igual que otra:
$A > B$, $A \leq B$, $B > A$, $B \leq A$,

También podría darse que una variable sea igual, diferente, mayor que o menor e igual que una constante especificada por FOIL, en este caso no se indica nada en el enunciado, por lo que esta condición no produciría ningún literal.

2)
Predicado a aprender:
Alcanzable: [(1,1), (1,2), (1,3), (1,4), (2,1), (2,2), (2,3), (2,4), (3,1), (3,2), (3,3), (3,4)]

Dominio:
Conectado: [(1,2), (1,4), (2,3), (2,4), (3,1), (3,4)]

El conjunto de ejemplos positivos del concepto meta, Alcanzable, es:
E^+ = [(1,1), (1,2), (1,3), (1,4), (2,1), (2,2), (2,3), (2,4), (3,1), (3,2), (3,3), (3,4)]
Card(E^+) = 12

El conjunto de ejemplos negativos del concepto meta, Alcanzable, es:
E^- = [(4,1), (4,2), (4,3), (4,4)]
Card(E^-) = 4

Por el enunciado suponemos que los 3 literales que producen mayor ganancia son:

{ Conectado(A,B), Conectado(A,C), Conectado(C,B) }

Si extendemos la regla utilizando el primer literal, L_1 = Conectado(A,B) es decir,

Alcanzable(A,B) :- Conectado(A,B)

Para calcular los ejemplos positivos de la nueva regla tenemos que ver una tupla donde A y B estén conectados y desde A se pueda alcanzar B.

Conectado(A, B): [(1,2), (1,4), (2,3), (2,4), (3,1), (3,4)]
Alcanzable(A, B): [(1,1), (1,2), (1,3), (1,4), (2,1), (2,2), (2,3), (2,4), (3,1), (3,2), (3,3), (3,4)]

E_L^+ tendrá la forma de una tupla (A,B)
E_L^+ = [(1,2), (1,4), (2,3), (2,4), (3,1), (3,4)]
Card(E_L^+) = 6

Para calcular los ejemplos negativos de la nueva regla tenemos que ver una tupla donde A y B están conectados pero que desde A no pueda alcanzar B.

Conectado(A, B): [(1,2), (1,4), (2,3), (2,4), (3,1), (3,4)]
! Alcanzable(A, B) : [(4,1), (4,2), (4,3), (4,4)]

E_L^- tendrá la forma de una tupla (A,B,C)
E_L^- = [ ]
Card(E_L^-) = 0

$E_K = E_L^+$ => Card(E_K) = 6 (número de ejemplos positivos cubiertos por la regla)

$$G(L_2)= \ Card(E_k)\cdot[-\log_2(\frac{Card(E^+)}{Card(E^+)+Card(E^-)})+\log_2(\frac{Card(E_L^+)}{Card(E_L^+)+Card(E_L^-)})]$$

G(L_2) = 6 · [- $\log_2$(12 / 12 + 4) + $\log_2$(6 / 6 + 0)] = 6 · (0.4150 - 0) = **2.49**

Si extendemos la regla utilizando el segundo literal, L_2 = Conectado(A,C), es decir,

Alcanzable(A,B) :- Conectado(A,C)

Para calcular los ejemplos positivos de la nueva regla tenemos que ver una tupla donde A y C estén conectados y desde A se pueda alcanzar B.

Conectado(A, C): [(1,2), (1,4), (2,3), (2,4), (3,1), (3,4)]
Alcanzable(A, B): [(1,1), (1,2), (1,3), (1,4), (2,1), (2,2), (2,3), (2,4), (3,1), (3,2), (3,3), (3,4)]

E_L^+ tendrá la forma de una tupla (A,B,C)
E_L^+ = [(1,2,1), (1,2,2), (1,2,3), (1,2,4),
(1,4,1), (1,4,2), (1,4,3), (1,4,4),
(2,3,1), (2,3,2), (2,3,3), (2,3,4),
(2,4,1), (2,4,2), (2,4,3), (2,4,4),
(3,1,1), (3,1,2), (3,1,3), (3,1,4),
(3,4,1), (3,4,2), (3,4,3), (3,4,4)]
Card(E_L^+) = 24

Para calcular los ejemplos negativos de la nueva regla tenemos que ver una tupla donde A y C están conectados pero que desde A no pueda alcanzar B.

Conectado(A, C): [(1,2), (1,4), (2,3), (2,4), (3,1), (3,4)]
! Alcanzable(A, B) : [(4,1), (4,2), (4,3), (4,4)]

E_L^- tendrá la forma de una tupla (A,B,C)
E_L^- = [ ]
Card(E_L^-) = 0

Card(E_K) = 12 (número de ejemplos positivos cubiertos por la regla)

$$G(L_2)= \ Card(E_k)\cdot[-\log_2(\frac{Card(E^+)}{Card(E^+)+Card(E^-)})+\log_2(\frac{Card(E_L^+)}{Card(E_L^+)+Card(E_L^-)})]$$

G(L_2) = 12 · [- $\log_2$(12 / 12 + 4) + $\log_2$(24 / 24 + 0)] = 12 · (0.4150 - 0) = **4.98**

Si extendemos la regla utilizando el tercer literal, L_3 = Conectado(C,B), es decir, Alcanzable(A,B) :- Conectado(C,B)

Para calcular los ejemplos positivos de la nueva regla tenemos que ver una tupla donde C y B estén conectados y desde A se pueda alcanzar B.

Conectado(C, B): [(1,2), (1,4), (2,3), (2,4), (3,1), (3,4)]
Alcanzable(A, B): [(1,1), (1,2), (1,3), (1,4), (2,1), (2,2), (2,3), (2,4), (3,1), (3,2), (3,3), (3,4)]

E_L^+ tendrá la forma de una tupla (A,B,C)
E_L^+ = [(1,1,3), (1,2,1), (1,3,2), (1,4,1), (1,4,2), (1,4,3), (2,1,3), (2,2,1), (2,3,2), (2,4,1), (2,4,2), (2,4,3), (3,1,3), (3,2,1), (3,3,2), (3,4,1), (3,4,2), (3,4,3)]
Card(E_L^+) = 18

Para calcular los ejemplos negativos de la nueva regla tenemos que ver una tupla donde A y B están conectados pero que desde C no pueda alcanzar D.

Conectado(C, B): [(1,2), (1,4), (2,3), (2,4), (3,1), (3,4)]
! Alcanzable(A, B) : [(4,1), (4,2), (4,3), (4,4)]

E_L^- tendrá la forma de una tupla (A,B,C)
E_L^- = [(1,2,4), (1,4,4), (2,3,4), (2,4,4), (3,1,4), (3,4,4)]
Card(E_L^-) = 6

Card(E_K) = 12 (número de ejemplos positivos cubiertos por la regla)

$$G(L_3) = Card(E_k) \cdot [-\log_2(\frac{Card(E^+)}{Card(E^+) + Card(E^-)}) + \log_2(\frac{Card(E_L^+)}{Card(E_L^+) + Card(E_L^-)})]$$

$G(L_3) = 12 \cdot [- \log_2(12 / 12 + 4) + \log_2(18 / 18 + 6)] = 12 \cdot (0.4150 - 0.4150) = \mathbf{0}$

El literal que mayor ganancia proporciona es Alcanzable(A,B) :- Conectado(A,C).
Con este literal no se cubre ningún ejemplo negativo, por lo que podemos confirmar que la regla está completa. Por otro lado, cubre todos los ejemplos positivos, por lo que no es necesaria ninguna otra regla adicional.

15SO

PREGUNTA 4. Valoración 3

Dado el grafo de la figura 1 en el que los arcos que conectan los nodos se describen mediante el predicado conectado(X,Y), utilizar el **algoritmo FOIL** para obtener la descripción, en forma de cláusulas de Horn, del concepto "cuándo un nodo B es alcanzable desde otro A", es decir, cuándo hay un camino que los conecta. Dicho concepto será representado por el predicado alcanzable(A,B). Para simplificar el problema, de todos los posibles literales, supóngase que los literales que producen las tres mayores ganancias son:

{Conectado(A, B), Conectado(A, C), Conectado(C, B)}

Se pide, para cada literal, obtener:

(a) Los conjuntos E0+ y E0-.
(b) Los conjuntos E0++, E1+ y E1- como resultado de añadir, en cada caso, cada uno de los literales indicados.
(c) La ganancia asociada al literal.

En vista de los resultados obtenidos en el apartado anterior, ¿sería necesario seguir iterando el algoritmo para aprender el concepto? Justifique la respuesta. Si la respuesta dada fuese un sí, describa cualitativamente cuáles serían los pasos a seguir.

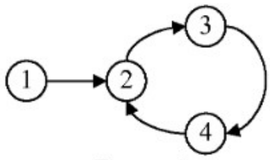


Figura 1

Predicado a aprender:
Alcanzable: [(1,2), (1,3), (1,4), (2,2), (2,3), (2,4), (3,2), (3,3), (3,4), (4,2), (4,3), (4,4)]

Dominio:
Conectado: [(1,2), (2,3), (3,4), (4,2)]

El conjunto de ejemplos positivos del concepto meta, Alcanzable, es:
E^+ = [(1,2), (1,3), (1,4), (2,2), (2,3), (2,4), (3,2), (3,3), (3,4), (4,2), (4,3), (4,4)]
Card(E^+) = 12

El conjunto de ejemplos negativos del concepto meta, Alcanzable, es:
E^- = [(1,1), (2,1), (3,1), (4,1)]
Card(E^-) = 4

Por el enunciado suponemos que los 3 literales que producen mayor ganancia son:

{ Conectado(A,B), Conectado(A,C), Conectado(C,B) }

Si extendemos la regla utilizando el primer literal, L_1 = Conectado(A,B) es decir,

Alcanzable(A,B) :- Conectado(A,B)

Para calcular los ejemplos positivos de la nueva regla tenemos que ver una tupla donde A y B estén conectados y desde A se pueda alcanzar B.

Conectado(A, B): [(1,2), (2,3), (3,4), (4,2)]
Alcanzable(A,B): [(1,2), (1,3), (1,4), (2,2), (2,3), (2,4), (3,2), (3,3), (3,4), (4,2), (4,3), (4,4)]

E_L^+ tendrá la forma de una tupla (A,B)
E_L^+ = [(1,2), (2,3), (3,4), (4,2)]
Card(E_L^+) = 4

Para calcular los ejemplos negativos de la nueva regla tenemos que ver una tupla donde A y B están conectados pero que desde A no pueda alcanzar B.

Conectado(A, B): [(1,2), (2,3), (3,4), (4,2)]
! Alcanzable(A, B) : [(1,1), (2,1), (3,1), (4,1)]

E_L^- tendrá la forma de una tupla (A,B)
E_L^- = [ ]
Card(E_L^-) = 0

$E_K = E^+$ => Card(E_K) = 12 (número de ejemplos positivos cubiertos por la regla)

G(L_1) = $Card(E_k)\cdot[-\log_2(\frac{Card(E^+)}{Card(E^+)+Card(E^-)})+\log_2(\frac{Card(E_L^+)}{Card(E_L^+)+Card(E_L^-)})]$
G(L_1) = 12 · [- $\log_2$(12 / 12 + 4) + $\log_2$(4 / 4 + 0)] = 12 · (0.4150 - 0) = **4.9804**

Si extendemos la regla utilizando el segundo literal, L_2 = Conectado(A,C), es decir,

Alcanzable(A,B) :- Conectado(A,C)

Para calcular los ejemplos positivos de la nueva regla tenemos que ver una tupla donde A y B estén conectados y desde C se pueda alcanzar D.

Conectado(A,C): [(1,2), (2,3), (3,4), (4,2)]
Alcanzable(A,B): [(1,2), (1,3), (1,4), (2,2), (2,3), (2,4), (3,2), (3,3), (3,4), (4,2), (4,3), (4,4)]

E_L^+ tendrá la forma de una tupla (A,C,B)
E_L^+ = [(1,2,2), (1,2,3), (1,2,4),
(2,3,2), (2,3,3), (2,3,4),
(3,4,2), (3,4,3), (3,4,4)]
Card(E_L^+) = 12

Para calcular los ejemplos negativos de la nueva regla tenemos que ver una tupla donde A y C están conectados pero que desde A no pueda alcanzar B.

Conectado(A,C): [(1,2), (2,3), (3,4), (4,2)]
! Alcanzable(A,B) : [(1,1), (2,1), (3,1), (4,1)]

E_L^- tendrá la forma de una tupla (A,C,B)
E_L^- = [(1,2,1),
(2,3,1),
(3,4,1),
(4,2,1)]
Card(E_L^-) = 4

$E_K = E^+ \Rightarrow Card(E_K) = 12$ (número de ejemplos positivos cubiertos por la regla)

G(L$_2$) = $Card(E_k) \cdot [-\log_2(\frac{Card(E^+)}{Card(E^+) + Card(E^-)}) + \log_2(\frac{Card(E_L^+)}{Card(E_L^+) + Card(E_L^-)})]$

G(L$_2$) = 12 · [- log$_2$(12 / 12 + 4) + log$_2$(12 / 12 + 4)] = 12 · (0.4150 – 0.4150) = **0**

Si extendemos la regla utilizando el tercer literal, L$_3$ = Conectado(C,B), es decir,

Alcanzable(A,B) :- Conectado(C,B)

Para calcular los ejemplos positivos de la nueva regla tenemos que ver una tupla donde A y B estén conectados y desde C se pueda alcanzar B.

Conectado(C, B): [(1,2), (2,3), (3,4), (4,2)]
Alcanzable(A, B): [(1,2), (1,3), (1,4), (2,2), (2,3), (2,4), (3,2), (3,3), (3,4), (4,2), (4,3), (4,4)]

E_L^+ tendrá la forma de una tupla (C,A,B)
E_L^+ = [(1,1,2), (1,2,2), (1,3,2), (1,4,2),
(2,1,3), (2,2,3), (2,3,3), (2,4,3),
(3,1,4), (3,2,4), (3,3,4), (3,4,4),
(4,1,2), (4,2,2), (4,3,2), (4,4,2)]
Card(E_L^+) = 16

Para calcular los ejemplos negativos de la nueva regla tenemos que ver una tupla donde C y B están conectados pero que desde A no pueda alcanzar B.

Conectado(C, B): [(1,2), (2,3), (3,4), (4,2)]
! Alcanzable(A, B) : [(1,1), (2,1), (3,1), (4,1)]

E_L^- tendrá la forma de una tupla (C,A,B)
E_L^- = [ ]
Card(E_L^-) = 0

$E_K = E^+ \Rightarrow Card(E_K) = 12$ (número de ejemplos positivos cubiertos por la regla)

G(L$_3$) = $Card(E_k) \cdot [-\log_2(\frac{Card(E^+)}{Card(E^+) + Card(E^-)}) + \log_2(\frac{Card(E_L^+)}{Card(E_L^+) + Card(E_L^-)})]$

G(L$_3$) = 12 · [- log$_2$(12 / 12 + 4) + log$_2$(16 / 16 + 0)] = 12 · (0.5305 - 0) = **4.9804**

El literal que mayor ganancia ofrece es el número 3, por ello la regla que formamos sería Alcanzable(A,B) :- Conectado(C,B). Con este literal no se cubre ningún ejemplo negativo, por lo que podemos confirmar que la regla está completa. Por otro lado, como no cubre todos los ejemplos positivos (solo cubre 4 de los 12 del concepto meta), sería necesaria una (o varias) regla(s) adicional(es), por lo que el algoritmo necesitaría hacer nuevas iteraciones descartando los ejemplos cubiertos por la regla formada a partir del literal anterior.

16SO

PREGUNTA 4. Valoración: 2

Dado el grafo de la figura 1 en el que los arcos que conectan los nodos se describen mediante el predicado *conectado(X,Y)*, utilizar el **algoritmo FOIL** para obtener la descripción, en forma de cláusulas de Horn, del concepto "cuándo un nodo *B* es alcanzable desde otro *A*", es decir, cuándo hay un camino que los conecta. Dicho concepto será representado por el predicado *alcanzable(A,B)*. Para simplificar el problema, de todos los posibles literales, supóngase que los literales que producen las tres mayores ganancias son:

{*Conectado(A, B), Conectado(A, C), Conectado(C, B)*}

Se pide, para cada literal, obtener:

(a) Los conjuntos $\mathbf{E^+}$ y $\mathbf{E^-}$.

(b) Los conjuntos $\mathbf{E_k}$, $\mathbf{E_L^+}$ y $\mathbf{E_L^-}$ como resultado de añadir, en cada caso, cada uno de los literales indicados.

(c) La ganancia asociada al literal.

¿Con el análisis realizado, se puede añadir ya alguna regla definitiva a la base de reglas? Si la respuesta a la anterior pregunta fue afirmativa, ¿la regla obtenida aprende el concepto buscado o sería necesario seguir iterando? Justifique su respuesta.

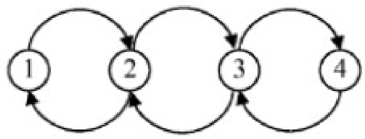


Figura 1

Disponemos como información del dominio los nodos conectados y los ejemplos positivos del concepto "Alcanzable que se quiere aprender":

Conectados: [(1,2), (2,1), (2,3), (3,2), (3,4), (4,3)]
Alcanzables: [(1,1), (1,2), (1,3), (1,4),
(2,1), (2,2), (2,3), (2,4),
(3,1), (3,2), (3,3), (3,4),
(4,1), (4,2), (4,3), (4,4)]

Por tanto, todos los conjuntos de ejemplos positivos (E^+) y negativos (E^-) serán los siguientes:

E^+ = Alcanzables   Card(E^+) = 16
E^- = [Ø]   Card(E^-) = 0

Como E^- está vacío, nunca se llega a entrar al bucle interior donde se evalúa cada literal, por lo que se creará laregla formada sólo por el consecuente:

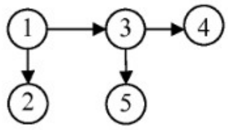
Alcanzable(A,B)

y que se cumplirá siempre. De hecho, de una forma sencilla, podemos comprobar que todos los nodos son alcanzables desde cualquier nodo, incluso desde el propio nodo.

17F1

PREGUNTA 3. Valoración: 2.5

Se utiliza el **algoritmo FOIL** para el aprendizaje del concepto *ALCANZABLE(A, B)*. El conjunto de ejemplos utilizado corresponde al grafo de la figura:



Conectado = [(1,2), (1,3), (3,4), (3,5)]

Alcanzable = [(1,2), (1,3), (1,4), (1,5), (3,4), (3,5)]

Contestar a las siguientes preguntas:

1. Suponiendo que los literales que producen las tres mayores ganancias son:

   $\{Conectado(A, B), Conectado(A, C), Conectado(C, B)\}$

   Se pide, para cada literal, obtener:
   (a) Los conjuntos $\mathbf{E^+}$ y $\mathbf{E^-}$, $\mathbf{E_k}$, $\mathbf{E_L}^+$ y $\mathbf{E_L}^+$.
   (b) La ganancia asociada.

2. En vista de los resultados obtenidos en el apartado anterior, ¿sería necesario seguir iterando el algoritmo para aprender el concepto? Justifique la respuesta. Si la respuesta fuese afirmativa, describa cualitativamente cuáles serían los pasos a seguir.

Predicado a aprender:
Alcanzable: [(1,2), (1,3), (1,4), (1,5), (3,4), (3,5)]

Dominio:
Conectado: [(1,2), (1,3), (3,4), (3,5)]

El conjunto de ejemplos positivos del concepto meta, Alcanzable, es:
E^+ = [(1,2), (1,3), (1,4), (1,5), (3,4), (3,5)]
Card(E^+) = 6

El conjunto de ejemplos negativos del concepto meta, Alcanzable, es:
E^- = [(1,1), (2,1), (2,2), (2,3), (2,4), (2,5), (3,1), (3,2), (3,3), (4,1), (4,2), (4,3), (4,4), (4,5), (5,1), (5,2), (5,3), (5,4), (5,5)]
Card(E^-) = 19

Por el enunciado suponemos que los 3 literales que producen mayor ganancia son:

{ Conectado(A,B), Conectado(A,C), Conectado(C,B) }

Si extendemos la regla utilizando el primer literal, L_1 = Conectado(A,B), es decir,

Alcanzable(A,B) :- Conectado(A,B)

Para calcular los ejemplos positivos de la nueva regla tenemos que ver una tupla donde A y B estén conectados y desde A se pueda alcanzar B.

Conectado(A,B): [(1,2), (1,3), (3,4), (3,5)]
Alcanzable(A,B): [(1,2), (1,3), (1,4), (1,5), (3,4), (3,5)]

E_L^+ tendrá la forma de una tupla (A,B)
E_L^+ = [(1,2), (1,3), (3,4), (3,5)]
Card(E_L^+) = 4

Para calcular los ejemplos negativos de la nueva regla tenemos que ver una tupla donde A y B están conectados pero que desde A no se pueda alcanzar B.
Conectado(A, B): [(1,2), (1,3), (3,4), (3,5)]
! Alcanzable(A,B) : [(1,1), (2,1), (2,2), (2,3), (2,4), (2,5), (3,1), (3,2), (3,3), (4,1), (4,2), (4,3), (4,4), (4,5), (5,1), (5,2), (5,3), (5,4), (5,5)]
E_L^- tendrá la forma de una tupla (A,B)
E_L^- = [ ]
Card(E_L^-) = 0

E_K = [(1,2), (1,3), (3,4), (3,5)] => Card(E_K) = 6 (número de ejemplos positivos cubiertos por la regla)

$$G(L_1) = \; Card(E_k) \cdot [-\log_2(\frac{Card(E^+)}{Card(E^+) + Card(E^-)}) + \log_2(\frac{Card(E_L^+)}{Card(E_L^+) + Card(E_L^-)})]$$
G(L_1) = 4 · [- $\log_2$(6 / 6 + 19) + $\log_2$(4 / 4 + 0)] = 4 · (2.0588) = **8.2352**

Si extendemos la regla utilizando el segundo literal, L_2 = Conectado(A,C), es decir,

Alcanzable(A,B) :- Conectado(A,C)

Para calcular los ejemplos positivos de la nueva regla tenemos que ver una tupla donde A y C estén conectados y desde A se pueda alcanzar B.

Conectado(A, C): [(1,2), (1,3), (3,4), (3,5)]
Alcanzable(A,B): [(1,2), (1,3), (1,4), (1,5), (3,4), (3,5)]

E_L^+ tendrá la forma de una tupla (A,B,C)
E_L^+ = [(1,2,2), (1,3,2), (1,4,2), (1,5,2),
(1,2,3), (1,3,3), (1,4,3), (1,5,3),
(3,4,4), (3,5,4),
(3,4,5), (3,5,5)]
Card(E_L^+) = 12

Para calcular los ejemplos negativos de la nueva regla tenemos que ver una tupla donde A y C están conectados pero que desde A no pueda alcanzar B.

Conectado(A,C): [(1,2), (1,3), (3,4), (3,5)]
! Alcanzable(A,B) : [(1,1), (2,1), (2,2), (2,3), (2,4), (2,5), (3,1), (3,2), (3,3), (4,1), (4,2), (4,3), (4,4), (4,5), (5,1), (5,2), (5,3), (5,4), (5,5)]

E_L^+ tendrá la forma de una tupla (A,B,C)
E_L^- = [(1,1,2),
(1,1,3),
(3,1,4), (3,2,4), (3,3,4),
(3,1,5), (3,2,5), (3,3,5)]
Card(E_L^-) = 8

E_K = [(1,2), (1,3), (1,4), (1,5), (3,4), (3,5)] => Card(E_K) = 6 (número de ejemplos positivos cubiertos por la regla)

G(L_2) = $Card(E_k) \cdot [-\log_2(\frac{Card(E^+)}{Card(E^+) + Card(E^-)}) + \log_2(\frac{Card(E_k^+)}{Card(E_k^+) + Card(E_k^-)})]$
G(L_2) = 6 · [- $\log_2$(6 / 6 + 19) + $\log_2$(12 / 12 + 8)] = 6 · (2.0588 - 0.7369) = **7.9317**

Si extendemos la regla utilizando el tercer literal, L_3 = Conectado(C,B), es decir,

Alcanzable(A,B) :- Conectado(C,B)

Para calcular los ejemplos positivos de la nueva regla tenemos que ver una tupla donde C y B estén conectados y desde A se pueda alcanzar B.

Conectado(C, B): [(1,2), (1,3), (3,4), (3,5)]
Alcanzable(A,B): [(1,2), (1,3), (1,4), (1,5), (3,4), (3,5)]

E_L^+ tendrá la forma de una tupla (C,A,B)
E_L^+ = [(1,1,2),
(1,1,3),
(3,1,4), (3,3,4),
(3,1,5), (3,3,5)]
Card(E_L^+) = 6

Para calcular los ejemplos negativos de la nueva regla tenemos que ver una tupla donde C y B están conectados pero que desde A no pueda alcanzar B.

Conectado(C, B): [(1,2), (1,3), (3,4), (3,5)]
! Alcanzable(A, B) : [(1,1), (2,1), (2,2), (2,3), (2,4), (2,5), (3,1), (3,2), (3,3), (4,1), (4,2), (4,3), (4,4), (4,5), (5,1), (5,2), (5,3), (5,4), (5,5)]

E_L^- tendrá la forma de una tupla (C,A,B)
E_L^- = [(1,2,2), (1,3,2), (1,4,2), (1,5,2),
(1,2,3), (1,3,3), (1,4,3), (1,5,3),
(3,2,4), (3,4,4), (3,5,4),
(3,2,5), (3,4,5), (3,5,5)]
Card(E_L^-) = 14

$E_K = E^+ \Rightarrow Card(E_K) = 6$ (número de ejemplos positivos cubiertos por la regla)

$$G(L_3) = Card(E_k) \cdot [-\log_2(\frac{Card(E^+)}{Card(E^+) + Card(E^-)}) + \log_2(\frac{Card(E_L^+)}{Card(E_L^+) + Card(E_L^-)})]$$

$G(L_3) = 6 \cdot [- \log_2(6 / 6 + 19) + \log_2(6 / 6 + 14)] = 6 \cdot (2.0588 - 1.7369) =$ **1.9314**

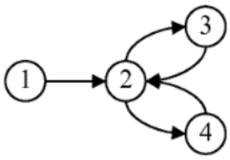
El literal elegido para comenzar a formar la nueva regla sería la L_1, la de mayor ganancia.

Una regla aprende el concepto si cubre todos los ejemplos positivos del concepto y no cubre ninguno negativo. El literal L_1 no cubre ejemplos negativos por lo que no es necesario seguir iterando y podemos añadir dicha regla a la base de reglas. Pero le faltan 2 ejemplos positivos por cubrir (1,4) y (1,5), por lo que debería buscarse otra (u otras) reglas para aprender el concepto.

17SR

PREGUNTA 3. Valoración: 4

Dado el grafo de la figura, en el que los arcos que conectan los nodos se describen mediante el predicado *conectado(A,B)*, se pretende utilizar el **algoritmo FOIL** para obtener la descripción, en forma de cláusulas de Horn, de cuándo un nodo *D* es alcanzable desde otro *C*, es decir, cuándo hay un camino que los conecta, y que será representado por el predicado *alcanzable(C,D)*.



Conectado = [(1,2), (2,3), (2,4), (3,2), (4,2)]

Alcanzable = [(1,2), (1,3), (1,4), (2,2), (2,3), (2,4), (3,2), (3,3), (3,4), (4,2), (4,3), (4,4)]

Contestar a las siguientes preguntas:

1. Si quisiéramos construir una regla con un único literal como antecedente (el consecuente sería justo el concepto a aprender: *Alcanzable(A,B)*, justifique cuáles serían todos los posibles literales a añadir a dicha regla.

2. Para reducir el tiempo de cálculo que el alumno debería emplear en calcular la ganancia de todos los posibles literales obtenidos en el apartado anterior, supóngase que los literales que producen las tres mayores ganancias son:

   {*Conectado(A, B), Conectado(A, C), Conectado(C, B)*}

   Se pide, para cada literal, obtener:
   (a) Los conjuntos $\mathbf{E^+}$ y $\mathbf{E^-}$.
   (b) Los conjuntos $\mathbf{E_k}$, $\mathbf{E_L^+}$ y $\mathbf{E_L^-}$ como resultado de añadir, en cada caso, cada uno de los literales indicados.
   (c) La ganancia asociada al literal.

3. En vista de los resultados obtenidos en el apartado anterior, ¿sería necesario seguir iterando el algoritmo para aprender el concepto? Justifique la respuesta. Si la respuesta dada fuese un *sí*, describa cualitativamente cuáles serían los pasos a seguir.

4. Describa las diferencias entre los algoritmos puramente inductivos y este tipo de algoritmos (FOIL) encuadrados dentro del paradigma de la Programación Lógica Inductiva. Además, respecto a este último, especifique cuáles son las *bias* y el comportamiento ante el ruido en las entradas y en la estructura aprendida.

Disponemos de la información del dominio de los conectados:

Conectado: [(1,2), (2,3), (2,4), (3,2), (4,2)]

Podemos obtener los conjuntos de ejemplos positivos (E^+) y negativos (E^-) del concepto "Alcanzable" a aprender:

E^+ = [(1,2), (1,3), (1,4), (2,2), (2,3), (2,4), (3,2), (3,3), (3,4), (4,2), (4,3), (4,4)]
Card(E^+) = 12

E^- = [(1,1), (2,1), (3,1), (4,1)]
Card(E^-) = 4

Para el Literal L_1 : Alcanzable(A,B) :- Conectado(A,B)

E_L^+ = [(1,2), (2,3), (2,4), (3,4), (4,2)]
Card(E_L^+) = 5

E_L^- = [Ø]
Card(E_L^-) = 0

$E_k = E_L^+$ => Card(E_k) = 5

G(L_1) = $Card(E_k) \cdot [-\log_2(\frac{Card(E^+)}{Card(E^+) + Card(E^-)}) + \log_2(\frac{Card(E_L^+)}{Card(E_L^+) + Card(E_L^-)})]$

G(L_1) = 5 · [- $\log_2$(12 / 12 + 4) + $\log_2$(5 / 5 + 0)] = **2.07**

Para el Literal L_2 : Alcanzable(A,B) :- Conectado(A,C)

Con tuplas de la forma (A,B,C):

E_L^+ = [(1,2,2), (1,2,3), (1,2,4),
(2,3,2), (2,3,3), (2,3,4),
(2,4,2), (2,4,3), (2,4,4),
(3,4,2), (3,4,3), (3,4,4),
(4,2,2), (4,2,3), (4,2,4)]
Card(E_L^+) = 15

E_L^- = [(1,2,1), (2,3,1), (2,4,1), (3,4,1), (4,2,1)]
Card(E_L^-) = 5

$E_k = E^+$ => Card(E_k) = 12

G(L_2) = $Card(E_k) \cdot [-\log_2(\frac{Card(E^+)}{Card(E^+) + Card(E^-)}) + \log_2(\frac{Card(E_L^+)}{Card(E_L^+) + Card(E_L^-)})]$

G(L_2) = 12 · [- $\log_2$(12 / 12 + 4) + $\log_2$(15 / 15 + 5)] = **0**

Para el Literal L_3 : Alcanzable(A,B) :- Conectado(C,B)

Con tuplas de la forma (A,C,B):

E_L^+ = [(1,1,2), (2,1,2), (3,1,2), (4,1,2),
(1,2,3), (2,2,3), (3,2,3), (4,2,3),
(1,2,4), (2,2,4), (3,2,4), (4,2,4),
(1,3,4), (2,3,4), (3,3,4), (4,3,4),
(1,4,2), (2,4,2), (3,4,2), (4,4,2)]
Card(E_L^+) = 20

E_L^- = [Ø]
Card(E_L^-) = 0

$E_k = E^+$ => Card(E_k) = 12

$$G(L_3) = Card(E_k) \cdot [-\log_2(\frac{Card(E^+)}{Card(E^+) + Card(E^-)}) + \log_2(\frac{Card(E_L^+)}{Card(E_L^+) + Card(E_L^-)})]$$

G(L_3) = 12 · [- $\log_2$(12 / 12 + 4) + $\log_2$(20 / 20 + 0)] = **4.98**

El literal de mayor ganancia sería L_3 , por tanto la regla aprendida es:

Alcanzable(A,B) :- Conectado(C,B)

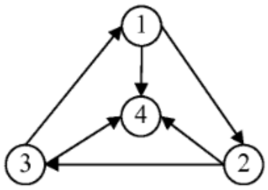
Dado que con esta regla no quedan ejemplos negativos, la regla está completa y no es necesario añadir otro literal a la misma.

Por otro lado, como la regla cubre todos los ejemplos positivos, no es necesario añadir una regla nueva, así que el algoritmo finalizaría.

18SR

PREGUNTA 3. Valoración: 5

Dado el grafo de la figura, en el que los arcos que conectan los nodos se describen mediante el predicado *conectado(A,B)*, se pretende utilizar el **algoritmo FOIL** para obtener la descripción, en forma de cláusulas de Horn, de cuándo un nodo *D* es alcanzable desde otro *C*, es decir, cuándo hay un camino que los conecta, y que será representado por el predicado *alcanzable(C,D)*.



Conectado = [(1,2), (2,3), (3,1), (1,4), (2,4), (3,4)]

Alcanzable = [(1,1), (1,2), (1,3), (1,4), (2,1), (2,2), (2,3), (2,4),

. . .

4. Describa las diferencias entre los algoritmos puramente inductivos y este tipo de algoritmos (FOIL) encuadrados dentro del paradigma de la Programación Lógica Inductiva. Además, respecto a este último, especifique cuáles son las *bias* y el comportamiento ante el ruido en las entradas y en la estructura aprendida.

Apartados 1, 2 y 3 hechos en **14SR** y **15SR**

4)

Los algoritmos inductivos reciben como entrada ejemplos expresados en lógica proposicional (valores asignados a atributos), en el caso de los ILP utilizan lógica de predicados (más expresiva) tanto en las entradas como en las salidas. .
Los ILP además de ejemplos, también pueden tomar como entradas teorías del dominio expresadas en términos de reglas lógicas. .
Cubre las deficiencias de los métodos inductivos (no utiliza conocimiento a priori del dominio) y deductivos (no generalizan más allá del conocimiento del dominio y los ejemplos) utilizando ideas de ambos.

Para la Programación lógica tenemos las siguientes características:

- Bias: Las heurísticas de búsquedas están descritas por la función de ganancia, se prefieren aquellos literales que se paren mejor los ejemplos positivos de los negativos.
  En cuanto a bías del tipo de restricciones del lenguaje utiliza bastantes como no permitir funciones ni constantes como términos en los literales, o que los literales deben tener al menos una variable utilizada anteriormente. También utiliza la suposición del mundo cerrado para generar los ejemplos negativos.
- Comportamiento ante el ruido: **En entradas**, las últimas versiones de FOIL admiten ruido en las entradas, con ejemplos etiquetados como positivos siendo negativos, para ello, la condición de parada ya no es que el conjunto de ejemplos esté vacío, sino que la ganancia

que se obtiene esté por debajo de un determinado umbral . **En la estructura aprendida**, es sensible a la eliminación de elementos aprendidos, ya que dejará de clasificar como positivos a los ejemplos definidos por las cláusulas eliminadas aunque seguirá pudiendo clasificar correctamente un subconjunto de los ejemplos.

15F1

PREGUNTA 4. Valoración: 2

Supóngase que se dispone de una colección de ejemplos, E={e_1(0,0), e_2(2,1), e_3(3,2), e_4(3,4), e_5(4,5), e_6(5,4)}, de la que se desconoce el valor de clase de cada uno de ellos. La información existente sobre el conocimiento del domino nos permite saber que dichos ejemplos pueden pertenecer sólo a una de tres clases posibles. Utilice entonces el **algoritmo k-medias** para:

(a) Encontrar la clase a la que pertenece cada ejemplo. Considere que, en el proceso de selección aleatorio de la semilla (primer paso del algoritmo), los ejemplos e1, e2 y e3 son los elegidos como centroides iniciales de cada clase.

(b) Suponga que se dispone de un ejemplo nuevo e_7(5,0), ¿a qué clase pertenecería según el modelo aprendido en el epígrafe anterior?

E = {e_1(0, 0), e_2(2, 1), e_3(3, 2), e_4(3, 4), e_5(4, 5), e_6(5, 4)}, que sólo pueden pertenecer a una de tres clases posibles.

Decido usar el criterio de convergencia que compruebe si durante 2 iteraciones no cambia las clasificaciones de los ejemplos.

a)

e_1, e_2 y e_3 son elegidos los centroides iniciales de cada clase, por lo que se calcula la distancia euclídea del resto de ejemplos con estos centroides.

$d(e_4,e_1) = \sqrt{(3-0)^2+(4-0)^2} = \sqrt{25} = 5$
$d(e_5,e_1) = \sqrt{(4-0)^2+(5-0)^2} = \sqrt{41} = 6.40$
$d(e_6,e_1) = \sqrt{(5-0)^2+(4-0)^2} = \sqrt{41} = 6.40$

$d(e_4,e_2) = \sqrt{(3-2)^2+(4-1)^2} = \sqrt{10} = 3.16$
$d(e_5,e_2) = \sqrt{(4-2)^2+(5-1)^2} = \sqrt{20} = 4.47$
$d(e_6,e_2) = \sqrt{(5-2)^2+(4-1)^2} = \sqrt{18} = 4.24$

$d(e_4,e_3) = \sqrt{(3-3)^2+(4-2)^2} = \sqrt{4} = 2$
$d(e_5,e_3) = \sqrt{(4-3)^2+(5-2)^2} = \sqrt{10} = 3.16$
$d(e_6,e_3) = \sqrt{(5-3)^2+(4-2)^2} = \sqrt{8} = 2.82$

e_4 está más cerca de e_3 que de e_1 o e_2
e_5 está más cerca de e_3 que de e_1 o e_2
e_6 está más cerca de e_3 que de e_1 o e_2

Se forman las agrupaciones (e_1), (e_2) y (e_3, e_4, e_5, e_6) y se calculan los centroides correspondientes de acuerdo a la fórmula:

$$\text{centroide} = \frac{\sum_{i=1}^{m} v_{i,1}}{m}, \frac{\sum_{i=1}^{m} v_{i,2}}{m}, \ldots, \frac{\sum_{i=1}^{m} v_{i,a}}{m},$$

con m = nº de ejemplos en el centroide y a = nº de atributos de los ejemplos

$$\text{centroide}(e_1) = c_1 = \left(\frac{\sum_{i=1}^{1} v_{i,1}}{1}, \frac{\sum_{i=1}^{1} v_{i,2}}{1}\right) = \left(\frac{(0)}{1}, \frac{(0)}{1}\right) = (0, 0)$$

$$\text{centroide}(e_2) = c_2 = \left(\frac{\sum_{i=1}^{1} v_{i,1}}{1}, \frac{\sum_{i=1}^{1} v_{i,2}}{1}\right) = \left(\frac{2}{1}, \frac{1}{1}\right) = (2, 1)$$

centroide(e_3, e_4 ,e_5, e_6) = c_2 = $\left(\frac{\sum_{i=1}^{4} v_{i,1}}{4}, \frac{\sum_{i=1}^{4} v_{i,2}}{4}\right) = \left(\frac{(3+3+4+5)}{4}, \frac{(2+4+5+4)}{4}\right)$ = (3.75, 3.75)

Se calcula la distancia euclídea del resto de ejemplos con estos centroides:

$d(e_4,c_1) = \sqrt{((3-0)^2+(4-0)^2)} = 5$
$d(e_5,c_1) = \sqrt{((4-0)^2+(5-0)^2)} = 6.40$
$d(e_6,c_1) = \sqrt{((5-0)^2+(4-0)^2)} = 6.40$

$d(e_4,c_2) = \sqrt{((3-2)^2+(4-1)^2)} = 3.16$
$d(e_5,c_2) = \sqrt{((4-2)^2+(5-1)^2)} = 4.47$
$d(e_6,c_2) = \sqrt{((5-2)^2+(4-1)^2)} = 4.24$

$d(e_4,c_3) = \sqrt{((3-3.75)^2+(4-3.75)^2)} = 0.79$
$d(e_5,c_3) = \sqrt{((4-3.75)^2+(5-3.75)^2)} = 1.27$
$d(e_6,c_3) = \sqrt{((5-3.75)^2+(4-3.75)^2)} = 1.27$

Se forman las agrupaciones (e_1), (e_2) y (e_4, e_5, e_6), que son las mismas que en la iteración anterior, por lo que acaba la ejecución del algoritmo (manteniéndose los mismos valores de los centroides anteriores).

b)

¿A qué clase pertenecería e_7(5, 0)? Pertenecería a la clase del centroide al que tuviera a menor distancia:

$d(e_7,c_1) = \sqrt{((5-0)^2+(0-0)^2)} = 5$
$d(e_7,c_2) = \sqrt{((5-2)^2+(0-1)^2)} = 3.16$
$d(e_7,c_3) = \sqrt{((5-3.75)^2+(0-3.75)^2)} = 3.95$

Como está más cerca de c_2, el ejemplo e_7 pertenecerá a la clase de este centroide c_2.

15SO

PREGUNTA 2. Valoración: 2.5

Supóngase que se dispone de una colección de ejemplos, E={e1(1,2), e2(2,1), e3(3,2), e4(3,4), e5(4,5), e6(5,4)}, de la que se desconoce el valor de clase de cada uno de ellos. La información implícita en el conocimiento del domino nos permite saber que dichos ejemplos pueden pertenecer sólo a una de dos clases posibles. Utilice entonces el **algoritmo k-medias** para:

(a) Encontrar la clase a la que pertenece cada ejemplo. Para ello, consideré que en el proceso de selección aleatorio de la semilla, en el primer paso del algoritmo, los ejemplos elegidos son e2 y e3.

(b) Suponga que se dispone de un ejemplo nuevo e6(3,3), ¿a qué clase pertenecería?

(c) Imagine que en el paso inicial de selección aleatoria de la semilla, se hubiera obtenido como valores los ejemplos e4 y e6 ¿Podría en este caso no estar asegurada la convergencia del algoritmo? Justifique su respuesta.

E = {e_1(1, 2), e_2(2, 1), e_3(3, 2), e_4(3, 4), e_5(4, 5), e_6(5, 4)}, que sólo pueden pertenecer a una de dos clases posibles.

Decido usar el criterio de convergencia que compruebe si durante 2 iteraciones no cambia las clasificaciones de los ejemplos.

a)

e_2 y e_3 son elegidos los centroides iniciales de cada clase, por lo que se calcula la distancia euclídea del resto de ejemplos con estos centroides.

$d(e_1,e_2) = \sqrt{(1-2)^2+(2-1)^2} = \sqrt{2} = 1.41$
$d(e_4,e_2) = \sqrt{(3-2)^2+(4-1)^2} = \sqrt{10} = 3.16$
$d(e_5,e_2) = \sqrt{(4-2)^2+(5-1)^2} = \sqrt{20} = 4.47$
$d(e_6,e_2) = \sqrt{(5-2)^2+(4-1)^2} = \sqrt{18} = 4.24$

$d(e_1,e_3) = \sqrt{(1-3)^2+(2-2)^2} = \sqrt{4} = 2$
$d(e_4,e_3) = \sqrt{(3-3)^2+(4-2)^2} = \sqrt{4} = 2$
$d(e_5,e_3) = \sqrt{(4-3)^2+(5-2)^2} = \sqrt{10} = 3.16$
$d(e_6,e_3) = \sqrt{(5-3)^2+(4-2)^2} = \sqrt{8} = 2.82$

e_1 está más cerca de e_2 que de e_3

e_4 está más cerca de e_3 que de e_2
e_5 está más cerca de e_3 que de e_2
e_6 está más cerca de e_3 que de e_2

Se forman las agrupaciones (e_1, e_2) y (e_3, e_4, e_5, e_6)

Se calculan los centroides correspondientes de acuerdo a la fórmula:

$$\text{centroide} = \frac{\sum_{i=1}^{m} v_{i,1}}{m}, \frac{\sum_{i=1}^{m} v_{i,2}}{m}, \ldots, \frac{\sum_{i=1}^{m} v_{i,a}}{m},$$

con m = nº de ejemplos en el centroide y a = nº de atributos de los ejemplos

$$\text{centroide}(e_1, e_2) = c_1 = \left(\frac{\sum_{i=1}^{2} v_{i,1}}{2}, \frac{\sum_{i=1}^{2} v_{i,2}}{2}\right) = \left(\frac{(1+2)}{2}, \frac{(2+1)}{2}\right) = (1.5, 1.5)$$

centroide(e_3, e_4 ,e_5, e_6) = c_2 = $\left(\frac{\sum_{i=1}^{4} v_{i,1}}{4}, \frac{\sum_{i=1}^{4} v_{i,2}}{4}\right)$ = $\left(\frac{(3+3+4+5)}{4}, \frac{(2+4+5+4)}{4}\right)$ = (3.75, 3.75)

Se calcula la distancia euclídea del resto de ejemplos con estos centroides:

$d(e_1,c_1) = \sqrt{((1-1.5)^2+(2-1.5)^2)} = 0.5$
$d(e_2,c_1) = \sqrt{((2-1.5)^2+(1-1.5)^2)} = 0.70$
$d(e_3,c_1) = \sqrt{((3-1.5)^2+(2-1.5)^2)} = 1.58$
$d(e_4,c_1) = \sqrt{((3-1.5)^2+(4-1.5)^2)} = 2.91$
$d(e_5,c_1) = \sqrt{((4-1.5)^2+(5-1.5)^2)} = 4.30$
$d(e_6,c_1) = \sqrt{((5-1.5)^2+(4-1.5)^2)} = 4.30$

$d(e_1,c_2) = \sqrt{((1-3.75)^2+(2-3.75)^2)} = 3.25$
$d(e_2,c_2) = \sqrt{((2-3.75)^2+(1-3.75)^2)} = 3.25$
$d(e_3,c_2) = \sqrt{((3-3.75)^2+(2-3.75)^2)} = 1.90$
$d(e_4,c_2) = \sqrt{((3-3.75)^2+(4-3.75)^2)} = 0.79$
$d(e_5,c_2) = \sqrt{((4-3.75)^2+(5-3.75)^2)} = 1.27$
$d(e_6,c_2) = \sqrt{((5-3.75)^2+(4-3.75)^2)} = 1.27$

e_1 está más cerca de c_1 que de c_2
e_2 está más cerca de c_1 que de c_2
e_3 está más cerca de c_1 que de c_2

e_4 está más cerca de c_2 que de c_1
e_5 está más cerca de c_2 que de c_1
e_6 está más cerca de c_2 que de c_1

Se forman las agrupaciones (e_1, e_2, e_3) y (e_4, e_5, e_6) y se calculan los centroides correspondientes:

centroide(e_1, e_2, e_3,) = c_3 = $\left(\frac{\sum_{i=1}^{3} v_{i,1}}{3}, \frac{\sum_{i=1}^{3} v_{i,2}}{3}\right)$ = $\left(\frac{(1+2+3)}{3}, \frac{(2+1+2)}{3}\right)$ = (2, 1.66)

centroide(e_4, e_5, e_6) = c_4 = $\left(\frac{\sum_{i=1}^{3} v_{i,1}}{3}, \frac{\sum_{i=1}^{3} v_{i,2}}{3}\right)$ = $\left(\frac{(3+4+5)}{3}, \frac{(4+5+4)}{3}\right)$ = (4, 4.33)

Se calcula la distancia euclídea del resto de ejemplos con estos centroides:

$d(e_1,c_3) = \sqrt{((1-2)^2+(2-1.66)^2)} = 1.05$
$d(e_2,c_3) = \sqrt{((2-2)^2+(1-1.66)^2)} = 0.66$
$d(e_3,c_3) = \sqrt{((3-2)^2+(2-1.66)^2)} = 1.05$
$d(e_4,c_3) = \sqrt{((3-2)^2+(4-1.66)^2)} = 2.53$
$d(e_5,c_3) = \sqrt{((4-2)^2+(5-1.66)^2)} = 3.88$
$d(e_6,c_3) = \sqrt{((5-2)^2+(4-1.66)^2)} = 3.80$

$d(e_1,c_4) = \sqrt{((1-4)^2+(2-4.33)^2)} = 3.80$
$d(e_2,c_4) = \sqrt{((2-4)^2+(1-4.33)^2)} = 3.88$
$d(e_3,c_4) = \sqrt{((3-4)^2+(2-4.33)^2)} = 2.53$
$d(e_4,c_4) = \sqrt{((3-4)^2+(4-4.33)^2)} = 1.05$
$d(e_5,c_4) = \sqrt{((4-4)^2+(5-4.33)^2)} = 0.66$
$d(e_6,c_4) = \sqrt{((5-4)^2+(4-4.33)^2)} = 1.05$

e_1 está más cerca de c_3 que de c_4
e_2 está más cerca de c_3 que de c_4
e_3 está más cerca de c_3 que de c_4

e_4 está más cerca de c_4 que de c_3
e_5 está más cerca de c_4 que de c_3
e_6 está más cerca de c_4 que de c_3

Se forman las agrupaciones (e_1, e_2, e_3) y (e_4, e_5, e_6), que son las mismas que en la iteración anterior, por lo que acaba la ejecución del algoritmo (manteniéndose los mismos valores de los centroides anteriores).

b)

¿A qué clase pertenecería e_7(3, 3)? Pertenecería a la clase del centroide al que tuviera a menor distancia:

$d(e_7,c_3) = \sqrt{((3-2)^2+(3-1.66)^2)} = 1.66$ $d(e_7,c_4) = \sqrt{((3-4)^2+(3-4.33)^2)} = 1.66$

Pero aquí tienen el mismo valor, lo que produce una indeterminación.

c)

Si e_4 y e_6 fuesen elegidos los centroides de cada clase, ¿podría estar asegurada la convergencia del algoritmo?

Tomar esos ejemplos como centroides iniciales hace que el ejemplo e_5 esté a la misma distancia de ambos y teniendo encuenta que el algoritmo básico no prevee qué hacer en caso de empate, llegaríamos a una indeterminación que no podría asegurar la convergencia del algoritmo.

17F2

<u>PREGUNTA 4</u>. Valoración: 2.5

Supóngase que se dispone de una colección de ejemplos, E={e_1(1,1), e_2(1,2), e_3(2,1), e_4(4,2), e_5(5,2), e_6(4,3)}, de la que se desconoce el valor de clase de cada uno de ellos. La información existente sobre el conocimiento del domino nos permite saber que dichos ejemplos pueden pertenecer sólo a una de dos clases posibles. Utilice entonces el **algoritmo k-medias** para:

(a) Encontrar la composición de ejemplos de las dos agrupaciones finales. Considere que, en el proceso de selección aleatorio de la semilla (primer paso del algoritmo), los ejemplos e1 y e3 son los elegidos como centroides iniciales de cada clase.

(b) Suponga que se dispone de un ejemplo nuevo e_7(5,2), ¿a qué clase pertenecería según el modelo aprendido en el epígrafe anterior?

<u>Nota importante</u>: Recuerde que, en cada iteración del algoritmo y hasta que se produzca la convergencia, debe mostrar y justificar la pertenencia de cada ejemplo al centroide correspondiente y el valor del nuevo centroide de cada una de las dos nuevas agrupaciones resultantes.

E = {e_1(1, 1), e_2(1, 2), e_3(2, 1), e_4(4, 2), e_5(5, 2), e_6(4, 3)}, que sólo pueden pertenecer a una de doss clases posibles.

Decido usar el criterio de convergencia que compruebe si durante 2 iteraciones no cambia las clasificaciones de los ejemplos.

a)

e_1 y e_3 son elegidos los centroides iniciales de cada clase, por lo que se calcula la distancia euclídea del resto de ejemplos con estos centroides.

$d(e_2,e_1) = \sqrt{(1-1)^2+(2-1)^2} = \sqrt{1} = 1$
$d(e_4,e_1) = \sqrt{(4-1)^2+(2-1)^2} = \sqrt{10} = 3.16$
$d(e_5,e_1) = \sqrt{(5-1)^2+(2-1)^2} = \sqrt{17} = 4.12$
$d(e_6,e_1) = \sqrt{(4-1)^2+(3-1)^2} = \sqrt{13} = 3.60$

$d(e_2,e_3) = \sqrt{(1-2)^2+(2-1)^2} = \sqrt{2} = 1.41$
$d(e_4,e_3) = \sqrt{(4-2)^2+(2-1)^2} = \sqrt{5} = 2.23$
$d(e_5,e_3) = \sqrt{(5-2)^2+(2-1)^2} = \sqrt{10} = 3.16$
$d(e_6,e_3) = \sqrt{(4-2)^2+(3-1)^2} = \sqrt{8} = 2.82$

e_2 está más cerca de e_1 que de e_3

e_4 está más cerca de e_3 que de e_1
e_5 está más cerca de e_3 que de e_1
e_6 está más cerca de e_3 que de e_1

Se forman las agrupaciones (e_1, e_2) y (e_3, e_4, e_5, e_6)

Se calculan los centroides correspondientes de acuerdo a la fórmula:

$$\text{centroide} = \frac{\sum_{i=1}^{m} v_{i,1}}{m}, \frac{\sum_{i=1}^{m} v_{i,2}}{m}, \ldots, \frac{\sum_{i=1}^{m} v_{i,a}}{m},$$

con m = nº de ejemplos en el centroide y a = nº de atributos de los ejemplos

$$\text{centroide}(e_1, e_2) = c_1 = \left(\frac{\sum_{i=1}^{2} v_{i,1}}{2}, \frac{\sum_{i=1}^{2} v_{i,2}}{2}\right) = \left(\frac{(1+1)}{2}, \frac{(1+2)}{2}\right) = (1, 1.5)$$

centroide(e_3, e_4 ,e_5, e_6) = c_2 = $\left(\frac{\sum_{i=1}^{4} v_{i,1}}{4}, \frac{\sum_{i=1}^{4} v_{i,2}}{4}\right) = \left(\frac{(2+4+5+4)}{4}, \frac{(1+2+2+3)}{4}\right)$ = (3.75, 2)

Se calcula la distancia euclídea del resto de ejemplos con estos centroides:

$d(e_1,c_1) = \sqrt{((1-1)^2+(1-1.5)^2)} = 0.5$
$d(e_2,c_1) = \sqrt{((1-1)^2+(2-1.5)^2)} = 0.5$
$d(e_3,c_1) = \sqrt{((2-1)^2+(1-1.5)^2)} = 1.11$
$d(e_4,c_1) = \sqrt{((4-1)^2+(2-1.5)^2)} = 3.04$
$d(e_5,c_1) = \sqrt{((5-1)^2+(2-1.5)^2)} = 4.03$
$d(e_6,c_1) = \sqrt{((4-1)^2+(3-1.5)^2)} = 3.35$

$d(e_1,c_2) = \sqrt{((1-3.75)^2+(1-2)^2)} = 2.29$
$d(e_2,c_2) = \sqrt{((1-3.75)^2+(2-2)^2)} = 2.75$
$d(e_3,c_2) = \sqrt{((2-3.75)^2+(1-2)^2)} = 2.01$
$d(e_4,c_2) = \sqrt{((4-3.75)^2+(2-2)^2)} = 0.25$
$d(e_5,c_2) = \sqrt{((5-3.75)^2+(2-2)^2)} = 1.25$
$d(e_6,c_2) = \sqrt{((4-3.75)^2+(3-2)^2)} = 1.03$

e_1 está más cerca de c_1 que de c_2
e_2 está más cerca de c_1 que de c_2
e_3 está más cerca de c_1 que de c_2

e_4 está más cerca de c_2 que de c_1
e_5 está más cerca de c_2 que de c_1
e_6 está más cerca de c_2 que de c_1

Se forman las agrupaciones (e_1, e_2, e_3) y (e_4, e_5, e_6) y se calculan los centroides correspondientes:

centroide(e_1, e_2, e_3,) = c_3 = $\left(\frac{\sum_{i=1}^{3} v_{i,1}}{3}, \frac{\sum_{i=1}^{3} v_{i,2}}{3}\right) = \left(\frac{(1+1+2)}{3}, \frac{(1+2+1)}{3}\right)$ = (1.33, 1.33)

centroide(e_4, e_5, e_6) = c_4 = $\left(\frac{\sum_{i=1}^{3} v_{i,1}}{3}, \frac{\sum_{i=1}^{3} v_{i,2}}{3}\right) = \left(\frac{(4+5+4)}{3}, \frac{(2+2+3)}{3}\right)$ = (4.33, 2.33)

Se calcula la distancia euclídea del resto de ejemplos con estos centroides:

$d(e_1,c_3) = \sqrt{((1-1.33)^2+(1-1.33)^2)} = 0.47$
$d(e_2,c_3) = \sqrt{((1-1.33)^2+(2-1.33)^2)} = 0.74$
$d(e_3,c_3) = \sqrt{((2-1.33)^2+(1-1.33)^2)} = 0.74$
$d(e_4,c_3) = \sqrt{((4-1.33)^2+(2-1.33)^2)} = 2.74$
$d(e_5,c_3) = \sqrt{((5-1.33)^2+(2-1.33)^2)} = 3.72$
$d(e_6,c_3) = \sqrt{((4-1.33)^2+(3-1.33)^2)} = 3.14$

$d(e_1,c_4) = \sqrt{((1-4.33)^2+(1-2.33)^2)} = 3.59$
$d(e_2,c_4) = \sqrt{((1-4.33)^2+(2-2.33)^2)} = 3.34$
$d(e_3,c_4) = \sqrt{((2-4.33)^2+(1-2.33)^2)} = 2.68$
$d(e_4,c_4) = \sqrt{((4-4.33)^2+(2-2.33)^2)} = 0.47$
$d(e_5,c_4) = \sqrt{((5-4.33)^2+(2-2.33)^2)} = 0.74$
$d(e_6,c_4) = \sqrt{((4-4.33)^2+(3-2.33)^2)} = 0.74$

e_1 está más cerca de c_3 que de c_4
e_2 está más cerca de c_3 que de c_4
e_3 está más cerca de c_3 que de c_4

e_4 está más cerca de c_4 que de c_3
e_5 está más cerca de c_4 que de c_3
e_6 está más cerca de c_4 que de c_3

Se forman las agrupaciones (e_1, e_2, e_3) y (e_4, e_5, e_6), que son las mismas que en la iteración anterior, por lo que acaba la ejecución del algoritmo (manteniéndose los mismos valores de los centroides anteriores).

b)

¿A qué clase pertenecería e_7(5, 2)? Pertenecería a la clase del centroide al que tuviera a menor distancia:

$d(e_7,c_3) = \sqrt{((5-1.33)^2+(2-1.33)^2)} = 3.72$

$d(e_7,c_4) = \sqrt{((5-4.33)^2+(2-2.33)^2)} = 0.74$

Como está más cerca de c_4, el ejemplo e_7 pertenecerá a la clase de este centroide c_4.

18F2

PREGUNTA 4. Valoración: 2.5

Supóngase que se dispone de una colección de ejemplos, E={e_1(0,1), e_2(1,0), e_3(2,1), e_4(2,3), e_5(3,4), e_6(4,3)}, de la que se desconoce el valor de clase de cada uno de ellos. La información existente sobre el conocimiento del domino nos permite saber que dichos ejemplos pueden pertenecer sólo a una de dos clases posibles. Utilice entonces el **algoritmo k-medias** para:

(a) Encontrar la clase a la que pertenece cada ejemplo. Para ello, consideré que en el proceso de selección aleatorio de la semilla, en el primer paso del algoritmo, los ejemplos elegidos son e2 y e3.

(b) Suponga que se dispone de un ejemplo nuevo e_0(2,2), ¿a qué clase pertenecería?

E = {e_1(0, 1), e_2(1, 0), e_3(2, 1), e_4(2, 3), e_5(3, 4), e_6(4, 3)}, que sólo pueden pertenecer a una de dos clases posibles.

Decido usar el criterio de convergencia que compruebe si durante 2 iteraciones no cambia las clasificaciones de los ejemplos.

a)

e_2 y e_3 son elegidos los centroides iniciales de cada clase, por lo que se calcula la distancia euclídea del resto de ejemplos con estos centroides.

$d(e_1,e_2) = \sqrt{(0-1)^2+(1-0)^2} = \sqrt{2} = 1.41$
$d(e_4,e_2) = \sqrt{(2-1)^2+(3-0)^2} = \sqrt{10} = 3.16$
$d(e_5,e_2) = \sqrt{(3-1)^2+(4-0)^2} = \sqrt{20} = 4.47$
$d(e_6,e_2) = \sqrt{(4-1)^2+(3-0)^2} = \sqrt{18} = 4.24$

$d(e_1,e_3) = \sqrt{(0-2)^2+(1-1)^2} = \sqrt{4} = 2$
$d(e_4,e_3) = \sqrt{(2-2)^2+(3-1)^2} = \sqrt{4} = 2$
$d(e_5,e_3) = \sqrt{(3-2)^2+(4-1)^2} = \sqrt{10} = 3.16$
$d(e_6,e_3) = \sqrt{(4-2)^2+(3-1)^2} = \sqrt{8} = 2.82$

e_1 está más cerca de e_2 que de e_3

e_4 está más cerca de e_3 que de e_2
e_5 está más cerca de e_3 que de e_2
e_6 está más cerca de e_3 que de e_2

Se forman las agrupaciones (e_1, e_2) y (e_3, e_4, e_5, e_6)

Se calculan los centroides correspondientes de acuerdo a la fórmula:

$$\text{centroide} = \frac{\sum_{i=1}^{m} v_{i,1}}{m}, \frac{\sum_{i=1}^{m} v_{i,2}}{m}, \ldots, \frac{\sum_{i=1}^{m} v_{i,a}}{m} ,$$

con m = nº de ejemplos en el centroide y a = nº de atributos de los ejemplos

$$\text{centroide}(e_1, e_2) = c_1 = \left(\frac{\sum_{i=1}^{2} v_{i,1}}{2}, \frac{\sum_{i=1}^{2} v_{i,2}}{2}\right) = \left(\frac{(2+1)}{2}, \frac{(1+2)}{2}\right) = (1.5, 1.5)$$

$$\text{centroide}(e_3, e_4, e_5, e_6) = c_2 = \left(\frac{\sum_{i=1}^{4} v_{i,1}}{4}, \frac{\sum_{i=1}^{4} v_{i,2}}{4}\right) = \left(\frac{(2+4+5+4)}{4}, \frac{(3+3+4+5)}{4}\right) = (3.75, 3.75)$$

Se calcula la distancia euclídea del resto de ejemplos con estos centroides:

$d(e_1,c_1) = \sqrt{((0-1.5)^2+(1-1.5)^2)} = 1.58$ $d(e_1,c_2) = \sqrt{((0-3.75)^2+(1-3.75)^2)} = 4.65$
$d(e_2,c_1) = \sqrt{((1-1.5)^2+(0-1.5)^2)} = 1.58$ $d(e_2,c_2) = \sqrt{((1-3.75)^2+(0-3.75)^2)} = 4.65$
$d(e_3,c_1) = \sqrt{((2-1.5)^2+(1-1.5)^2)} = 0.70$ $d(e_3,c_2) = \sqrt{((2-3.75)^2+(1-3.75)^2)} = 3.25$
$d(e_4,c_1) = \sqrt{((2-1.5)^2+(3-1.5)^2)} = 1.58$ $d(e_4,c_2) = \sqrt{((2-3.75)^2+(3-3.75)^2)} = 1.90$
$d(e_5,c_1) = \sqrt{((3-1.5)^2+(4-1.5)^2)} = 2.91$ $d(e_5,c_2) = \sqrt{((3-3.75)^2+(4-3.75)^2)} = 0.79$
$d(e_6,c_1) = \sqrt{((4-1.5)^2+(3-1.5)^2)} = 2.91$ $d(e_6,c_2) = \sqrt{((4-3.75)^2+(3-3.75)^2)} = 0.79$

e_1 está más cerca de c_1 que de c_2 e_5 está más cerca de c_2 que de c_1
e_2 está más cerca de c_1 que de c_2 e_6 está más cerca de c_2 que de c_1
e_3 está más cerca de c_1 que de c_2
e_4 está más cerca de c_1 que de c_2

Se forman las agrupaciones (e_1, e_2, e_3, e_4) y (e_5, e_6) y se calculan los centroides correspondientes:

$$\text{centroide}(e_1, e_2, e_3, e_4) = c_3 = \left(\frac{\sum_{i=1}^{4} v_{i,1}}{4}, \frac{\sum_{i=1}^{4} v_{i,2}}{4}\right) = \left(\frac{(0+1+2+2)}{4}, \frac{(1+0+1+3)}{4}\right) = (1.25, 1.25)$$

$$\text{centroide}(e_5, e_6) = c_4 = \left(\frac{\sum_{i=1}^{2} v_{i,1}}{2}, \frac{\sum_{i=1}^{2} v_{i,2}}{2}\right) = \left(\frac{(3+4)}{2}, \frac{(4+3)}{2}\right) = (3.5, 3.5)$$

Se calcula la distancia euclídea del resto de ejemplos con estos centroides:

$d(e_1,c_1) = \sqrt{((0-1.25)^2+(1-1.25)^2)} = 1.27$ $d(e_1,c_2) = \sqrt{((0-3.5)^2+(1-3.5)^2)} = 4.30$
$d(e_2,c_1) = \sqrt{((1-1.25)^2+(0-1.25)^2)} = 1.27$ $d(e_2,c_2) = \sqrt{((1-3.5)^2+(0-3.5)^2)} = 4.30$
$d(e_3,c_1) = \sqrt{((2-1.25)^2+(1-1.25)^2)} = 0.79$ $d(e_3,c_2) = \sqrt{((2-3.5)^2+(1-3.5)^2)} = 2.91$
$d(e_4,c_1) = \sqrt{((2-1.25)^2+(3-1.25)^2)} = 1.90$ $d(e_4,c_2) = \sqrt{((2-3.5)^2+(3-3.5)^2)} = 1.58$
$d(e_5,c_1) = \sqrt{((3-1.25)^2+(4-1.25)^2)} = 3.25$ $d(e_5,c_2) = \sqrt{((3-3.5)^2+(4-3.5)^2)} = 0.70$
$d(e_6,c_1) = \sqrt{((4-1.25)^2+(3-1.25)^2)} = 3.25$ $d(e_6,c_2) = \sqrt{((4-3.5)^2+(3-3.5)^2)} = 0.70$

e_1 está más cerca de c_1 que de c_2 e_4 está más cerca de c_1 que de c_2
e_2 está más cerca de c_1 que de c_2 e_5 está más cerca de c_2 que de c_1
e_3 está más cerca de c_1 que de c_2 e_6 está más cerca de c_2 que de c_1

Se forman las agrupaciones (e_1, e_2, e_3) y (e_4, e_5, e_6) y se calculan los centroides correspondientes:

$$\text{centroide}(e_1, e_2, e_3,) = c_3 = \left(\frac{\sum_{i=1}^{3} v_{i,1}}{3}, \frac{\sum_{i=1}^{3} v_{i,2}}{3}\right) = \left(\frac{(0+1+2)}{3}, \frac{(1+0+1)}{3}\right) = (1, 0.66)$$

$$\text{centroide}(e_4, e_5, e_6) = c_4 = \left(\frac{\sum_{i=1}^{3} v_{i,1}}{3}, \frac{\sum_{i=1}^{3} v_{i,2}}{3}\right) = \left(\frac{(2+3+4)}{3}, \frac{(3+4+3)}{3}\right) = (1, 3.33)$$

Se calcula la distancia euclídea del resto de ejemplos con estos centroides:

$d(e_1,c_3) = \sqrt{((0-1)^2+(1-0.66)^2)} = 1.05$ $d(e_1,c_4) = \sqrt{((0-1)^2+(1-3.33)^2)} = 2.53$
$d(e_2,c_3) = \sqrt{((1-1)^2+(0-0.66)^2)} = 0.66$ $d(e_2,c_4) = \sqrt{((1-1)^2+(0-3.33)^2)} = 3.33$
$d(e_3,c_3) = \sqrt{((2-1)^2+(1-0.66)^2)} = 1.05$ $d(e_3,c_4) = \sqrt{((2-1)^2+(1-3.33)^2)} = 2.53$
$d(e_4,c_3) = \sqrt{((2-1)^2+(3-0.66)^2)} = 2.53$ $d(e_4,c_4) = \sqrt{((2-1)^2+(3-3.33)^2)} = 1.05$
$d(e_5,c_3) = \sqrt{((3-1)^2+(4-0.66)^2)} = 3.88$ $d(e_5,c_4) = \sqrt{((3-1)^2+(4-3.33)^2)} = 2.10$
$d(e_6,c_3) = \sqrt{((4-1)^2+(3-0.66)^2)} = 3.80$ $d(e_6,c_4) = \sqrt{((4-1)^2+(3-3.33)^2)} = 3.01$

e_1 está más cerca de c_3 que de c_4
e_2 está más cerca de c_3 que de c_4
e_3 está más cerca de c_3 que de c_4

e_4 está más cerca de c_4 que de c_3
e_5 está más cerca de c_4 que de c_3
e_6 está más cerca de c_4 que de c_3

Se forman las agrupaciones (e_1, e_2, e_3) y (e_4, e_5, e_6), que son las mismas que en la iteración anterior, por lo que acaba la ejecución del algoritmo (manteniéndose los mismos valores de los centroides anteriores).

b)
¿A qué clase pertenecería $e_0(2, 2)$? Pertenecería a la clase del centroide al que tuviera a menor distancia:

$d(e_7,c_3) = \sqrt{((2-1)^2+(2-0.66)^2)} = 1.66$

$d(e_7,c_4) = \sqrt{((2-1)^2+(2-3.33)^2)} = 1.66$

Pero aquí tienen el mismo valor, lo que produce una indeterminación.

18SO

<u>PREGUNTA 1</u>. Valoración: 2

Describa las características del **algoritmo k-Medias** atendiendo a los siguientes aspectos:

a) *Bias.*

b) Cómo se comparte ante el ruido: en entradas y en estructura.

c) Complejidad: espacial y temporal.

a) *Bias*

El número k, que especifica el número de clases necesaria y la función de similitud, que selecciona la clase de la cuál está más cercana.

b) *Cómo se comporta ante el ruido: en entradas y en estructura*

<u>En entradas</u>: no lo trata directamete, pero al incluir ele ejemplo en la clase más parecida, lo suaviza.

<u>En estructura</u>: la desaparición de una clase afecta solo a los ejemplos parecidosa esa clase, aunque siempre se podrá clasificar con la siguiente clas emás parecida.

c) *Complejidad: espacial y temporal*

<u>Espacio</u>: al guardar solo los centroides de cada clase, el espacio es proporcional a k (nº de clases)

<u>Tiempo</u>: en la inicialización se tienen k pasos y en el procedimiento, por cada iteración, k comparaciones (total N * k), en las que se pregunta por el valor de cada atributo (N * k * A). Y para un numero I de iteraciones: I * N * k * A.

19F2 – Pregunta 3.a

PREGUNTA 3. Valoración: 3

Sea la siguiente colección de ejemplos: E={e_1(0,1,3), e_2(1,0,2), e_3(2,1,1), e_4(2,3,0), e_5(3,4,2), e_6(4,3,2)}.

(a) Suponga que se desconoce el valor de clase de cada ejemplo y que se sabe que dichos ejemplos pueden pertenecer sólo a una de dos clases posibles. Utilice entonces el **algoritmo k-medias** para encontrar la clase a la que pertenece cada ejemplo. Para ello, consideré que, en el proceso de selección aleatorio de la semilla en el primer paso del algoritmo, los ejemplos elegidos son e2 y e3.

(b) Suponga ahora que se conoce la clase a la que pertenece cada ejemplo y es la siguiente: e_1(+), e_2(-), e_3(-), e_4(+), e_5(+) y e_6(-). Utilice el **algoritmo k-NN**, con k=3, para conocer la clase del ejemplo de clasificación desconocida: e_0(4,3,1; ?). Para ello, use la *clase de la mayoría* como función de asignación de clase de e_0.

E = {e_1(0, 1, 3), e_2(1, 0, 2), e_3(2, 1, 1), e_4(2, 3, 0), e_5(3, 4, 2), e_6(4, 3, 2)}, que sólo pueden pertenecer a una de dos clases posibles.

Decido usar el criterio de convergencia que compruebe si durante 2 iteraciones no cambia las clasificaciones de los ejemplos.

a)

e_2 y e_3 son elegidos los centroides iniciales de cada clase, por lo que se calcula la distancia euclídea del resto de ejemplos con estos centroides.

$d(e_1,e_2) = \sqrt{((0-1)^2+(1-0)^2+(3-2)^2)} = 1.73$
$d(e_4,e_2) = \sqrt{((2-1)^2+(3-0)^2+(0-2)^2)} = 3.74$
$d(e_5,e_2) = \sqrt{((3-1)^2+(4-0)^2+(2-2)^2)} = 4.47$
$d(e_6,e_2) = \sqrt{((4-1)^2+(3-0)^2+(2-2)^2)} = 4.24$

$d(e_1,e_3) = \sqrt{((0-2)^2+(1-1)^2+(1-3)^2)} = 2.82$
$d(e_4,e_3) = \sqrt{((2-2)^2+(3-1)^2+(0-3)^2)} = 3.60$
$d(e_5,e_3) = \sqrt{((3-2)^2+(4-1)^2+(2-3)^2)} = 3.31$
$d(e_6,e_3) = \sqrt{((4-2)^2+(3-1)^2+(2-3)^2)} = 3$

e_1 está más cerca de e_2 que de e_3

e_4 está más cerca de e_3 que de e_2
e_5 está más cerca de e_3 que de e_2
e_6 está más cerca de e_3 que de e_2

Se forman las agrupaciones (e_1, e_2) y (e_3, e_4, e_5, e_6) y se calculan los centroides correspondientes de acuerdo a la fórmula:

$$\text{centroide} = \frac{\sum_{i=1}^{m} v_{i,1}}{m}, \frac{\sum_{i=1}^{m} v_{i,2}}{m}, \ldots, \frac{\sum_{i=1}^{m} v_{i,a}}{m},$$

con m = nº de ejemplos en el centroide y a = nº de atributos de los ejemplos

$$\text{centroide}(e_1, e_2) = c_1 = \left(\frac{\sum_{i=1}^{2} v_{i,1}}{2}, \frac{\sum_{i=1}^{2} v_{i,2}}{2}\right) = \left(\frac{(0+1)}{2}, \frac{(1+0)}{2}, \frac{(3+2)}{2}\right)$$
$$= (0.5, 0.5, 2.5)$$

centroide(e_3, e_4 ,e_5, e_6) = c_2 = $\left(\frac{\sum_{i=1}^{4} v_{i,1}}{4}, \frac{\sum_{i=1}^{4} v_{i,2}}{4}\right)$ = $\left(\frac{(2+2+3+4)}{4}, \frac{(1+3+4+3)}{4}, \frac{(3+0+2+2)}{4}\right)$
= (2.75, 2.75, 1.75)

se calcula la distancia euclídea de los centroides al resto de ejemplos:

$d(e_1,c_1) = \sqrt{((0-0.5)^2+(1-0.5)^2+(3-2.5)^2)} = 0.86$
$d(e_2,c_1) = \sqrt{((1-0.5)^2+(0-0.5)^2+(2-2.5)^2)} = 0.86$
$d(e_3,c_1) = \sqrt{((2-0.5)^2+(1-0.5)^2+(1-2.5)^2)} = 2.17$
$d(e_4,c_1) = \sqrt{((2-0.5)^2+(3-0.5)^2+(0-2.5)^2)} = 2.56$
$d(e_5,c_1) = \sqrt{((3-0.5)^2+(4-0.5)^2+(2-2.5)^2)} = 4.33$
$d(e_6,c_1) = \sqrt{((4-0.5)^2+(3-0.5)^2+(2-2.5)^2)} = 4.33$

$d(e_1,c_2) = \sqrt{((0-2.75)^2+(1-2.75)^2+(3-1.75)^2)} = 3.49$
$d(e_2,c_2) = \sqrt{((1-2.75)^2+(0-2.75)^2+(2-1.75)^2)} = 3.26$
$d(e_3,c_2) = \sqrt{((2-2.75)^2+(1-2.75)^2+(1-1.75)^2)} = 2.04$
$d(e_4,c_2) = \sqrt{((2-2.75)^2+(3-2.75)^2+(0-1.75)^2)} = 1.92$
$d(e_5,c_2) = \sqrt{((3-2.75)^2+(4-2.75)^2+(2-1.75)^2)} = 1.29$
$d(e_6,c_2) = \sqrt{((4-2.75)^2+(3-2.75)^2+(2-1.75)^2)} = 1.29$

e_1 está más cerca de c_1 que de c_2
e_2 está más cerca de c_1 que de c_2

e_3 está más cerca de c_1 que de c_2
e_4 está más cerca de c_2 que de c_1
e_5 está más cerca de c_2 que de c_1
e_6 está más cerca de c_2 que de c_1

Se forman las agrupaciones (e_1, e_2) y (e_3, e_4, e_5, e_6), que son las mismas que en la iteración anterior, por lo que acaba la ejecución del algoritmo (manteniéndose los mismos valores de los centroides anteriores).

19SO

PREGUNTA 3. Valoración: 2

Supóngase que se dispone de la siguiente colección de ejemplos:

$$E=\{e_1(2,1), e_2(1,2), e_3(2,3), e_4(4,3), e_5(5,4), e_6(4,5)\},$$

de la que se desconoce el valor de clase de cada uno de ellos. La información implícita en el conocimiento del domino nos permite saber que dichos ejemplos pueden pertenecer sólo a una de dos clases posibles. Utilice entonces el **algoritmo k-medias** para:

(a) Encontrar la clase a la que pertenece cada ejemplo. Para ello, consideré que en el proceso de selección aleatorio de la semilla, en el primer paso del algoritmo, los ejemplos elegidos son e_2 y e_3.

(b) Suponga que se dispone de un ejemplo nuevo $e_7(3,3)$. Clasifique dicho ejemplo de acuerdo al modelo obtenido en el paso anterior.

E = $\{e_1(2, 1), e_2(1, 2), e_3(2, 3), e_4(4, 3), e_5(5, 4), e_6(4, 5)\}$, que sólo pueden pertenecer a una de dos clases posibles.

Decido usar el criterio de convergencia que compruebe si durante 2 iteraciones no cambia las clasificaciones de los ejemplos.

a)

e_2 y e_3 son elegidos los centroides iniciales de cada clase, por lo que se calcula la distancia euclídea del resto de ejemplos con estos centroides.

$d(e_1,e_2) = \sqrt{((2-1)^2+(1-2)^2)} = \sqrt{2} = 1.41$
$d(e_4,e_2) = \sqrt{((4-1)^2+(3-2)^2)} = \sqrt{10} = 3.16$
$d(e_5,e_2) = \sqrt{((5-1)^2+(4-2)^2)} = \sqrt{20} = 4.47$
$d(e_6,e_2) = \sqrt{((4-1)^2+(5-2)^2)} = \sqrt{18} = 4.24$

$d(e_1,e_3) = \sqrt{((2-2)^2+(1-3)^2)} = \sqrt{4} = 2$
$d(e_4,e_3) = \sqrt{((4-2)^2+(3-3)^2)} = \sqrt{4} = 2$
$d(e_5,e_3) = \sqrt{((5-2)^2+(4-3)^2)} = \sqrt{10} = 3.16$
$d(e_6,e_3) = \sqrt{((4-2)^2+(5-3)^2)} = \sqrt{8} = 2.82$

e_1 está más cerca de e_2 que de e_3

e_4 está más cerca de e_3 que de e_2
e_5 está más cerca de e_3 que de e_2
e_6 está más cerca de e_3 que de e_2

Se forman las agrupaciones (e_1, e_2) y (e_3, e_4, e_5, e_6)

Se calculan los centroides correspondientes de acuerdo a la fórmula:

$$\text{centroide} = \frac{\sum_{i=1}^{m} v_{i,1}}{m}, \frac{\sum_{i=1}^{m} v_{i,2}}{m}, \ldots, \frac{\sum_{i=1}^{m} v_{i,a}}{m},$$

con m = nº de ejemplos en el centroide y a = nº de atributos de los ejemplos

$$\text{centroide}(e_1, e_2) = c_1 = \left(\frac{\sum_{i=1}^{2} v_{i,1}}{2}, \frac{\sum_{i=1}^{2} v_{i,2}}{2}\right) = \left(\frac{(2+1)}{2}, \frac{(1+2)}{2}\right) = (1.5, 1.5)$$

$$\text{centroide}(e_3, e_4, e_5, e_6) = c_2 = \left(\frac{\sum_{i=1}^{4} v_{i,1}}{4}, \frac{\sum_{i=1}^{4} v_{i,2}}{4}\right) = \left(\frac{(2+4+5+4)}{4}, \frac{(3+3+4+5)}{4}\right) = (3.75, 3.75)$$

Se calcula la distancia euclídea del resto de ejemplos con estos centroides:

$d(e_1,c_1) = \sqrt{((2-1.5)^2+(1-1.5)^2)} = 0.70$
$d(e_2,c_1) = \sqrt{((1-1.5)^2+(2-1.5)^2)} = 0.70$
$d(e_3,c_1) = \sqrt{((2-1.5)^2+(3-1.5)^2)} = 1.58$
$d(e_4,c_1) = \sqrt{((4-1.5)^2+(3-1.5)^2)} = 2.91$
$d(e_5,c_1) = \sqrt{((5-1.5)^2+(4-1.5)^2)} = 4.30$
$d(e_6,c_1) = \sqrt{((4-1.5)^2+(5-1.5)^2)} = 4.30$

$d(e_1,c_2) = \sqrt{((2-3.75)^2+(1-3.75)^2)} = 3.25$
$d(e_2,c_2) = \sqrt{((1-3.75)^2+(2-3.75)^2)} = 3.25$
$d(e_3,c_2) = \sqrt{((2-3.75)^2+(3-3.75)^2)} = 1.90$
$d(e_4,c_2) = \sqrt{((4-3.75)^2+(3-3.75)^2)} = 0.79$
$d(e_5,c_2) = \sqrt{((5-3.75)^2+(4-3.75)^2)} = 1.27$
$d(e_6,c_2) = \sqrt{((4-3.75)^2+(5-3.75)^2)} = 1.27$

e_1 está más cerca de c_1 que de c_2
e_2 está más cerca de c_1 que de c_2
e_3 está más cerca de c_1 que de c_2

e_4 está más cerca de c_2 que de c_3
e_5 está más cerca de c_2 que de c_3
e_6 está más cerca de c_2 que de c_3

Se forman las agrupaciones (e_1, e_2 ,e_3) y (e_4, e_5, e_6) y se calculan los centroides correspondientes:

$$\text{centroide}(e_1, e_2, e_3) = c_3 = \left(\frac{\sum_{i=1}^{3} v_{i,1}}{3}, \frac{\sum_{i=1}^{3} v_{i,2}}{3}\right) = \left(\frac{(2+1+2)}{3}, \frac{(1+2+3)}{3}\right) = (1.66, 2)$$

$$\text{centroide}(e_4, e_5, e_6) = c_4 = \left(\frac{\sum_{i=1}^{3} v_{i,1}}{3}, \frac{\sum_{i=1}^{3} v_{i,2}}{3}\right) = \left(\frac{(4+5+4)}{3}, \frac{(3+4+5)}{3}\right) = (4.33, 4)$$

Se calcula la distancia euclídea del resto de ejemplos con estos centroides:

$d(e_1,c_3) = \sqrt{((2-1.66)^2+(1-2)^2)} = 1.05$
$d(e_2,c_3) = \sqrt{((1-1.66)^2+(2-2)^2)} = 0.66$
$d(e_3,c_3) = \sqrt{((2-1.66)^2+(3-2)^2)} = 1.05$
$d(e_4,c_3) = \sqrt{((4-1.66)^2+(3-2)^2)} = 2.53$
$d(e_5,c_3) = \sqrt{((5-1.66)^2+(4-2)^2)} = 3.88$
$d(e_6,c_3) = \sqrt{((4-1.66)^2+(5-2)^2)} = 3.80$

$d(e_1,c_4) = \sqrt{((2-4.33)^2+(1-4)^2)} = 3.80$
$d(e_2,c_4) = \sqrt{((1-4.33)^2+(2-4)^2)} = 3.88$
$d(e_3,c_4) = \sqrt{((2-4.33)^2+(3-4)^2)} = 2.53$
$d(e_4,c_4) = \sqrt{((4-4.33)^2+(3-4)^2)} = 1.05$
$d(e_5,c_4) = \sqrt{((5-4.33)^2+(4-4)^2)} = 0.66$
$d(e_6,c_4) = \sqrt{((4-4.33)^2+(5-4)^2)} = 1.05$

e_1 está más cerca de c_3 que de c_4
e_2 está más cerca de c_3 que de c_4
e_3 está más cerca de c_3 que de c_4

e_4 está más cerca de c_4 que de c_3
e_5 está más cerca de c_4 que de c_3
e_6 está más cerca de c_4 que de c_3

Se forman las agrupaciones (e_1, e_2 ,e_3) y (e_4, e_5, e_6), que son las mismas que en la iteración anterior, por lo que acaba la ejecución del algoritmo (manteniéndose los mismos valores de los centroides anteriores).

b)

¿A qué clase pertenecería $e_7(3, 3)$?

$d(e_7,c_3) = \sqrt{((3-1.66)^2+(3-2)^2)} = 1.66$

$d(e_7,c_4) = \sqrt{((3-4.33)^2+(3-4)^2)} = 1.66$

Tienen el mismo valor, se produciría una indeterminación.

15F1

PREGUNTA 3. Valoración: 3

En el contexto del algoritmo k-NN (**k-vecinos más cercanos**), supóngase que se dispone de una colección de ejemplos, *E*, cada uno de ellos denotados por *e(x, y; c)* donde *x, y* representan los valores de dos atributos y *c* el valor de clase. Sea E={e_1(1, 4; +), e_2(5, 5.5; -), e_3(2, 5; +), e_4(3, 1; -), e_5(3, 9; -)} el conjunto de ejemplos de clasificación conocida y e_0(3,5; ?) la instancia a clasificar. Responda a las siguientes preguntas:

(a) Si se utiliza como medida de similitud la función $s(e_i,e_j)=1/[1+d(e_i,e_j)]$, donde $d(e_i,e_j)$ es la distancia euclídea, y como función de salida la clase de la mayoría, establezca y justifique numéricamente a qué clase pertenece e_0 en el caso de ejecutar el algoritmo con *k=3*.

(b) Compruebe que, si se vuelve a ejecutar nuevamente el algoritmo con *k*=2, se produce una indeterminación. En este caso (*k*=2), para obtener la clase de e_0, utilice como función de salida aquella que pondera los votos de los *k*-vecinos por su similitud.

E = { e_1(1, 4; +), e_2(5, 5.5; –), e_3(2, 5; +), e_4(3, 1; –), e_5(3, 9; –) }
e_0(3, 5; ?)

a)

Al tener el valor de k = 3, tenemos que comparar e_0 con las 3 instancias que más similitud con él. La función de salida indicada en el enunciado es la de *la clase de la mayoría*.

Comenzamos calculando la similitud de todos los ejemplos de E con e_0.
La función de similitud de dos ejemplos viene dada por $s(e_i, e_j)=\frac{1}{1+d(e_i, e_j)}$ donde, para ello, debemos calcular también la distancia euclídea a través de la fórmula: $d(x, y)=\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$

Ejemplos	$d(e_0, e_i)$, k= [1..5] → Distancia a e_0(3, 5; ?)	$s(e_0, e_i)$, k= [1..5] → Similitud con e_0(3, 5; ?)
e_1(1, 4; +)	$\sqrt{(1-3)^2+(4-5)^2} = \sqrt{4 + 1} = 2.23606$	1 / (1 + 2.23606) = 0.30901
e_2(5, 5.5; –)	$\sqrt{(5-3)^2+(5.5-5)^2} = \sqrt{4 + 0.25} = 2.06155$	1 / (1 + 2.06155) = 0.32037
e_3(2, 5; +)	$\sqrt{(2-3)^2+(5-5)^2} = \sqrt{1 + 0} = 1$	1 / (1 + 1) = 0.5
e_4(3, 1; –)	$\sqrt{(3-3)^2+(1-5)^2} = \sqrt{0 + 16} = 4$	1 / (1 + 4) = 0.2
e_5(3, 9; –)	$\sqrt{(3-3)^2+(9-5)^2} = \sqrt{0 + 16} = 4$	1 / (1 + 4) = 0.2

Los k = 3 ejemplos con mayor similitud son: e_1(1, 4; +), e_2(5, 5.5; –) y e_3(2, 5; +), donde hay 2 ejemplos clasificados como + y uno clasificado como –

Por tanto, la instancia se clasificaría como positivo: e_0(3, 5; +)

b)

Con k =2 tendríamos como instancias más similares a $e_3(2, 5; +)$ y a $e_2(5, 5.5; –)$, que produciría una indeterminación si la función de salida fuera la de *la clase de la mayoría*, ya que no hay clase mayoritaría, ambas tienen el mismo valor, solo un ejemplo.

Para evitar esta indeterminación, se busca una función de salida que *pondera los votos de los k-vecinos por su similitud*. Esto quiere decir que dependiendo del valor de la similitud de una instancia, ésta tendrá tendrá mayor o menor peso en la decisión de clasificación de la instancia e_0.

Para ello se suman, para cada valor de clase, los valores de la similitud de los k = 2 ejemplos que producen clase positiva y los que producen clase negativa:

votos(+) = ∑ (similitudes clase +) = similitud(e_0, e_3) = 0.5
votos(–) = ∑ (similitudes clase –) = similitud(e_0, e_2) = 0.32037

F = arg max (votos(+) , votos(–)) = positivo

Y de ambos valores, se clasifica el ejemplo como $e_0(3, 5; F) = e_0(3, 5; +)$

16SO

PREGUNTA 1. Valoración: 2

Compare el **algoritmo k-medias** con el **algoritmo k-NN (k-vecinos más cercanos)** respecto de los siguientes aspectos:

a) Bias.

b) Tratamiento del ruido: en entrada y en estructura.

a) *Bias*

El valor de k y el concepto de similitud.

b) *Tratamiento del ruido: en entrada y en estructura*

En entradas: el ruido lo suaviza la función de similitud.

En estructura: si se elimina o cambia alguno de los ejemplos, como se hace al eliminar las instancias atípicas en la variante del modelo de mejores ejemplos.

16SR – Pregunta 4 (apartado a) igual que en **15F1 – Pregunta 4**

PREGUNTA 4. Valoración: 2
Supóngase que se dispone de una colección de ejemplos, *E*, cada uno de ellos denotados por *e(x, y; c)* donde *x*, *y* representan los dos atributos que permite caracterizarlos y *c* el valor de clase a la que pertenecen. Si E={e_1(1,4; +), e_2(5, 5.5; -), e_3(2, 5; +), e_4(3, 1; -), e_5(3, 9; -)} es el conjunto de ejemplos de clasificación conocida, utilice el **algoritmo k-NN** para clasificar la instancia e_0(3,5; ?). Para realizar la clasificación final, utilice **k=3** y aplique las dos siguientes estrategias:

(a) Usando el criterio de la mayoría.

(b) Ponderando los votos de los vecinos por su similitud.

E = { e_1(1, 4; +), e_2(5, 5.5; –), e_3(2, 5; +), e_4(3, 1; –), e_5(3, 9; –) }
e_0(3, 5; ?)

a)

Al tener el valor de k = 3, tenemos que comparar e_0 con las 3 instancias que más similitud con él. La función de salida indicada en el enunciado es la de *la clase de la mayoría*.

Comenzamos calculando la similitud de todos los ejemplos de E con e_0.
La función de similitud de dos ejemplos viene dada por $s(e_i, e_j) = \frac{1}{1 + d(e_i, e_j)}$ donde, para ello, debemos calcular también la distancia euclídea a través de la fórmula: $d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$

Ejemplos	d(e_0, e_i), k= [1..5] → Distancia a e_0(3, 5; ?)	s(e_0, e_i), k= [1..5] → Similitud con e_0(3, 5; ?)
e_1(1, 4; +)	$\sqrt{(1-3)^2+(4-5)^2} = \sqrt{4 + 1} = 2.23606$	1 / (1 + 2.23606) = 0.30901
e_2(5, 5.5; –)	$\sqrt{(5-3)^2+(5.5-5)^2} = \sqrt{4 + 0.25} = 2.06155$	1 / (1 + 2.06155) = 0.32663
e_3(2, 5; +)	$\sqrt{(2-3)^2+(5-5)^2} = \sqrt{1 + 0} = 1$	1 / (1 + 1) = 0.5
e_4(3, 1; –)	$\sqrt{(3-3)^2+(1-5)^2} = \sqrt{0 + 16} = 4$	1 / (1 + 4) = 0.2
e_5(3, 9; –)	$\sqrt{(3-3)^2+(9-5)^2} = \sqrt{0 + 16} = 4$	1 / (1 + 4) = 0.2

Los k = 3 ejemplos con mayor similitud son: e_1(1, 4; +), e_2(5, 5.5; –) y e_3(2, 5; +), donde hay 2 ejemplos clasificados como + y uno clasificado como –

Por tanto, la instancia se clasificaría como <u>positivo</u>: e_0(3, 5; +)

b)

Con k = 3 y como función de salida la que *pondera los votos de los k-vecinos por su similitud* tendríamos la clasificación de e_0 de acuerdo al mayor valor de las sumas, para cada valor de clase, de los valores de la similitud de los k = 3 ejemplos que producen clase positiva y los que producen clase negativa:

votos(+) = ∑ (similitudes clase +) = similitud(e_0, e_1) + similitud(e_0, e_3) = 0.30901 + 0.5 = 0.80901
votos(–) = ∑ (similitudes clase –) = similitud(e_0, e_2) = 0.32663

F = arg max (votos(+) , votos(–)) = positivo

Y de ambos valores, se clasifica el ejemplo como e_0(3, 5; F) = e_0(3, 5; +)

18F2

PREGUNTA 2. Valoración: 2.5
Supóngase que se dispone de la siguiente colección de ejemplos:

E={ e_1(1,4,2,2; +), e_2(5,6,2,2; -), e_3(2,5,2,3; +), e_4(2,5,1,3; -), e_5(3,9,2,2; -), e_6(3,4,1,1 ;+)}

Se propone utilizar el **algoritmo k-NN**, con *k*=3, para conocer la clase del ejemplo de clasificación desconocida: e_0(3,5,2,2; ?). Para ello, tenga en cuenta los siguientes apartados:

(a) Utilice como función de asignación de clase de e_0 la *clase de la mayoría*.

(b) Pondere los votos de los *k*=3 vecinos más cercanos por su similitud para establecer el valor de clase de e_0.

E={e_1(1, 4, 2, 2; +),e_2(5, 6, 2, 2; –),e_3(2, 5, 2, 3; +), e_4(2, 5, 1, 3; –), e_5(3, 9, 2, 2; –), e_6(3, 4, 1, 1; +)}
e_0(3, 5, 2, 2; ?)

a)

Al tener el valor de k = 3, tenemos que comparar e_0 con las 3 instancias que más similitud con él. La función de salida indicada en el enunciado es la de *la clase de la mayoría*.

Comenzamos calculando la similitud de todos los ejemplos de E con e_0.

La función de similitud de dos ejemplos viene dada por $s(e_i, e_j) = \frac{1}{1 + d(e_i, e_j)}$ donde, para ello, debemos calcular también la distancia euclídea a través de la fórmula: $d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$

Ejemplos	$d(e_0, e_i)$, k= [1..6] → Distancia a e_0(3, 5, 2, 2; ?)	$s(e_0, e_i)$, k= [1..6] → Similitud con e_0(3, 5, 2, 2; ?)
e_1(1, 4, 2, 2; +)	$\sqrt{(1-3)^2+(4-5)^2+(2-2)^2+(2-2)^2}$ = $\sqrt{4+1+0+0}$ = 2.23606	1 / (1 + 2.23606) = 0.30901
e_2(5, 6, 2, 2; –)	$\sqrt{(5-3)^2+(6-5)^2+(2-2)^2+(2-2)^2}$ = $\sqrt{4+1+0+0}$ = 2.23606	1 / (1 + 2.23606) = 0.30901
e_3(2, 5, 2, 3; +)	$\sqrt{(2-3)^2+(5-5)^2+(2-2)^2+(3-2)^2}$ = $\sqrt{1+0+0+1}$ = 1.41421	1 / (1 + 1.41421) = 0.41421
e_4(2, 5, 1, 3; –)	$\sqrt{(2-3)^2+(2-5)^2+(1-2)^2+(3-2)^2}$ = $\sqrt{1+9+1+1}$ = 3.46410	1 / (1 + 3.46410) = 0.22401
e_5(3, 9, 2, 2; –)	$\sqrt{(3-3)^2+(9-5)^2+(2-2)^2+(2-2)^2}$ = $\sqrt{0+16+0+0}$ = 4	1 / (1 + 4) = 0.2
e_6(3, 4, 1, 1; +)	$\sqrt{(3-3)^2+(4-5)^2+(1-2)^2+(1-2)^2}$ = $\sqrt{0+1+1+1}$ = 1.73205	1 / (1 + 1.73205) = 0.366025

Los k = 3 ejemplos con mayor similitud son: e_1(1, 4, 2, 2; +), e_2(5, 6, 2, 2; –) y hay un empate en el tercer ejemplo, de clases distintas.

Si se utiliza criterio de la clase de la mayoría sin ponderación, no habría forma de determinar la clase.

b)

Con k = 3 y como función de salida la que *pondera los votos de los k-vecinos por su similitud* y empate en el valor de similitud del tercer ejemplo, tendríamos la misma situación que en el paso anterior, falta de información para elegir el tercer ejemplo, por lo que el algoritmo no podría clasificar el ejemplo.
En caso de empate se podría recurrir a otros criterios, como el del *siguiente ejemplo más cercano* para deshacer dicho empate.

18SO

PREGUNTA 3. Valoración: 2
Supóngase que se dispone de la siguiente colección de ejemplos:

E={ e_1(1,4,2,2,1; **+**), e_2(5,6,2,2,2; **-**), e_3(2,5,2,3,1; **+**),
e_4(2,5,1,3,1; **-**), e_5(3,9,2,2,1; **-**), e_6(3,4,1,1,1 ;**+**)}

Se propone utilizar el **algoritmo k-NN**, con *k*=3, para conocer la clase del ejemplo de clasificación desconocida e_0(3,5,2,2,0; ?). Para ello, conteste a cada uno de los dos siguientes apartados:

(a) Utilice como función de asignación de clase a e_0 la *clase de la mayoría*.

(b) Pondere los votos de los *k*=3 vecinos más cercanos por su similitud para establecer el valor de clase de e_0.

E = { e_1(1, 4, 2, 2, 1; +), e_2(5, 6, 2, 2, 2; –), e_3(2, 5, 2, 3, 1; +), e_4(2, 5, 1, 3, 1; –), e_5(3, 9, 2, 2, 1; –), e_6(3, 4, 1, 1, 1; +) }
e_0(3, 5, 2, 2, 0; ?)

a)

Al tener el valor de k = 3, tenemos que comparar e_0 con las 3 instancias que más similitud con él. La función de salida indicada en el enunciado es la de *la clase de la mayoría*.

Comenzamos calculando la similitud de todos los ejemplos de E con e_0.
La función de similitud de dos ejemplos viene dada por $s(e_i, e_j) = \frac{1}{1 + d(e_i, e_j)}$ donde, para ello, debemos calcular también la distancia euclídea a través de la fórmula: $d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$

Ejemplos	d(e_0, e_i), k= [1..6] → Distancia a e_0(3, 5, 2, 2, 0; ?)	s(e_0, e_i), k= [1..6] → Similitud con e_0(3, 5, 2, 2, 0; ?)
e_1(1, 4, 2, 2, 1; +)	√((1-3)²+(4-5)²+(2-2)²+(2-2)²+(1-0)) = √(4 + 1 + 0 + 0 + 1) = 2.44948	1 / (1 + 2.44948) = 0.28989
e_2(5, 6, 2, 2, 2; –)	√((5-3)²+(6-5)²+(2-2)²+(2-2)² +(2-0)) = √(4 + 1 + 0 + 0 + 2) = 2.64575	1 / (1 + 2.64575) = 0.27429
e_3(2, 5, 2, 3, 1; +)	√((2-3)²+(5-5)²+(2-2)²+(3-2)² +(1-0)) = √(1 + 0 + 0 + 1 + 1) = 1.73205	1 / (1 + 1.73205) = 0.36602
e_4(2, 5, 1, 3, 1; –)	√((2-3)²+(2-5)²+(1-2)²+(3-2)² +(1-0)) = √(1 + 9 + 1 + 1 + 1) = 3.60555	1 / (1 + 3.60555) = 0.21712
e_5(3, 9, 2, 2, 1; –)	√((3-3)²+(9-5)²+(2-2)²+(2-2)² +(1-0)) = √(0 + 16 + 0 + 0 + 1) = 4.12310	1 / (1 + 4.12310) = 0.19519
e_6(3, 4, 1, 1, 1; +)	√((3-3)²+(4-5)²+(1-2)²+(1-2)² +(1-0)) = √(0 + 1 + 1 + 1 + 1) = 2	1 / (1 + 2) = 0.33333

Los k=3 ejemplos con mayor similitud son e_1(1, 4, 2, 2, 1; +), e_3(2, 5, 2, 3, 1; +) y e_6(3, 4, 1, 1, 1; +)

y donde hay 3 ejemplos clasificados como + y ninguno como –

Por tanto, la instancia se clasificaría como positivo: $e_0(3, 5, 2, 2, 0; +)$

b)

Con k = 3 y como función de salida la que *pondera los votos de los k-vecinos por su similitud* tendríamos la clasificación de e_0 de acuerdo al mayor valor de las sumas, para cada valor de clase, de los valores de la similitud de los k = 3 ejemplos que producen clase positiva y los que producen clase negativa:

votos(+) = $\sum$(similitudes clase +) = 0.28989 + 0.36602 + 0,33333 = 0.98924
votos(–) = $\sum$ (similitudes clase –) = 0

F = arg max (votos(+) , votos(–)) = positivo
Y de ambos valores, se clasifica el ejemplo como $e_0(3, 5, 2, 2, 0; F) = e_0(3, 5, 2, 2; +)$

19F2 – Pregunta 3.b

PREGUNTA 3. Valoración: 3
Sea la siguiente colección de ejemplos: E={e_1(0,1,3), e_2(1,0,2), e_3(2,1,1), e_4(2,3,0), e_5(3,4,2), e_6(4,3,2)}.

(a) Suponga que se desconoce el valor de clase de cada ejemplo y que se sabe que dichos ejemplos pueden pertenecer sólo a una de dos clases posibles. Utilice entonces el **algoritmo k-medias** para encontrar la clase a la que pertenece cada ejemplo. Para ello, consideré que, en el proceso de selección aleatorio de la semilla en el primer paso del algoritmo, los ejemplos elegidos son e2 y e3.

(b) Suponga ahora que se conoce la clase a la que pertenece cada ejemplo y es la siguiente: e_1(+), e_2(-), e_3(-), e_4(+), e_5(+) y e_6(-). Utilice el **algoritmo k-NN**, con *k*=3, para conocer la clase del ejemplo de clasificación desconocida: e_0(4,3,1; ?). Para ello, use la *clase de la mayoría* como función de asignación de clase de e_0.

b)

E = { e_1(0, 1, 3; +), e_2(1, 0, 2; –), e_3(2, 1, 1; –), e_4(2, 3, 0; +), e_5(3, 4, 2; +), e_6(4, 3, 2; –) }

e_0(4, 3, 1; ?)

Al tener el valor de k = 3, tenemos que comparar e_0 con las 3 instancias que más similitud con él. La función de salida indicada en el enunciado es la de *la clase de la mayoría*.
Comenzamos calculando la similitud de todos los ejemplos de E con e_0.

La función de similitud de dos ejemplos viene dada por $s(e_i, e_j) = \frac{1}{1 + d(e_i, e_j)}$ donde, para ello,

debemos calcular también la distancia euclídea a través de la fórmula: $d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$

Ejemplos	$d(e_0, e_i)$, k= [1..6] → Distancia a e_0(4, 3, 1; ?)	$s(e_0, e_i)$, k= [1..6] → Similitud con e_0(4, 3, 1; ?)
e_1(0, 1, 3; +)	$\sqrt{(0-4)^2+(1-3)^2+(3-1)^2}$ = $\sqrt{16 + 4 + 4}$ = 4.89897	1 / (1 + 4.89897) = 0.16952
e_2(1, 0, 2; –)	$\sqrt{(1-4)^2+(0-3)^2+(2-1)^2}$ = $\sqrt{9 + 9 + 1}$ = 4.35889	1 / (1 + 4.35889) = 0.18660
e_3(2, 1, 1; –)	$\sqrt{(2-4)^2+(1-3)^2+(1-1)^2}$ = $\sqrt{4 + 4 + 0}$ = 2.82842	1 / (1 + 2.82842) = 0.26120
e_4(2, 3, 0; +)	$\sqrt{(2-4)^2+(3-3)^2+(0-1)^2}$ = $\sqrt{4 + 0 + 1}$ = 2.23606	1 / (1 + 2.23606) = 0.30901
e_5(3, 4, 2; +)	$\sqrt{(3-4)^2+(4-3)^2+(2-1)^2}$ = $\sqrt{1 + 1 + 1}$ = 1.73205	1 / (1 + 1.73205) = 0.36602
e_6(4, 3, 2; –)	$\sqrt{(4-4)^2+(3-3)^2+(2-1)^2}$ = $\sqrt{0 + 0 + 1}$ = 1	1 / (1 + 1) = 0.5

Los k = 3 ejemplos con mayor similitud son: e_4(2, 3, 0; +), e_5(3, 4, 2; +) y e_6(4, 3, 2; –), donde hay 2 ejemplos clasificados como + y uno clasificado como –
Por tanto, la instancia se clasificaría como positivo: e_0(4, 3, 1; +)

19SR

PREGUNTA 4. Valoración: 2

Supóngase que se dispone de una colección de ejemplos, *E*, cada uno de ellos denotados por *e(x, y; c)* donde *x, y* representan los dos atributos que permite caracterizarlos y *c* el valor de clase a la que pertenecen. Si E={e_1(1,4; +), e_2(5, 5.5; -), e_3(2, 5; +), e_4(3, 1; -), e_5(3, 9; -)} es el conjunto de ejemplos de clasificación conocida, utilice el **algoritmo k-NN** para clasificar la instancia e_0(3,5; ?). Para realizar la clasificación final, utilice **k=3** y aplique las dos siguientes estrategias:

(a) Usando el criterio de la mayoría.

(b) Ponderando los votos de los vecinos por su similitud.

E = { e_1(1, 4; +), e_2(5, 5.5; –), e_3(2, 5; +), e_4(3, 1; –), e_5(3, 9; –) }
e_0(3, 5; ?)

a)

Al tener el valor de k = 3, tenemos que comparar e_0 con las 3 instancias que más similitud con él. La función de salida indicada en el enunciado es la de *la clase de la mayoría*.

Comenzamos calculando la similitud de todos los ejemplos de E con e_0.

La función de similitud de dos ejemplos viene dada por $s(e_i, e_j) = \frac{1}{1 + d(e_i, e_j)}$ donde, para ello, debemos calcular también la distancia euclídea a través de la fórmula: $d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$

Ejemplos	d(e_0, e_i), k= [1..5] → Distancia a e_0(3, 5; ?)	s(e_0, e_i), k= [1..5] → Similitud con e_0(3, 5; ?)
e_1(1, 4; +)	√((1-3)²+(4-5)²) = √(4 + 1) = 2.23606	1 / (1 + 2.23606) = 0.30901
e_2(5, 5.5; –)	√((5-3)²+(5.5-5)²) = √(4 + 0.25) = 2.06155	1 / (1 + 2.06155) = 0.32037
e_3(2, 5; +)	√((2-3)²+(5-5)²) = √(1 + 0) = 1	1 / (1 + 1) = 0.5
e_4(3, 1; –)	√((3-3)²+(1-5)²) = √(0 + 16) = 4	1 / (1 + 4) = 0.2
e_5(3, 9; –)	√((3-3)²+(9-5)²) = √(0 + 16) = 4	1 / (1 + 4) = 0.2

Los k = 3 ejemplos con mayor similitud son: e_1(1, 4; +), e_2(5, 5.5; –) y e_3(2, 5; +), donde hay 2 ejemplos clasificados como + y uno clasificado como –

Por tanto, la instancia se clasificaría como positivo: e_0(3, 5; +)

b)

Con k = 3 y como función de salida la que *pondera los votos de los k-vecinos por su similitud* tendríamos la clasificación de e_0 de acuerdo al mayor valor de las sumas, para cada valor de clase, de los valores de la similitud de los k = 3 ejemplos que producen clase positiva y los que producen clase negativa:

votos(+) = $\sum$ (similitudes clase +) = similitud(e_0, e_1) + similitud(e_0, e_3) = 0.30901 + 0.5 =0.80901
votos(–) = $\sum$ (similitudes clase –) = similitud(e_0, e_2) = 0.32037

F = arg max (votos(+) , votos(–)) = positivo

Y de ambos valores, se clasifica el ejemplo como e_0(3, 5; F) = e_0(3, 5; +)

14SO – Pregunta 2 y 17SO – Pregunta 2

PREGUNTA 2. Valoración: 4

Aplique el algoritmo ID3 a los datos reflejados en la tabla adjunta, suponiendo que éstos son ejemplos pertenecientes o no a la clase: *Administrar un tratamiento T_0.* Muestre el árbol obtenido y los pasos seguidos para alcanzarlo. Extraiga el conjunto final de reglas que reflejan la descripción del concepto aprendido.

Muy importante: Debe justificar cada paso del algoritmo incluyendo los cálculos numéricos requeridos.

Paciente	Presión Aterial	Urea en sangre	Gota	Hipotiroidismo	Administrar Tratamiento
1	Alta	Alta	Sí	No	No
2	Alta	Alta	Sí	Sí	No
3	Normal	Alta	Sí	No	Sí
4	Baja	Normal	Sí	No	Sí
5	Baja	Baja	No	No	Sí
6	Baja	Baja	No	Sí	No
7	Normal	Baja	No	Sí	Sí
8	Alta	Normal	Sí	No	No
9	Alta	Baja	No	No	Sí
10	Baja	Normal	No	No	Sí
11	Alta	Normal	No	Sí	Sí
12	Normal	Normal	Sí	Sí	Sí
13	Normal	Alta	No	No	Sí
14	Baja	Normal	Sí	Sí	No

Conjunto de atributos posibles y sus posibles valores:

Presión Arterial (A_1) = Alta, Normal, Baja
Urea en Sangre (A_2) = Alta, Normal, Baja
Gota (A_3) = Sí, No
Hipotiroidismo (A_4) = Sí, No

Para determinar qué atributo coger, calculamos la ganancia de información de cada uno de ellos:

$$I(A_1) = \sum_{j=1}^{nv(A_1)} \frac{n_{ij}}{n} I_{ij} = \frac{n_{1\,Alta}}{14}\cdot I_{1\,Alta} + \frac{n_{1\,Normal}}{14}\cdot I_{1\,Normal} + \frac{n_{1\,Baja}}{14}\cdot I_{1\,Baja}$$

$$= \frac{5}{14}\cdot I_{1\,Alta} + \frac{4}{14}\cdot I_{1\,Normal} + \frac{5}{14}\cdot I_{1\,Baja} =$$

$$I_{1\,Alta} = -\sum_{k=1}^{(2)} \frac{n_{1\,Alta\,k}}{n_{1\,Alta}} \log_2\left(\frac{n_{1\,Alta\,k}}{n_{1\,Alta}}\right) = -\frac{n_{1\,Alta\,Sí}}{n_{1\,Alta}}\cdot\log_2\left(\frac{n_{1\,Alta\,Sí}}{n_{1\,Alta}}\right) - \frac{n_{1\,Alta\,No}}{n_{1\,Alta}}\cdot\log_2\left(\frac{n_{1\,Alta\,No}}{n_{1\,Alta}}\right) =$$

$$= -\frac{2}{5}\cdot\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\cdot\log_2\left(\frac{3}{5}\right) = 0{,}97093$$

$$I_{1\,Normal} = -\sum_{k=1}^{(2)} \frac{n_{1\,Normal\,k}}{n_{1\,Normal}} \log_2\left(\frac{n_{1\,Normal\,k}}{n_{1\,Normal}}\right) =$$

$$= -\frac{n_{1\,Normal\,Sí}}{n_{1\,Normal}}\cdot\log_2\left(\frac{n_{1\,Normal\,Sí}}{n_{1\,Normal}}\right) - \frac{n_{1\,Normal\,No}}{n_{1\,Normal}}\cdot\log_2\left(\frac{n_{1\,Normal\,No}}{n_{1\,Normal}}\right) =$$

$$= -\frac{4}{4}\cdot\log_2\left(\frac{4}{4}\right) - \frac{0}{4}\cdot\log_2\cdot\left(\frac{0}{4}\right) = 0$$

$$I_{1\,Baja} = -\sum_{k=1}^{(2)} \frac{n_{1\,Baja\,k}}{n_{1\,Baja}} \log_2\left(\frac{n_{1\,Baja\,k}}{n_{1\,Baja}}\right) = -\frac{n_{1\,Baja\,Sí}}{n_{1\,Baja}}\cdot\log_2\left(\frac{n_{1\,Baja\,Sí}}{n_{1\,Baja}}\right) - \frac{n_{1\,Baja\,No}}{n_{1\,Baja}}\cdot\log_2\left(\frac{n_{1\,Baja\,No}}{n_{1\,Baja}}\right)$$

$$= -\frac{3}{5}\cdot\log_2\left(\frac{3}{5}\right) - \frac{2}{5}\cdot\log_2\left(\frac{2}{5}\right) = 0{,}97093$$

$$\mathbf{I(A_1)} = \frac{5}{14}\, I_{1\,Alta} + \frac{4}{14}\, I_{1\,Normal} + \frac{5}{14}\, I_{1\,Baja} = \mathbf{0{,}69352}$$

$$I(A_2) = \frac{4}{14}\cdot I_{2\,Alta} + \frac{6}{14}\cdot I_{2\,Normal} + \frac{4}{14}\cdot I_{2\,Baja} =$$

$$I_{2\,Alta} = -\frac{2}{4}\cdot\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\cdot\log_2\left(\frac{2}{4}\right) = 1$$

$$I_{2\,Normal} = -\frac{4}{6}\cdot\log_2\left(\frac{4}{6}\right) - \frac{2}{6}\cdot\log_2\left(\frac{2}{6}\right) = 0{,}91829$$

$$I_{2\,Baja} = -\frac{3}{4}\cdot\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\cdot\log_2\left(\frac{1}{4}\right) = 0{,}81127$$

$$\mathbf{I(A_2)} = \frac{4}{14}\, I_{1\,Alta} + \frac{6}{14}\, I_{1\,Normal} + \frac{4}{14}\, I_{1\,Baja} = \mathbf{0{,}91105}$$

$$I(A_3) = \frac{7}{14}\cdot I_{3\,Sí} + \frac{7}{14}\cdot I_{3\,No} =$$

$$I_{3\,Sí} = -\frac{3}{7}\cdot\log_2\left(\frac{3}{7}\right) - \frac{4}{7}\cdot\log_2\left(\frac{4}{7}\right) = 0{,}98523$$

$$I_{3\,No} = -\frac{6}{7}\cdot\log_2\left(\frac{6}{7}\right) - \frac{1}{7}\log_2\left(\frac{1}{7}\right) = 0{,}59168$$

$$\mathbf{I(A_3)} = \frac{7}{14}\, I_{3\,Sí} + \frac{7}{14}\, I_{3\,No} = \mathbf{0{,}78845}$$

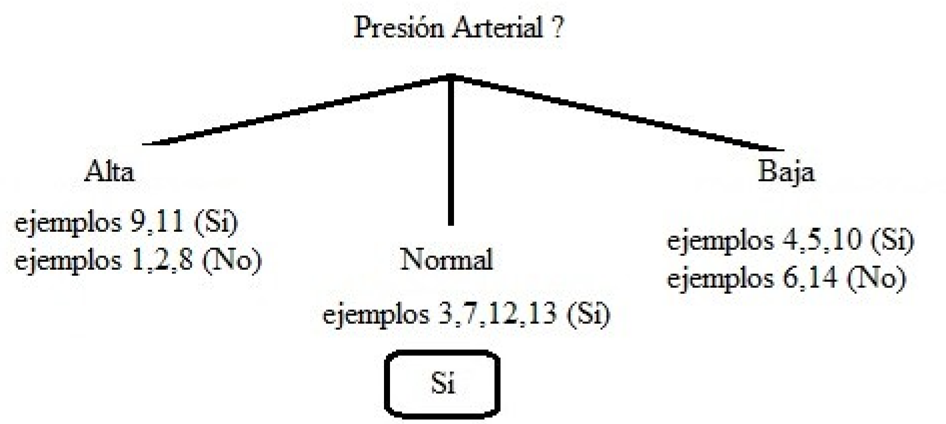
$$I(A_4) = \frac{6}{14}\cdot I_{4\,Sí} + \frac{8}{14}\cdot I_{4\,No} =$$

$$I_{4\,Sí} = -\frac{3}{6}\cdot\log_2\left(\frac{3}{6}\right) - \frac{3}{6}\cdot\log_2\left(\frac{3}{6}\right) = 1$$

$$I_{4\,No} = -\frac{6}{8}\cdot\log_2\left(\frac{6}{8}\right) - \frac{2}{8}\log_2\left(\frac{2}{8}\right) = 0{,}81127$$

$$\mathbf{I(A_4)} = \frac{6}{14}\, I_{4\,Sí} + \frac{8}{14}\, I_{4\,No} = \mathbf{0{,}89215}$$

Se elegiría el atributo de menor $I(A_i)$, que sería Presión Arterial, A_1. Que nos daría el siguiente árbol de decisión.



Y volvemos a calcular las ganancias de información con los atributos restantes (al quitar el Atributo Presión Arterial, A_1) y los ejemplos que tenían A_1 = Alta por un lado (recursividad 2.a) y A_1 = Baja (recursividad 2.b) por otro :

Recursividad 2.a

Paciente	Urea en sangre	Gota	Hipotiroidismo	Administrar Tratamiento
1	Alta	Sí	No	No
2	Alta	Sí	Sí	No
8	Normal	Sí	No	No
9	Baja	No	No	Sí
11	Normal	No	Sí	Sí

$$I(A_2) = \frac{2}{5}\cdot I_{2\,Alta} + \frac{2}{5}\cdot I_{2\,Normal} + \frac{1}{5}\cdot I_{2\,Baja} =$$

$$I_{2\,Alta} = -\frac{0}{2}\cdot\log_2(\frac{0}{2}) - \frac{2}{2}\cdot\log_2(\frac{2}{2}) = 0$$

$$I_{2\,Normal} = -\frac{1}{2}\cdot\log_2(\frac{1}{2}) - \frac{1}{2}\cdot\log_2(\frac{1}{2}) = 1$$

$$I_{2\,Baja} = -\frac{1}{1}\cdot\log_2(\frac{1}{1}) - \frac{0}{1}\cdot\log_2(\frac{0}{1}) = 0$$

$$\mathbf{I(A_2)} = \frac{2}{5}\cdot 0 + \frac{2}{5}\cdot 1 + \frac{1}{5}\cdot 0 = \mathbf{0,4}$$

$$I(A_3) = \frac{3}{5}\cdot I_{3\,Sí} + \frac{2}{5}\cdot I_{3\,No} =$$

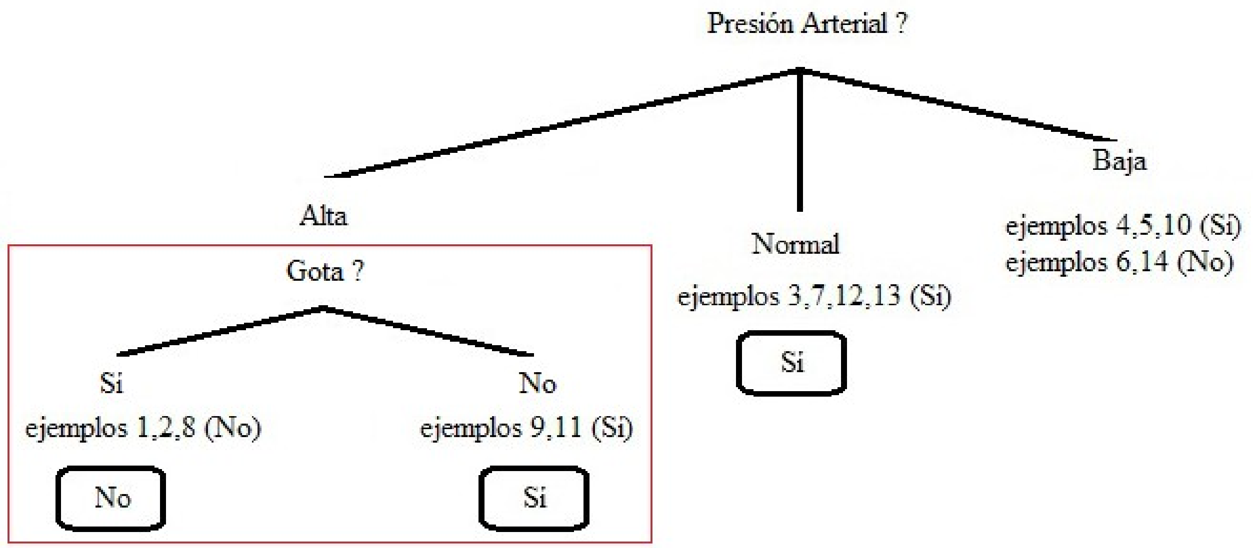
$$I_{3\,Sí} = -\frac{0}{3}\cdot\log_2(\frac{0}{3}) - \frac{3}{3}\cdot\log_2(\frac{3}{3}) = 0$$

$$I_{3\,No} = -\frac{2}{2}\cdot\log_2(\frac{2}{2}) - \frac{0}{2}\cdot\log_2\cdot(\frac{0}{2}) = 0$$

$$\mathbf{I(A_3)} = \frac{3}{5} \cdot 0 + \frac{2}{5} \cdot 0 = \mathbf{0}$$

Se podría seguir calculando $I(A_4)$ pero como ya tenemos a A_3 con la entropía más baja y, por tanto, la ganancia más alta, lo consideramos como el mejor atributo. A_4 a lo más, tendría la misma ganancia de A_3 y cualquiera de los dos nos valdría para expandir el árbol.

Se elegiría el atributo de menor $I(A_i)$, que produce la mayor ganancia, en este caso el de 'Gota', el A_3, que nos daría el siguiente árbol de decisión:



Al llegar al final de esta rama, se continúa con la llamadas recursivas de la otra rama anterior:

Recursividad 2.b

Paciente	Urea en sangre	Gota	Hipotiroidismo	Administrar Tratamiento
4	Normal	Sí	No	Sí
5	Baja	No	No	Sí
6	Baja	No	Sí	No
10	Normal	No	No	Sí
14	Normal	Sí	Sí	No

$$I(A_2) = \frac{0}{5} \cdot I_{2\,Alta} + \frac{3}{5} \cdot I_{2\,Normal} + \frac{2}{5} \cdot I_{2\,Baja} =$$

$$I_{2\,Normal} = -\frac{2}{3} \cdot \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \cdot \log_2 \cdot \left(\frac{1}{3}\right) = 0{,}91829$$

$$I_{2\,Baja} = -\frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) = 1$$

$$\mathbf{I(A_2)} = 0 + \frac{3}{5} \cdot I_{2\,Normal} + \frac{2}{5} \cdot I_{2\,Baja} = \mathbf{0{,}95097}$$

$$I(A_3) = \frac{2}{5} \cdot I_{3\,Si} + \frac{3}{5} \cdot I_{3\,No} =$$

$$I_{3\,Sí} = -\frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) = 1$$

$$I_{3\,No} = -\frac{2}{3} \cdot \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \cdot \log_2\left(\frac{1}{3}\right) = 0{,}95097$$

$$\mathbf{I(A_3)} = \frac{2}{5} \cdot I_{3\,Sí} + \frac{3}{5} \cdot I_{3\,No} = \mathbf{0{,}95097}$$

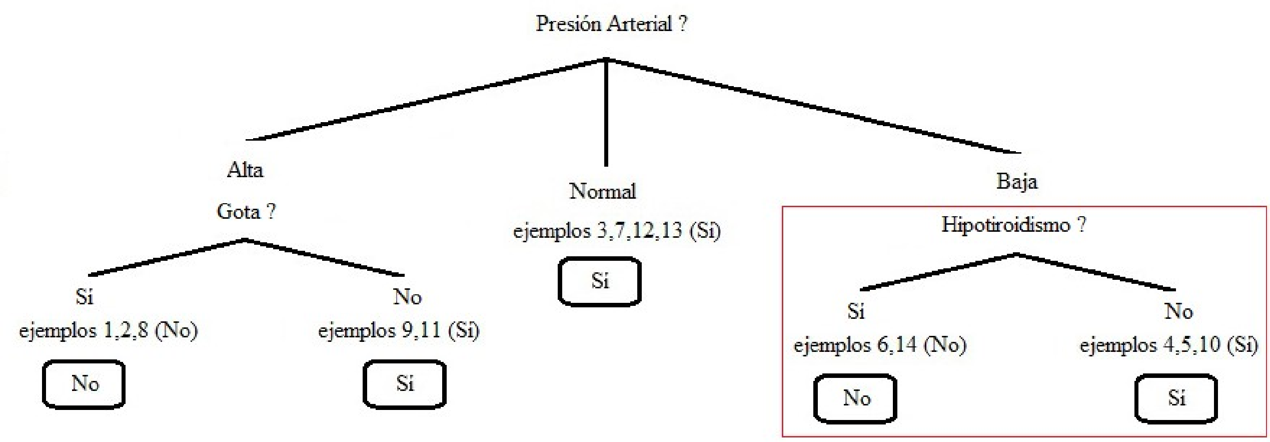
$$I(A_4) = \frac{2}{5} \cdot I_{4\,Si} + \frac{3}{5} \cdot I_{4\,No} =$$

$$I_{4\,Sí} = -\frac{0}{2} \cdot \log_2\left(\frac{0}{2}\right) - \frac{2}{2} \cdot \log_2\left(\frac{2}{2}\right) = 0$$

$$I_{4\,No} = -\frac{3}{3} \cdot \log_2\left(\frac{3}{3}\right) - \frac{0}{3} \log_2\left(\frac{0}{3}\right) = 0$$

$$\mathbf{I(A_4)} = \mathbf{0}$$

Se elegiría el atributo de menor $I(A_i)$, que sería el A_4 , Hipotiroidismo. Que nos daría el siguiente árbol de decisión:



Reglas de descripción del concepto aprendido (examen Septiembre Original de 2017):

- Si Presión Arterial es Alta y Sí se tiene Gota entonces No administrar tratamiento T0 .
- Si Presión Arterial es Alta y No se tiene Gota entonces Sí administrar tratamiento T0.
- Si Presión Arterial es Normal entonces Sí administrar tratamiento T0
- Si Presión Arterial es Baja y se tiene Hipotiroidismo entonces No administrar tratamiento T0.
- Si Presión Arterial es Baja y no se tiene Hipotiroidismo entonces Sí administrar tratamiento T0.

14SR

PREGUNTA 2. Valoración: 2.5

Para la siguiente función booleana, utilice el **algoritmo ID3** para obtener el árbol de decisión correspondiente a su tabla de verdad. Justifique en todo momento la elección de las variables en la construcción del árbol en términos de entropía y de ganancia de información.

f(A,B) = A ∨ (B ∧ C)

ejemplos	**A**	**B**	**C**	B ∧ C	f(A,B) = A ∨ (B ∧ C)
1	V	V	V	V	V
2	V	V	F	F	V
3	V	F	V	F	V
4	V	F	F	F	V
5	F	V	V	V	V
6	F	V	F	F	F
7	F	F	V	F	F
8	F	F	F	F	F

Conjunto de atributos posibles y sus posibles valores:

A = V, F;   B = V, F;   C = V, F

Para determinar qué atributo coger, calculamos la ganancia de información de cada uno de ellos:

$$I(A) = \sum_{j=1}^{nv(A)} \frac{n_{ij}}{n} I_{ij} = \frac{n_{AV}}{8} \cdot I_{AV} + \frac{n_{AF}}{8} \cdot I_{AF} =$$

$$= \frac{4}{8} \cdot I_{AV} + \frac{4}{8} \cdot I_{AF}$$

$$I_{AV} = -\sum_{k=1}^{(2)} \frac{n_{AVk}}{n_{AF}} \log_2\left(\frac{n_{AVk}}{n_{AV}}\right) = -\frac{n_{AVV}}{n_{AV}} \cdot \log_2\left(\frac{n_{AVV}}{n_{AV}}\right) - \frac{n_{AVF}}{n_{AF}} \cdot \log_2\left(\frac{n_{AVF}}{n_{AF}}\right) =$$

$$= -\frac{4}{4} \cdot \log_2\left(\frac{4}{4}\right) - \frac{0}{4} \cdot \log_2\left(\frac{0}{4}\right) = 0$$

$$I_{AF} = -\sum_{k=1}^{(2)} \frac{n_{AFk}}{n_{AF}} \log_2\left(\frac{n_{AFk}}{n_{AF}}\right) =$$

$$= -\frac{n_{AFV}}{n_{AF}} \cdot \log_2\left(\frac{n_{AFV}}{n_{AF}}\right) - \frac{n_{AFF}}{n_{AF}} \cdot \log_2\left(\frac{n_{AFF}}{n_{AF}}\right) =$$

$$= -\frac{1}{4} \cdot \log_2\left(\frac{1}{4}\right) - \frac{3}{4} \cdot \log_2 \cdot \left(\frac{3}{4}\right) = 0{,}81127$$

$$\mathbf{I(A)} = \frac{4}{8} \; I_{AV} + \frac{4}{8} \; I_{AF} = \mathbf{0{,}40563}$$

I(B) $= \frac{4}{8} \cdot I_{BV} + \frac{6}{8} \cdot I_{BF} =$

$$I_{BV} = -\frac{3}{4} \cdot \log_2 \left(\frac{3}{4}\right) - \frac{1}{4} \cdot \log_2 \left(\frac{1}{4}\right) = 0{,}81127$$

$$I_{BF} = -\frac{1}{4} \cdot \log_2 \left(\frac{1}{4}\right) - \frac{2}{4} \cdot \log_2 \left(\frac{2}{4}\right) = 1$$

I(B) $= \frac{4}{8} \quad I_{BV} + \frac{4}{8} \quad I_{BF} =$ **0,90563**

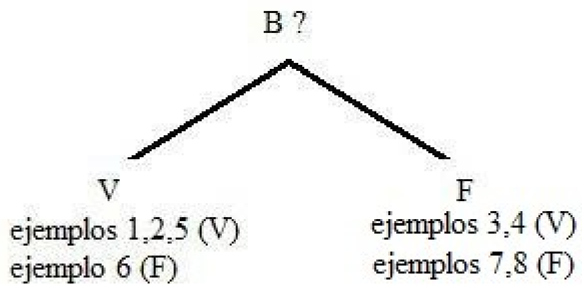
I(C) $= \frac{4}{8} \cdot I_{CV} + \frac{4}{8} \cdot I_{CF} =$

$$I_{CV} = -\frac{3}{4} \cdot \log_2 \left(\frac{3}{4}\right) - \frac{1}{4} \cdot \log_2 \left(\frac{1}{4}\right) = 0.81127$$

$$I_{CF} = -\frac{2}{4} \cdot \log_2 \left(\frac{2}{4}\right) - \frac{2}{4} \log_2 \left(\frac{2}{4}\right) = 1$$

I(C) $= \frac{4}{8} \cdot I_{CV} + \frac{4}{8} \cdot I_{CF} =$ **0,90563**

Se elegiría el atributo de menor I, que sería el B. Que nos daría el siguiente árbol de decisión.



Y volvemos a calcular las ganancias de información con los atributos restantes (al quitar el Atributo B) y los ejemplos que tenían B = V (recursividad 2a) por un lado y B = F por otro (recursividad 2b).

Recursividad 2a

ejemplos	**A**	**C**	B ^ C	f(A,B) = A v (B ^ C)
1	V	V	V	V
2	V	F	F	V
5	F	V	V	V
6	F	F	F	F

I(A) $= \frac{2}{4} \cdot I_{AV} + \frac{2}{4} \cdot I_{AF} =$

$$I_{AV} = -\frac{2}{2} \cdot \log_2 \left(\frac{2}{2}\right) - \frac{0}{2} \cdot \log_2 \left(\frac{0}{2}\right) = 0$$

$$I_{AF} = -\frac{0}{2} \cdot \log_2 \left(\frac{0}{2}\right) - \frac{2}{2} \cdot \log_2 \left(\frac{2}{2}\right) = 0$$

$$\textbf{I(A)} \quad = \quad \frac{2}{4} \quad I_{A\,V} + \frac{2}{4} \quad I_{A\,F} = \mathbf{0}$$

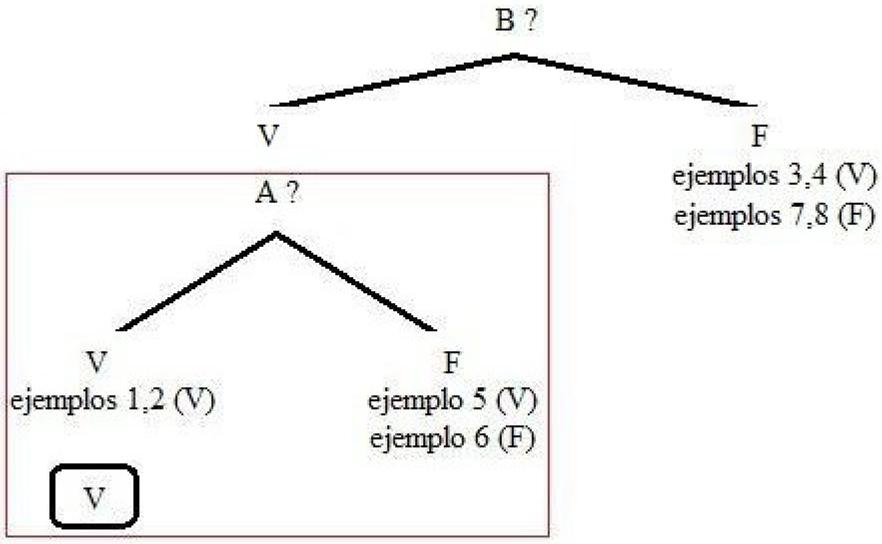
$$\text{I(C)} \quad = \quad \frac{2}{4} \cdot I_{C\,V} + \frac{2}{4} \cdot I_{C\,F} \quad =$$

$$I_{C\,V} \quad = \quad -\frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) \quad = \quad 1$$

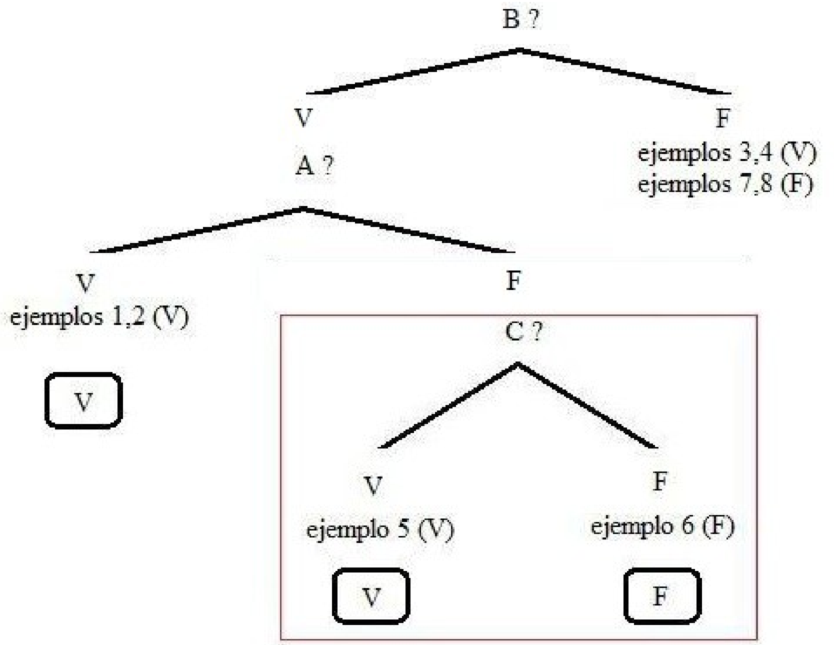
$$I_{C\,F} \quad = \quad -\frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) \quad = 1$$

$$I_{C\,F} \quad = \quad \frac{2}{4} \cdot I_{C\,V} + \frac{2}{4} \cdot I_{C\,F} = \mathbf{1}$$

Se elegiría el atributo de menor I, que sería el A. Que nos daría el siguiente árbol de decisión:



Al quedar solo un atributo, el C, que es el que se elige (no es necesario calcular ganancia de información), el árbol de decisión sería el siguiente:



Al llegar al final de esta rama, se continúa con la llamadas recursivas.

Recursividad 2b:

ejemplos	**A**	**C**	B ^ C	f(A,B) = A v (B ^ C)
3	V	V	F	V
4	V	F	F	V
7	F	V	F	F
8	F	F	F	F

$$I(A) = \frac{2}{4} \cdot I_{AV} + \frac{2}{4} \cdot I_{AF} =$$

$$I_{AV} = -\frac{2}{2} \cdot \log_2\left(\frac{2}{2}\right) - \frac{0}{2} \cdot \log_2\left(\frac{0}{2}\right) = 0$$

$$I_{AF} = -\frac{0}{2} \cdot \log_2\left(\frac{0}{2}\right) - \frac{2}{2} \cdot \log_2\left(\frac{2}{2}\right) = 0$$

$$\mathbf{I(A)} = \frac{2}{4} \quad I_{BV} + \frac{2}{4} \quad I_{BF} = \mathbf{0}$$

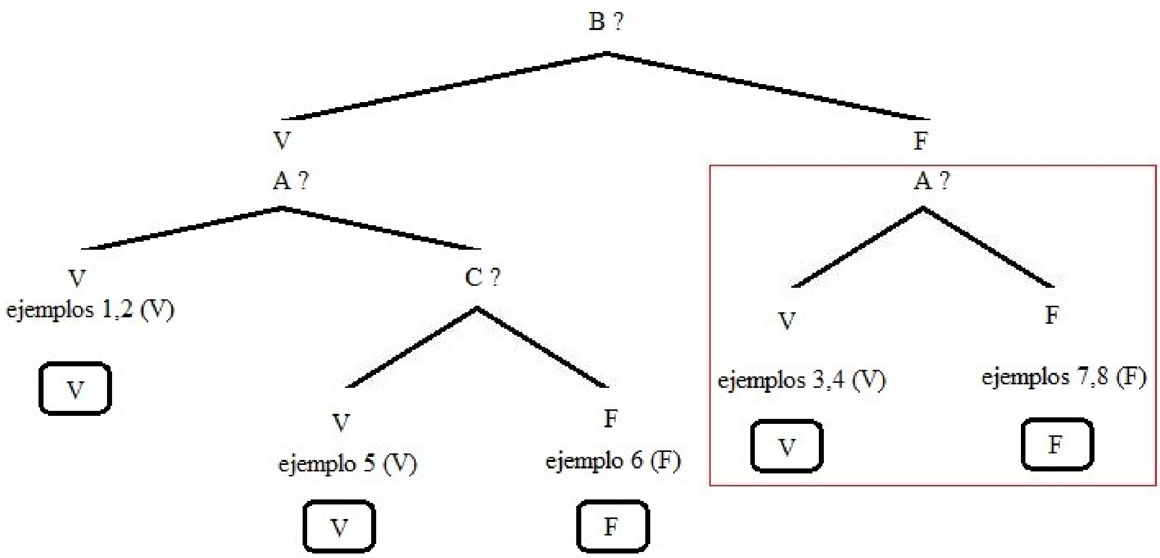
$$I(C) = \frac{2}{4} \cdot I_{CV} + \frac{2}{4} \cdot I_{CF} =$$

$$I_{CV} = -\frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) = 1$$

$$I_{CF} = -\frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) = 1$$

$$\mathbf{I(C)} = \frac{2}{4} \cdot I_{CV} + \frac{2}{4} \cdot I_{CF} = \mathbf{1}$$

Se elegiría el atributo de menor I, que sería el A. Que nos daría el siguiente árbol de decisión:



15SR

PREGUNTA 2. Valoración: 2.5

Para cada una de las siguientes funciones booleanas, utilice al **algoritmo ID3** para obtener el árbol de decisión correspondiente que las representa. Justifique en todo momento la elección de las variables en la construcción del árbol en términos de entropía y de ganancia de información.

(a) F(A,B)= ¬A ∨ B

(b) F(A,B)= A ∧ (¬B ∨ C)

a) F(A,B) = ¬A v B

ejemplos	**A**	¬A	**B**	F(A,B) = ¬A v B
1	V	F	V	V
2	V	F	F	F
3	F	V	V	V
4	F	V	F	V

Conjunto de atributos posibles y sus posibles valores:

A = V, F
B = V, F

$$I(A) = \frac{2}{4} \cdot I_{AV} + \frac{2}{4} \cdot I_{AF} =$$

$$I_{AV} = -\frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) = 1$$

$$I_{AF} = -\frac{2}{2} \cdot \log_2\left(\frac{2}{2}\right) - \frac{0}{2} \cdot \log_2\left(\frac{0}{2}\right) = 0$$

$$\mathbf{I(A)} = \frac{2}{4} \cdot I_{AV} + \frac{2}{4} \cdot I_{AF} = \mathbf{0{,}5}$$

$$I(B) = \frac{2}{4} \cdot I_{BV} + \frac{2}{4} \cdot I_{BF} =$$

$$I_{BV} = -\frac{2}{2} \cdot \log_2\left(\frac{2}{2}\right) - \frac{0}{2} \cdot \log_2\left(\frac{0}{2}\right) = 0$$

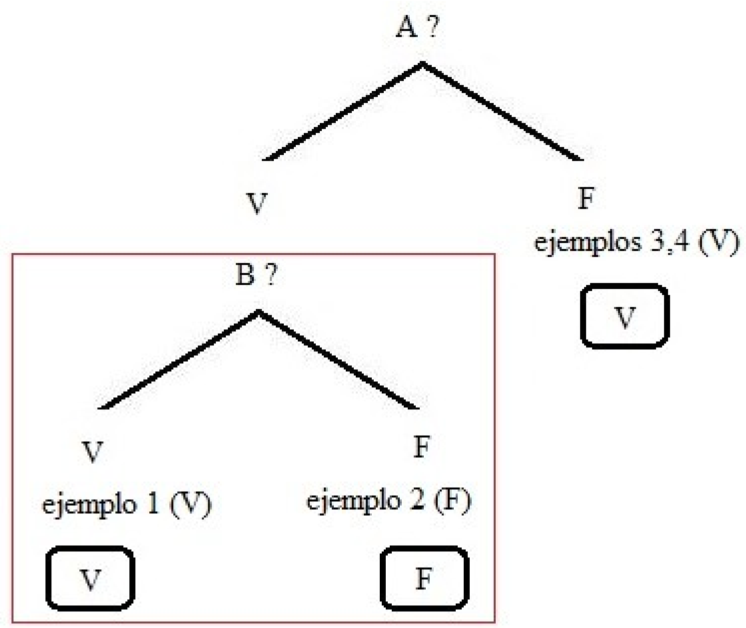
$$I_{BF} = -\frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) = 1$$

$$\mathbf{I(B)} = \frac{2}{4} \cdot I_{CV} + \frac{2}{4} \cdot I_{CF} = \mathbf{0{,}5}$$

Se elegiría el atributo de menor I, pero ambos tienen el mismo valor, por lo que tomaríamos uno caleatoriamente, por ejemplo el A. Que nos daría el siguiente árbol de decisión:

ejemplos	**B**	F(A,B) = ¬A v B
1	V	V
2	F	F

Al quedar solo un atributo, el B, que es el que se elige (no es necesario calcular ganancia de información), el árbol de decisión sería el siguiente:



b) F(A,B) = A ∧ (¬B ∨ C)

ejemplos	**A**	**B**	**C**	¬B	¬B ∨ C	F(A,B) = A ∧ (¬B ∨ C)
1	V	V	V	F	V	V
2	V	V	F	F	F	F
3	V	F	V	V	V	V
4	V	F	F	V	V	V
5	F	V	V	F	V	F
6	F	V	F	F	F	F
7	F	F	V	V	V	F
8	F	F	F	V	V	F

Conjunto de atributos posibles y sus posibles valores:

A = V, F
B = V, F
C = V, F
Para determinar qué atributo coger, calculamos la ganancia de información de cada uno de ellos:

$$I(A) \quad = \quad \sum_{j=1}^{nv(A)} \frac{n_{ij}}{n} I_{ij} \quad = \quad \frac{n_{AV}}{8} \cdot I_{AV} + \frac{n_{AF}}{8} \cdot I_{AF} \quad =$$
$$= \quad \frac{4}{8} \cdot I_{AV} + \frac{4}{8} \cdot I_{AF}$$

$$I_{AV} \quad = \quad -\sum_{k=1}^{(2)} \frac{n_{AVk}}{n_{AF}} \log_2\left(\frac{n_{AVk}}{n_{AV}}\right) \quad = \quad -\frac{n_{AVV}}{n_{AV}} \cdot \log_2\left(\frac{n_{AVV}}{n_{AV}}\right) - \frac{n_{AVF}}{n_{AF}} \cdot \log_2\left(\frac{n_{AVF}}{n_{AF}}\right) \quad =$$
$$= \quad -\frac{3}{4} \cdot \log_2\left(\frac{3}{4}\right) - \frac{1}{4} \cdot \log_2\left(\frac{1}{4}\right) \quad = 0{,}81127$$

$$I_{AF} \quad = \quad -\sum_{k=1}^{(2)} \frac{n_{AFk}}{n_{AF}} \log_2\left(\frac{n_{AFk}}{n_{AF}}\right) \quad =$$
$$= \quad -\frac{n_{AFV}}{n_{AF}} \cdot \log_2\left(\frac{n_{AFV}}{n_{AF}}\right) - \frac{n_{AFF}}{n_{AF}} \cdot \log_2\left(\frac{n_{AFF}}{n_{AF}}\right) \quad =$$
$$= \quad -\frac{0}{4} \cdot \log_2\left(\frac{0}{4}\right) - \frac{4}{4} \cdot \log_2 \cdot \left(\frac{4}{4}\right) \quad = 0$$

I(A) $= \frac{4}{8} \cdot I_{AV} + \frac{4}{8} \cdot I_{AF} =$ **0,40563**

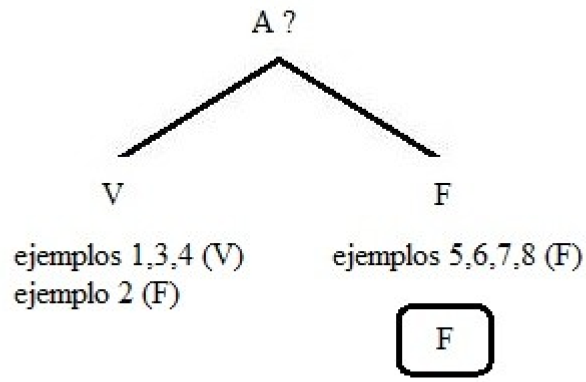
$$I(B) \quad = \quad \frac{4}{8} \cdot I_{BV} + \frac{4}{8} \cdot I_{BF} \quad =$$
$$I_{BV} \quad = \quad -\frac{1}{4} \cdot \log_2\left(\frac{1}{4}\right) - \frac{3}{4} \cdot \log_2\left(\frac{3}{4}\right) \quad = 0{,}81127$$
$$I_{BF} \quad = \quad -\frac{2}{4} \cdot \log_2\left(\frac{2}{4}\right) - \frac{2}{4} \cdot \log_2\left(\frac{2}{4}\right) \quad = 1$$

I(B) $= \frac{4}{8} \cdot I_{BV} + \frac{4}{8} \cdot I_{BF} =$ **0,90563**

$$I(C) \quad = \quad \frac{4}{8} \cdot I_{CV} + \frac{4}{8} \cdot I_{CF} \quad =$$
$$I_{CV} \quad = \quad -\frac{2}{4} \cdot \log_2\left(\frac{2}{4}\right) - \frac{2}{4} \cdot \log_2\left(\frac{2}{4}\right) \quad = 1$$
$$I_{CF} \quad = \quad -\frac{1}{4} \cdot \log_2\left(\frac{1}{4}\right) - \frac{3}{4} \log_2\left(\frac{3}{4}\right) \quad = 0{,}81127$$

I(C) $= \frac{4}{8} \cdot I_{CV} + \frac{4}{8} \cdot I_{CF} =$ **0,90563**

Se elegiría el atributo de menor I, que sería el A. Que nos daría el siguiente árbol de decisión:



Y volvemos a calcular las ganancias de información con los atributos restantes (al quitar el Atributo A) y los ejemplos que tenían A = F.

ejemplos	**B**	**C**	¬B	¬B ∨ C	F(A,B) = A ∧ (¬B ∨ C)
1	V	V	F	V	V
2	V	F	F	F	F
3	F	V	V	V	V
4	F	F	V	V	V

I(B) $= \frac{2}{4}\cdot I_{BV} + \frac{2}{4}\cdot I_{BF} =$

$$I_{BV} = -\frac{1}{2}\cdot\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\cdot\log_2\left(\frac{1}{2}\right) = 1$$

$$I_{BF} = -\frac{2}{2}\cdot\log_2\left(\frac{2}{2}\right) - \frac{0}{2}\cdot\log_2\left(\frac{0}{2}\right) = 0$$

I(B) $= \frac{2}{4}\ I_{BV} + \frac{2}{4}\ I_{BF} =$ **0,5**

I(C) $= \frac{2}{4}\cdot I_{CV} + \frac{2}{4}\cdot I_{CF} =$

$$I_{CV} = -\frac{2}{2}\cdot\log_2\left(\frac{2}{2}\right) - \frac{0}{2}\cdot\log_2\left(\frac{0}{2}\right) = 0$$

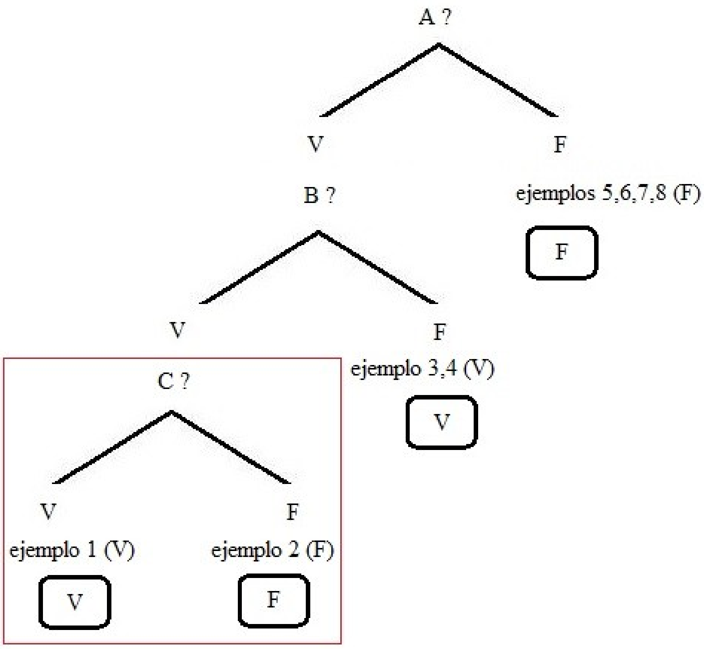
$$I_{CF} = -\frac{1}{2}\cdot\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right) = 1$$

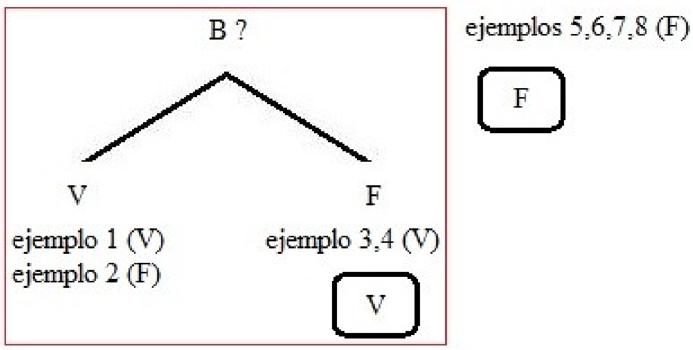
I(C) $= \frac{2}{4}\ I_{CV} + \frac{2}{4}\ I_{CF} =$ **0,5**

Se elegiría el atributo de menor I, pero en este caso tanto B como C tienen el mismo valor de I, por eso elegimos uno aleatoriamente, por ejemplo B. Que nos daría el siguiente árbol de decisión:

Al quedar solo un atributo, el C, que es el que se elige (no es necesario calcular ganancia de información), el árbol de decisión sería el siguiente:

ejemplos	**C**	¬B	¬B ∨ C	F(A,B) = A ∧ (¬B ∨ C)
1	V	F	V	V
2	F	F	F	F





16SO

PREGUNTA 3. Valoración: 3

Para la siguiente función booleana de tres variables, aplique el **algoritmo ID3** para obtener el árbol de decisión asociado a su tabla de verdad. Justifique en todo momento y de forma cuantitativa (no sólo cualitativa) la elección de las variables en la construcción del árbol en términos de entropía y de ganancia de información.

$$F(A,B,C)=(A \vee B) \wedge (A \vee C)$$

ejemplos	**A**	**B**	**C**	A ∨ B	A ∨ C	F(A,B,C) = (A ∨ B) ∧ (A ∨ C)
1	V	V	V	V	V	V
2	V	V	F	V	V	V
3	V	F	V	V	V	V
4	V	F	F	V	V	V
5	F	V	V	V	V	V
6	F	V	F	V	F	F
7	F	F	V	F	V	F
8	F	F	F	F	F	F

Conjunto de atributos posibles y sus posibles valores:

A = V, F
B = V, F
C = V, F

Para determinar qué atributo coger, calculamos la ganancia de información de cada uno de ellos:

$$I(A) = \sum_{j=1}^{nv(A)} \frac{n_{ij}}{n} I_{ij} = \frac{n_{AV}}{8} \cdot I_{AV} + \frac{n_{AF}}{8} \cdot I_{AF} =$$

$$= \frac{4}{8} \cdot I_{AV} + \frac{4}{8} \cdot I_{AF}$$

$$I_{AV} = -\sum_{k=1}^{(2)} \frac{n_{AVk}}{n_{AF}} \log_2 \left(\frac{n_{AVk}}{n_{AV}}\right) = -\frac{n_{AVV}}{n_{AV}} \cdot \log_2 \left(\frac{n_{AVV}}{n_{AV}}\right) - \frac{n_{AVF}}{n_{AF}} \cdot \log_2 \left(\frac{n_{AVF}}{n_{AF}}\right) =$$

$$= -\frac{4}{4} \cdot \log_2 \left(\frac{4}{4}\right) - \frac{0}{4} \cdot \log_2 \left(\frac{0}{4}\right) = 0$$

$$I_{AF} = -\sum_{k=1}^{(2)} \frac{n_{AFk}}{n_{AF}} \log_2 \left(\frac{n_{AFk}}{n_{AF}}\right) =$$

$$= -\frac{n_{AFV}}{n_{AF}} \cdot \log_2 \left(\frac{n_{AFV}}{n_{AF}}\right) - \frac{n_{AFF}}{n_{AF}} \cdot \log_2 \left(\frac{n_{AFF}}{n_{AF}}\right) =$$

$$= -\frac{1}{4} \cdot \log_2 \left(\frac{1}{4}\right) - \frac{3}{4} \cdot \log_2 \cdot \left(\frac{3}{4}\right) = 0.81127$$

I(A) $= \frac{4}{8}$ $I_{A\,V} + \frac{4}{8}$ $I_{A\,F}$ $=$ **0,40563**

I(B) $= \frac{4}{8} \cdot I_{BV} + \frac{4}{8} \cdot I_{BF} =$

$I_{B\,V} = -\frac{3}{4} \cdot \log_2(\frac{3}{4}) - \frac{1}{4} \cdot \log_2(\frac{1}{4}) = 0{,}81127$

$I_{B\,F} = -\frac{2}{4} \cdot \log_2(\frac{2}{4}) - \frac{2}{4} \cdot \log_2(\frac{2}{4}) = 1$

I(B) $= \frac{4}{8}$ $I_{B\,V} + \frac{4}{8}$ $I_{B\,F} =$ **0,90563**

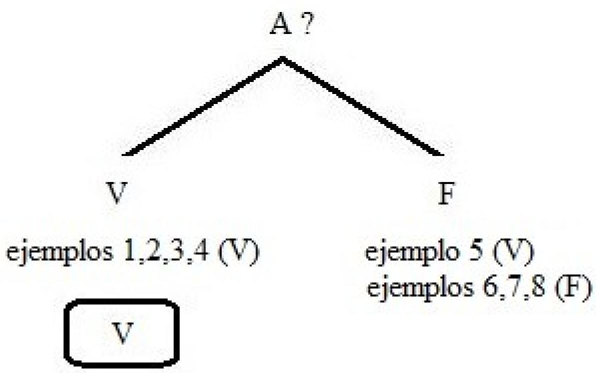
I(C) $= \frac{4}{8} \cdot I_{CV} + \frac{4}{8} \cdot I_{CF} =$

$I_{C\,V} = -\frac{3}{4} \cdot \log_2(\frac{3}{4}) - \frac{1}{4} \cdot \log_2(\frac{1}{4}) = 0{,}81127$

$I_{C\,F} = -\frac{2}{4} \cdot \log_2(\frac{2}{4}) - \frac{2}{4} \log_2(\frac{2}{4}) = 1$

I(C) $= \frac{4}{8}$ $I_{C\,V} + \frac{4}{8}$ $I_{C\,F} =$ **0,90563**

Se elegiría el atributo de menor I, que sería el A. Que nos daría el siguiente árbol de decisión:



Y volvemos a calcular las ganancias de información con los atributos restantes (al quitar el Atributo A) y los ejemplos que tenían A = F.

ejemplos	**B**	**C**	A ∨ B	A ∨ C	F(A,B,C) = (A ∨ B) ∧ (A ∨ C)
5	V	V	V	V	V
6	V	F	V	F	F
7	F	V	F	V	F
8	F	F	F	F	F

$$I(B) \quad = \quad \frac{2}{4} \cdot I_{BV} + \frac{2}{4} \cdot I_{BF} \quad =$$

$$I_{BV} \quad = \quad -\frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) \quad = 1$$

$$I_{BF} \quad = \quad -\frac{0}{2} \cdot \log_2\left(\frac{0}{2}\right) - \frac{2}{2} \cdot \log_2\left(\frac{2}{2}\right) \quad = 0$$

$$\mathbf{I(B)} \quad = \quad \frac{2}{4} \quad I_{BV} + \frac{2}{4} \quad I_{BF} = \mathbf{0{,}5}$$

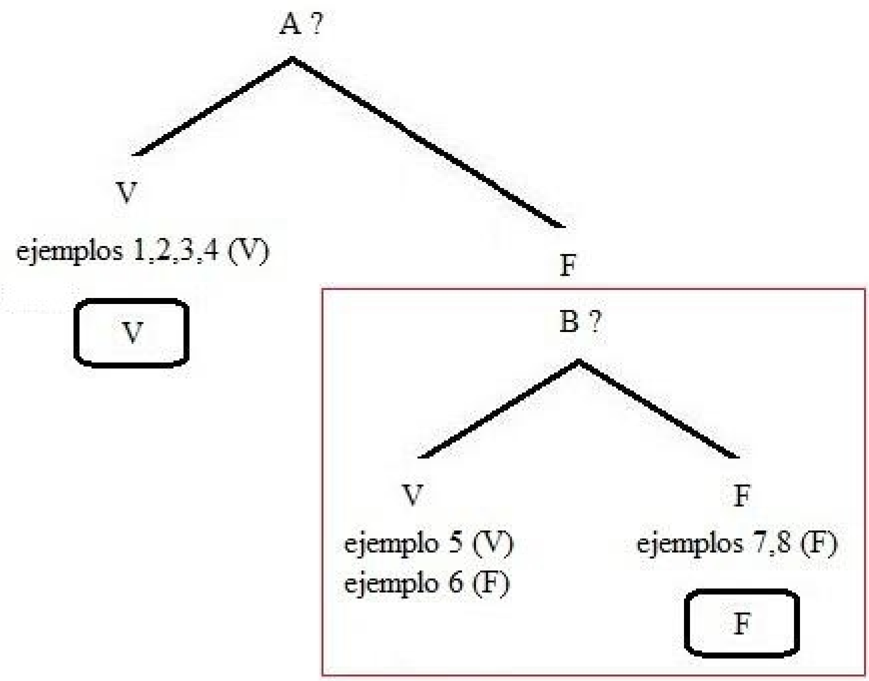
$$I(C) \quad = \quad \frac{2}{4} \cdot I_{CV} + \frac{2}{4} \cdot I_{CF} \quad =$$

$$I_{CV} \quad = \quad -\frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) \quad = 1$$

$$I_{CF} \quad = \quad -\frac{0}{2} \cdot \log_2\left(\frac{0}{2}\right) - \frac{2}{2} \log_2\left(\frac{2}{2}\right) \quad = 0$$

$$\mathbf{I(C)} \quad = \quad \frac{2}{4} \quad I_{CV} + \frac{2}{4} \quad I_{CF} = \mathbf{0{,}90563}$$

Se elegiría el atributo de menor I, en el que tendríamos el mismo valor en B y C. Elegimos aleatoriamente el B, que nos daría el siguiente árbol de decisión:

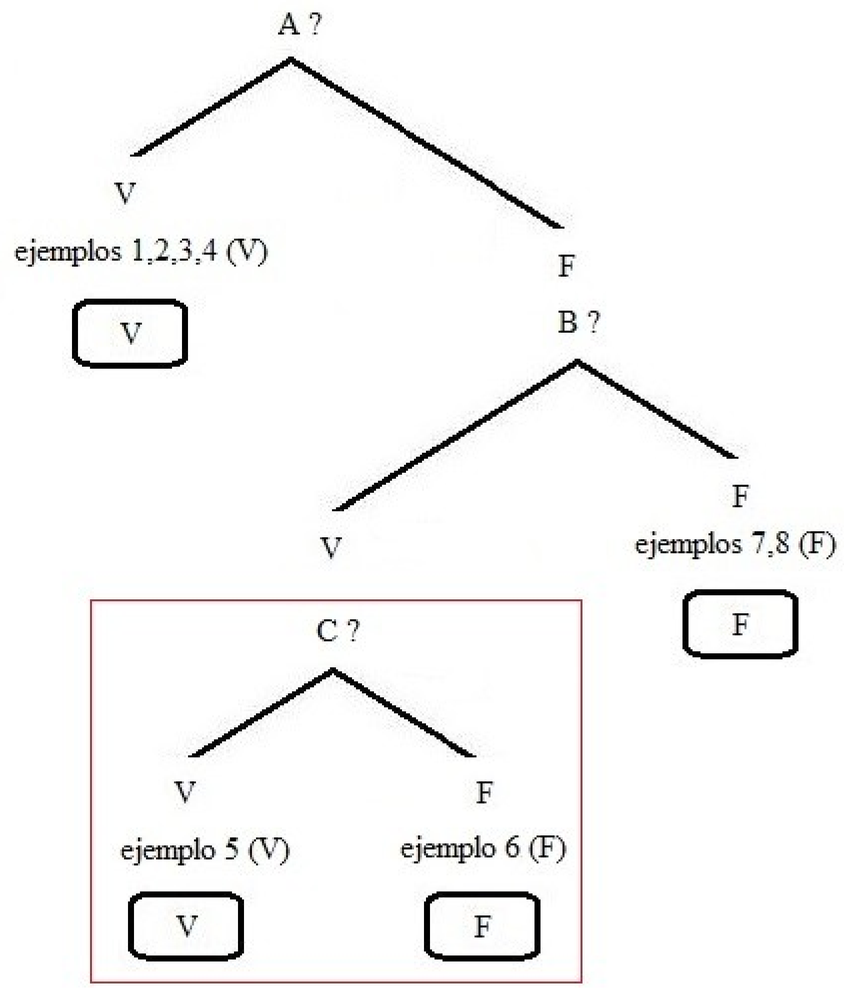


Y volvemos a calcular las ganancias de información con los atributos restantes (al quitar el Atributo B) y los ejemplos que tenían B = V. En este caso solo tendríamos un atributo, por lo que no sería necesario calcular las ganancias.

ejemplos	**C**	A ∨ B	A ∨ C	F(A,B,C) = (A ∨ B) ∧ (A ∨

				C)
5	V	V	V	V
6	F	V	F	F

Con lo que tendríamos el siguiente árbol de decisión:



16 SR y 19SR

PREGUNTA 1. Valoración: 2

Describa el **algoritmo ID3** (árboles de decisión) indicando:

a) Bias.
b) Cómo le afecta el ruido en entradas y en estructura.
c) Su complejidad espacial y temporal.

a) *Bias*.
La bias viene marcada por la entropía/ganancia de información, en la que se prefieren árboles que tengan atributos con más información más cerca de la raíz y que sean más cortos. También favorece por su heurística, atributos con muchos valores.

b) *Cómo le afecta el ruido en entradas y estructura*.
En entradas, el algoritmo es capaz de tratar ejemplos con ruidos debido a que no es necesario que todos los ejemplos pertenezcan a la misma clase para generar un nodo hoja.
En estructura, es tolerante a fallos, ya que si se elimina algún nodo del árbol, todavía se podría clasificar correcamente las instancias restantes.

c) *Su complejidad espacial y temporal*.
Espacio: el ocupado por el árbol de decisión, con el peor caso con un nodo hoja por cada ejemplo. En el mejor caso, cuando todos los ejemplos pertenecen a la misma clase, habrá un único nodo.
Tiempo: crece linealmente con el número de ejemplos de entrenamiento y exponencialmente con el número de atributos.

17F1

PREGUNTA 4. Valoración: 2.5
El estado de una máquina industrial se determina a partir de tres leds indicadores, denominados A, B y C. Teniendo en cuenta que la máquina sufre una parada automática en los casos mostrados en la Tabla 2, construya un **árbol de decisión** que permita predecir la parada de la máquina en función del valor de dichos leds.

Nota: Justifique en todo momento la elección de las variables en la construcción del árbol en términos de entropía y/o ganancia de información.

	A	B	C	Parada
1	0	0	0	No
2	0	0	1	No
3	0	1	0	No
4	0	1	1	No
5	1	0	0	No
6	1	0	1	Sí
7	1	1	0	Sí
8	1	1	1	Sí

Tabla 2

Conjunto de atributos posibles y sus posibles valores:

A = 0,1
B = 0,1
C = 0,1

Para determinar qué atributo coger, calculamos la ganancia de información de cada uno de ellos:

$$I(A) = \sum_{j=1}^{nv(A)} \frac{n_{ij}}{n} I_{ij} = \frac{n_{AV}}{8} \cdot I_{AV} + \frac{n_{AF}}{8} \cdot I_{AF} =$$

$$= \frac{4}{8} \cdot I_{A\,No} + \frac{4}{8} \cdot I_{A\,Sí}$$

$$I_{A\,0} = -\sum_{k=1}^{(2)} \frac{n_{A0k}}{n_{A0}} \log_2\left(\frac{n_{A0k}}{n_{A0}}\right) = -\frac{n_{A0\,No}}{n_{A0}} \cdot \log_2\left(\frac{n_{A0No}}{n_{A0}}\right) - \frac{n_{A0Sí}}{n_{A0}} \cdot \log_2\left(\frac{n_{A0Sí}}{n_{A0}}\right) =$$

$$= -\frac{4}{4} \cdot \log_2\left(\frac{4}{4}\right) - \frac{0}{4} \cdot \log_2\left(\frac{0}{4}\right) = 0$$

$$I_{A\,1} = -\sum_{k=1}^{(2)} \frac{n_{AFk}}{n_{AF}} \log_2\left(\frac{n_{AFk}}{n_{AF}}\right) =$$

$$= -\frac{n_{A1Sí}}{n_{A1}} \cdot \log_2\left(\frac{n_{A1Sí}}{n_{A1}}\right) - \frac{n_{A1\,No}}{n_{A1}} \cdot \log_2\left(\frac{n_{A1\,No}}{n_{A1}}\right) =$$

$$= -\frac{3}{4}\cdot\log_2(\frac{3}{4}) - \frac{1}{4}\cdot\log_2\cdot(\frac{1}{4}) \quad = 0.81127$$

I(A) $= \frac{4}{8}$ $I_{A\,0} + \frac{4}{8}$ $I_{A\,1}$ = **0,40563**

I(B) $= \frac{4}{8}\cdot I_{B0} + \frac{4}{8}\cdot I_{B1} \quad =$

$$I_{B\,0} \quad = -\frac{3}{4}\cdot\log_2(\frac{3}{4}) - \frac{1}{4}\cdot\log_2(\frac{1}{4}) \quad = 0{,}81127$$

$$I_{B\,1} \quad = -\frac{2}{4}\cdot\log_2(\frac{2}{4}) - \frac{2}{4}\cdot\log_2(\frac{2}{4}) \quad = 1$$

I(B) $= \frac{4}{8}$ $I_{B\,0} + \frac{4}{8}$ $I_{B\,1}$ = **0,90563**

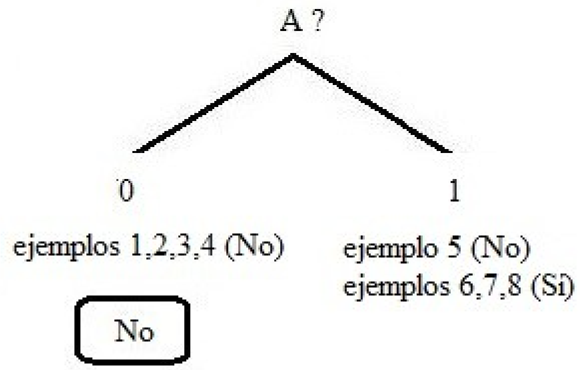
I(C) $= \frac{4}{8}\cdot I_{C0} + \frac{4}{8}\cdot I_{C1} \quad =$

$$I_{C\,0} \quad = -\frac{3}{4}\cdot\log_2(\frac{3}{4}) - \frac{1}{4}\cdot\log_2(\frac{1}{4}) \quad = 0{,}81127$$

$$I_{C\,1} \quad = -\frac{2}{4}\cdot\log_2(\frac{2}{4}) - \frac{2}{4}\log_2(\frac{2}{4}) \quad = 1$$

I(C) $= \frac{4}{8}$ $I_{C\,0} + \frac{4}{8}$ $I_{C\,1}$ = **0,90563**

Se elegiría el atributo de menor I, que sería el A. Que nos daría el siguiente árbol de decisión:



Y volvemos a calcular las ganancias de información con los atributos restantes (al quitar el Atributo A) y los ejemplos que tenían A = 1.

ejemplos	**B**	**C**	Parada
5	0	0	No
6	0	1	Sí
7	1	0	Sí
8	1	1	Sí

$$I(B) = \frac{2}{4} \cdot I_{B0} + \frac{2}{4} \cdot I_{B1} =$$

$$I_{B0} = -\frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) = 1$$

$$I_{B1} = -\frac{0}{2} \cdot \log_2\left(\frac{0}{2}\right) - \frac{2}{2} \cdot \log_2\left(\frac{2}{2}\right) = 0$$

$$\mathbf{I(B)} = \frac{2}{4} \; I_{B0} + \frac{2}{4} \; I_{B1} = \mathbf{0{,}5}$$

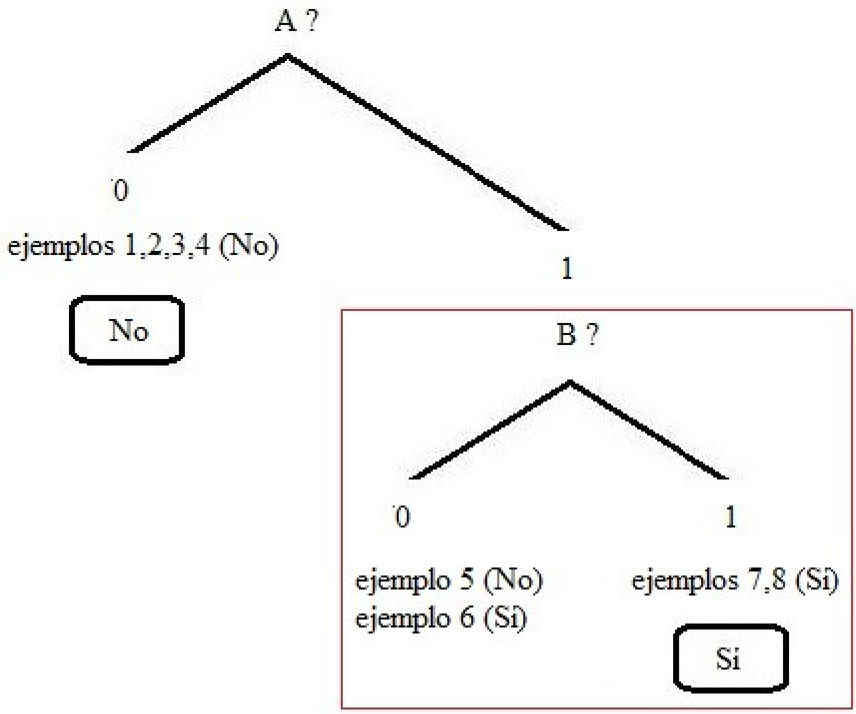
$$I(C) = \frac{2}{4} \cdot I_{C0} + \frac{2}{4} \cdot I_{C1} =$$

$$I_{C0} = -\frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) = 1$$

$$I_{C1} = -\frac{0}{2} \cdot \log_2\left(\frac{0}{2}\right) - \frac{2}{2} \log_2\left(\frac{2}{2}\right) = 0$$

$$\mathbf{I(C)} = \frac{2}{4} \; I_{C0} + \frac{2}{4} \; I_{C1} = \mathbf{0{,}5}$$

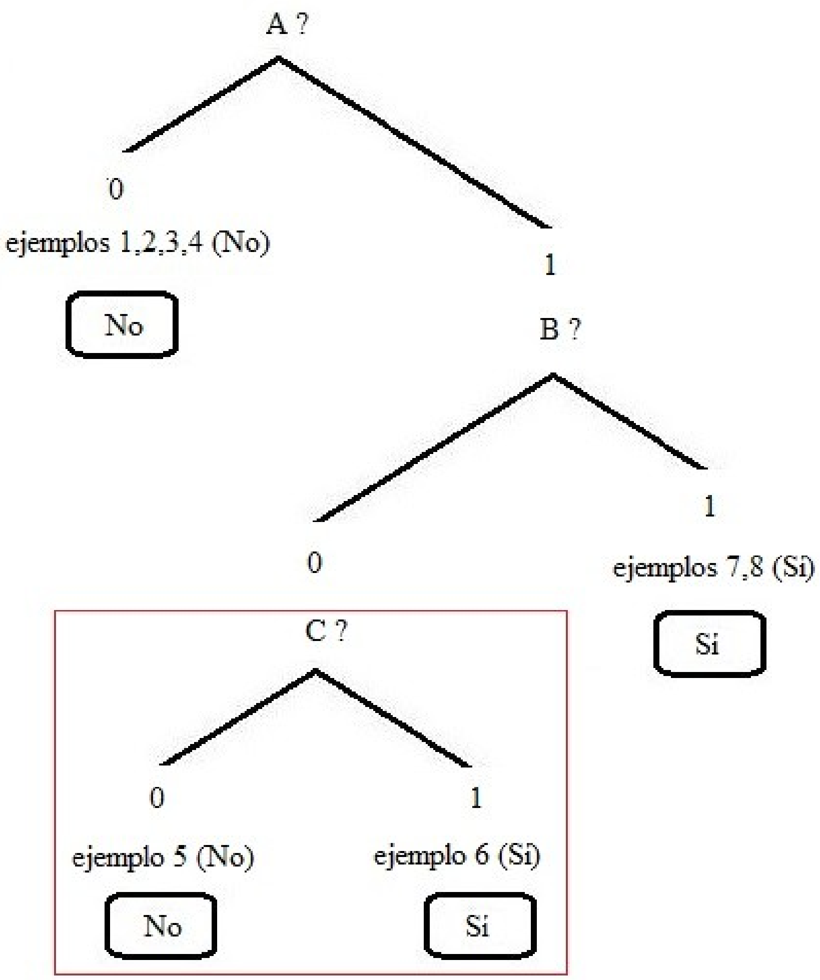
Se elegiría el atributo de menor I, en el que tendríamos el mismo valor en B y C. Elegimos aleatoriamente el B, que nos daría el siguiente árbol de decisión:



Y volvemos a calcular las ganancias de información con los atributos restantes (al quitar el Atributo B) y los ejemplos que tenían B = 0. En este caso solo tendríamos un atributo, por lo que no sería necesario calcular las ganancias.

ejemplos	**C**	Parada
5	0	No
6	1	Sí

Y el árbol de decisión definitivo sería:



17SR

PREGUNTA 2. Valoración: 3

Para la siguiente función booleana, utilice el **algoritmo ID3** para obtener el árbol de decisión correspondiente a su tabla de verdad. Justifique en todo momento la elección de las variables en la construcción del árbol en términos de entropía y de ganancia de información.

$f(A,B) = A \wedge (B \vee C)$

ejemplos	**A**	**B**	**C**	B ∨ C	F(A,B) = A ∧ (B ∨ C)
1	V	V	V	V	V
2	V	V	F	V	V
3	V	F	V	V	V
4	V	F	F	F	F
5	F	V	V	V	F
6	F	V	F	V	F
7	F	F	V	V	F
8	F	F	F	F	F

Conjunto de atributos posibles y sus posibles valores:

A = V, F
B = V, F
C = V, F

Para determinar qué atributo coger, calculamos la ganancia de información de cada uno de ellos:

$$I(A) = \sum_{j=1}^{nv(A)} \frac{n_{ij}}{n} I_{ij} = \frac{n_{AV}}{8} \cdot I_{AV} + \frac{n_{AF}}{8} \cdot I_{AF} =$$

$$= \frac{4}{8} \cdot I_{AV} + \frac{4}{8} \cdot I_{AF}$$

$$I_{AV} = -\sum_{k=1}^{(2)} \frac{n_{AVk}}{n_{AF}} \log_2\left(\frac{n_{AVk}}{n_{AV}}\right) = -\frac{n_{AVV}}{n_{AV}} \cdot \log_2\left(\frac{n_{AVV}}{n_{AV}}\right) - \frac{n_{AVF}}{n_{AF}} \cdot \log_2\left(\frac{n_{AVF}}{n_{AF}}\right) =$$

$$= -\frac{3}{4} \cdot \log_2\left(\frac{3}{4}\right) - \frac{1}{4} \cdot \log_2\left(\frac{1}{4}\right) = 0{,}81127$$

$$I_{AF} = -\sum_{k=1}^{(2)} \frac{n_{AFk}}{n_{AF}} \log_2\left(\frac{n_{AFk}}{n_{AF}}\right) =$$

$$= \ -\frac{n_{AFV}}{n_{AF}} \cdot \log_2\left(\frac{n_{AFV}}{n_{AF}}\right) - \frac{n_{AFF}}{n_{AF}} \cdot \log_2\left(\frac{n_{AFF}}{n_{AF}}\right) \ =$$

$$= \ -\frac{0}{4} \cdot \log_2\left(\frac{0}{4}\right) - \frac{4}{4} \cdot \log_2 \cdot \left(\frac{4}{4}\right) \ = 0$$

I(A) $= \ \frac{4}{8} \ I_{A\,V} + \frac{4}{8} \ I_{A\,F} \ =$ **0,40563**

$$I(B) \ = \ \frac{4}{8} \cdot I_{BV} + \frac{4}{8} \cdot I_{BF} \ =$$

$$I_{B\,V} \ = \ -\frac{2}{4} \cdot \log_2\left(\frac{2}{4}\right) - \frac{2}{4} \cdot \log_2\left(\frac{2}{4}\right) \ = 1$$

$$I_{B\,F} \ = \ -\frac{1}{4} \cdot \log_2\left(\frac{1}{4}\right) - \frac{3}{4} \cdot \log_2\left(\frac{3}{4}\right) \ = 0{,}81127$$

I(B) $= \ \frac{4}{8} \ I_{B\,V} + \frac{4}{8} \ I_{B\,F} =$ **0,90563**

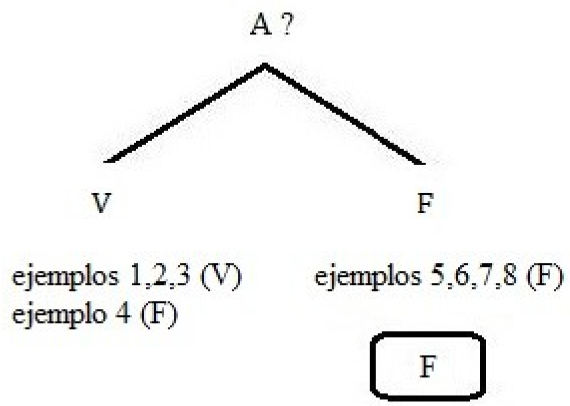
$$I(C) \ = \ \frac{4}{8} \cdot I_{CV} + \frac{4}{8} \cdot I_{CF} \ =$$

$$I_{C\,V} \ = \ -\frac{2}{4} \cdot \log_2\left(\frac{2}{4}\right) - \frac{2}{4} \cdot \log_2\left(\frac{2}{4}\right) \ = \ 1$$

$$I_{C\,F} \ = \ -\frac{1}{4} \cdot \log_2\left(\frac{1}{4}\right) - \frac{3}{4} \log_2\left(\frac{3}{4}\right) \ = 0{,}81127$$

I(C) $= \ \frac{4}{8} \ I_{C\,V} + \frac{4}{8} \ I_{C\,F} =$ **0,92563**

Se elegiría el atributo de menor I, que sería el A. Que nos daría el siguiente árbol de decisión:



Y volvemos a calcular las ganancias de información con los atributos restantes (al quitar el Atributo A) y los ejemplos que tenían A = V.

ejemplos	**B**	**C**	B ∨ C	F(A,B) = A ∧ (B ∨ C)

1	V	V	V	V
2	V	F	V	V
3	F	V	V	V
4	F	F	F	F

I(B) $= \frac{2}{4} \cdot I_{BV} + \frac{2}{4} \cdot I_{BF} =$

$I_{BV} = -\frac{2}{2} \cdot \log_2\left(\frac{2}{2}\right) - \frac{0}{2} \cdot \log_2\left(\frac{0}{2}\right) = 0$

$I_{BF} = -\frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) = 1$

I(B) $= \frac{2}{4}$ $I_{BV} + \frac{2}{4}$ $I_{BF} =$ **0,5**

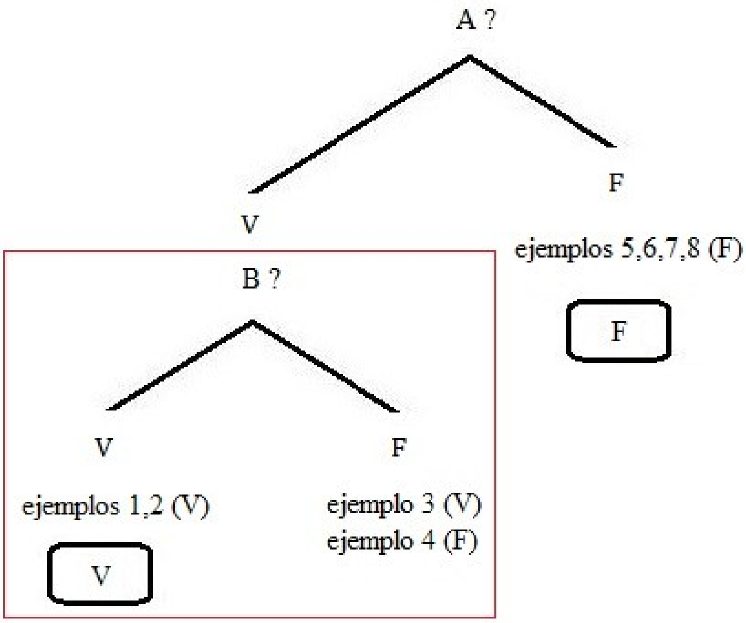
I(C) $= \frac{2}{4} \cdot I_{CV} + \frac{2}{4} \cdot I_{CF} =$

$I_{CV} = -\frac{2}{2} \cdot \log_2\left(\frac{2}{2}\right) - \frac{0}{2} \cdot \log_2\left(\frac{0}{2}\right) = 0$

$I_{CF} = -\frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) = 1$

I(C) $= \frac{2}{4}$ $I_{CV} + \frac{2}{4}$ $I_{CF} =$ **0,5**

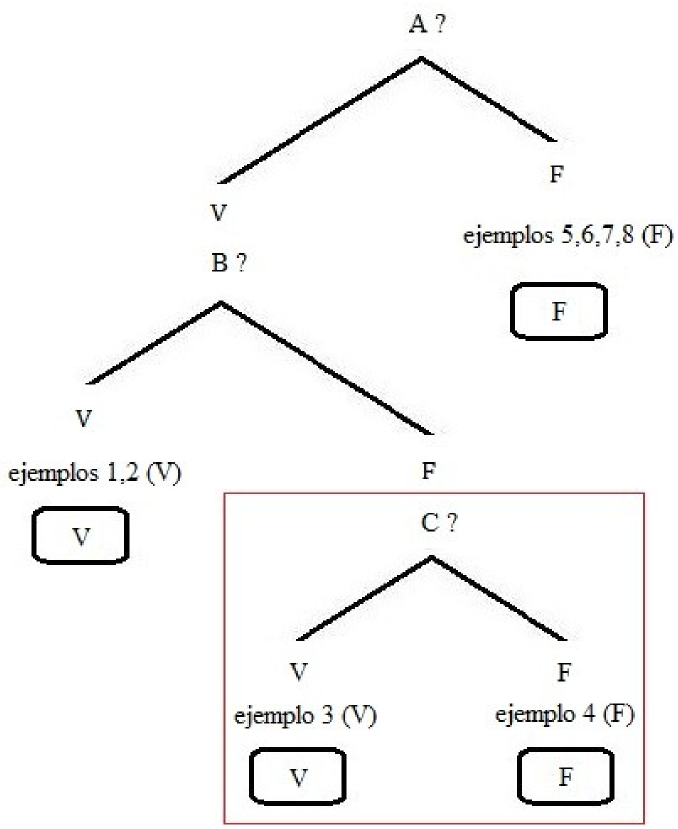
Se elegiría el atributo de menor I, en el que tendríamos el mismo valor en B y C. Elegimos aleatoriamente el B, que nos daría el siguiente árbol de decisión:



Y volvemos a calcular las ganancias de información con los atributos restantes (al quitar el Atributo B) y los ejemplos que tenían B = F. En este caso solo tendríamos un atributo, por lo que no sería necesario calcular las ganancias.

ejemplos	**C**	B ∨ C	F(A,B) = A ∧ (B ∨ C)
3	V	V	V
4	F	F	F

Con lo que tendríamos el siguiente árbol de decisión:



18SR

PREGUNTA 2. Valoración: 2.5

Para cada una de las siguientes funciones booleanas, utilice al **algoritmo ID3** para obtener el árbol de decisión correspondiente que las representa. Justifique en todo momento la elección de las variables en la construcción del árbol en términos de entropía y de ganancia de información.

(a) $F(A,B)= A \vee \neg B$
(b) $F(A,B)= A \wedge (B \vee \neg C)$

donde $\vee$, $\wedge$ y $\neg$ denotan, respectivamente, los operadores OR, AND y NOT.

a) F(A,B) = A v ¬B

ejemplos	**A**	**B**	**¬B**	F(A,B) = ¬A v B
1	V	V	F	V
2	V	F	V	V
3	F	V	F	F
4	F	F	V	V

Conjunto de atributos posibles y sus posibles valores:

A = V, F
B = V, F

$$I(A) = \frac{2}{4} \cdot I_{AV} + \frac{2}{4} \cdot I_{AF} =$$

$$I_{AV} = -\frac{2}{2} \cdot \log_2\left(\frac{2}{2}\right) - \frac{0}{2} \cdot \log_2\left(\frac{0}{2}\right) = 0$$

$$I_{AF} = -\frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) = 1$$

$$\mathbf{I(A)} = \frac{2}{4} \cdot I_{BV} + \frac{2}{4} \cdot I_{BF} = \mathbf{0{,}5}$$

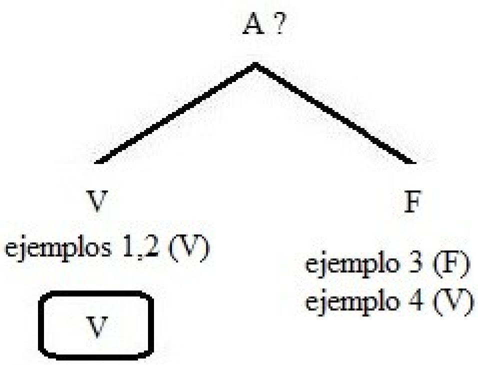
$$I(B) = \frac{2}{4} \cdot I_{BV} + \frac{2}{4} \cdot I_{BF} =$$

$$I_{BV} = -\frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) = 1$$

$$I_{BF} = -\frac{2}{2} \cdot \log_2\left(\frac{2}{2}\right) - \frac{0}{2} \log_2\left(\frac{0}{2}\right) = 0$$

$$\mathbf{I(B)} = \frac{2}{4} \cdot I_{BV} + \frac{2}{4} \cdot I_{BF} = \mathbf{0{,}5}$$

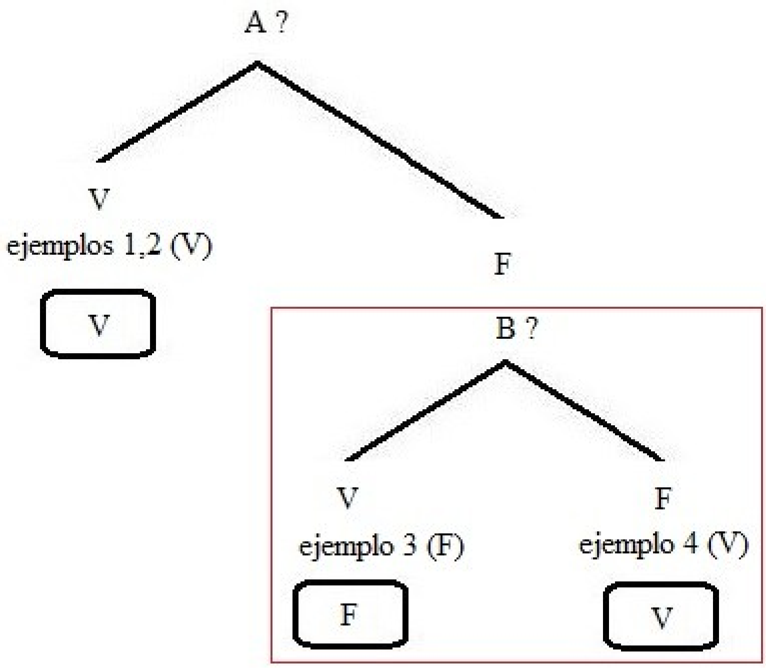
Se elegiría el atributo de menor I, pero ambos tienen el mismo valor, por lo que tomaríamos uno caleatoriamente, por ejemplo el A. Que nos daría el siguiente árbol de decisión:



Y volvemos a calcular las ganancias de información con los atributos restantes (al quitar el Atributo A) y los ejemplos que tenían A = F. En este caso solo tendríamos un atributo, por lo que no sería necesario calcular las ganancias.

ejemplos	**B**	**¬B**	F(A,B) = ¬A v B
3	V	F	F
4	F	V	V

Y el árbol de decisión correspondiente sería:



b) F(A,B) = A ∧ (B ∨ ¬C)

ejemplos	**A**	**B**	**C**	¬C	B ∨ ¬C	F(A,B) = A ∧ (B ∨ ¬C)
1	V	V	V	F	V	V
2	V	V	F	V	V	V
3	V	F	V	F	F	F
4	V	F	F	V	V	V
5	F	V	V	F	V	F
6	F	V	F	V	V	F
7	F	F	V	F	F	F
8	F	F	F	V	V	F

Conjunto de atributos posibles y sus posibles valores:

A = V, F;   B = V, F;   C = V, F

Para determinar qué atributo coger, calculamos la ganancia de información de cada uno de ellos:

$$I(A) = \sum_{j=1}^{nv(A)} \frac{n_{ij}}{n} I_{ij} = \frac{n_{AV}}{8} \cdot I_{AV} + \frac{n_{AF}}{8} \cdot I_{AF} =$$
$$= \frac{4}{8} \cdot I_{AV} + \frac{4}{8} \cdot I_{AF}$$

$$I_{AV} = -\sum_{k=1}^{(2)} \frac{n_{AVk}}{n_{AF}} \log_2(\frac{n_{AVk}}{n_{AV}}) = -\frac{n_{AVV}}{n_{AV}} \cdot \log_2(\frac{n_{AVV}}{n_{AV}}) - \frac{n_{AVF}}{n_{AF}} \cdot \log_2(\frac{n_{AVF}}{n_{AF}}) =$$
$$= -\frac{3}{4} \cdot \log_2(\frac{3}{4}) - \frac{1}{4} \cdot \log_2(\frac{1}{4}) = 0{,}81127$$

$$I_{AF} = -\sum_{k=1}^{(2)} \frac{n_{AFk}}{n_{AF}} \log_2(\frac{n_{AFk}}{n_{AF}}) =$$
$$= -\frac{n_{AFV}}{n_{AF}} \cdot \log_2(\frac{n_{AFV}}{n_{AF}}) - \frac{n_{AFF}}{n_{AF}} \cdot \log_2(\frac{n_{AFF}}{n_{AF}}) =$$
$$= -\frac{0}{4} \cdot \log_2(\frac{0}{4}) - \frac{4}{4} \cdot \log_2 \cdot (\frac{4}{4}) = 0$$

$$\mathbf{I(A)} = \frac{4}{8} \cdot I_{AV} + \frac{4}{8} \cdot I_{AF} = \mathbf{0{,}40563}$$

$$I(B) = \frac{4}{8} \cdot I_{BV} + \frac{4}{8} \cdot I_{BF} =$$

$$I_{BV} = -\frac{2}{4} \cdot \log_2(\frac{2}{4}) - \frac{2}{4} \cdot \log_2(\frac{2}{4}) = 1$$

$$I_{BF} = -\frac{1}{4} \cdot \log_2(\frac{1}{4}) - \frac{3}{4} \cdot \log_2(\frac{3}{4}) = 0{,}81127$$

$$\mathbf{I(B)} = \frac{4}{8} \cdot I_{BV} + \frac{4}{8} \cdot I_{BF} = \mathbf{0{,}90563}$$

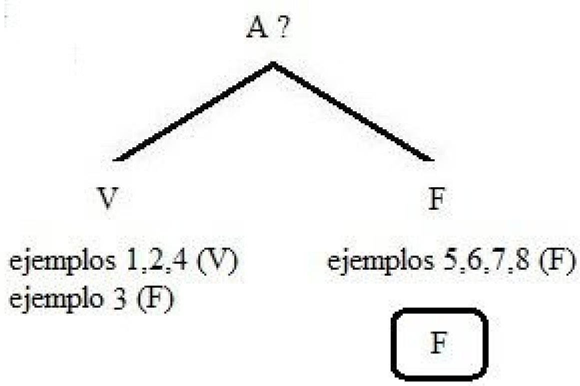
$$I(C) = \frac{4}{8} \cdot I_{CV} + \frac{4}{8} \cdot I_{CF} =$$

$$I_{CV} = -\frac{1}{4}\cdot\log_2\left(\frac{1}{4}\right) - \frac{3}{4}\log_2\left(\frac{3}{4}\right) = 0{,}81127$$

$$I_{CF} = -\frac{2}{4}\cdot\log_2\left(\frac{2}{4}\right) - \frac{2}{4}\cdot\log_2\left(\frac{2}{4}\right) = 1$$

$$\mathbf{I(C)} = \frac{4}{8}\cdot I_{CV} + \frac{4}{8}\cdot I_{CF} = \mathbf{0{,}90563}$$

Se elegiría el atributo de menor I, que sería el A. Que nos daría el siguiente árbol de decisión:



Y volvemos a calcular las ganancias de información con los atributos restantes (al quitar el Atributo A) y los ejemplos que tenían A = V.

ejemplos	**B**	**C**	¬C	B ∨ ¬C	F(A,B) = A ∧ (B ∨ ¬C)
1	V	V	F	V	V
2	V	F	V	V	V
3	F	V	F	F	F
4	F	F	V	V	V

$$I(B) = \frac{2}{4}\cdot I_{BV} + \frac{2}{4}\cdot I_{BF} =$$

$$I_{BV} = -\frac{2}{2}\cdot\log_2\left(\frac{2}{2}\right) - \frac{0}{2}\cdot\log_2\left(\frac{0}{2}\right) = 0$$

$$I_{BF} = -\frac{1}{2}\cdot\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\cdot\log_2\left(\frac{1}{2}\right) = 1$$

$$\mathbf{I(B)} = \frac{2}{4}\ I_{BV} + \frac{2}{4}\ I_{BF} = \mathbf{0{,}5}$$

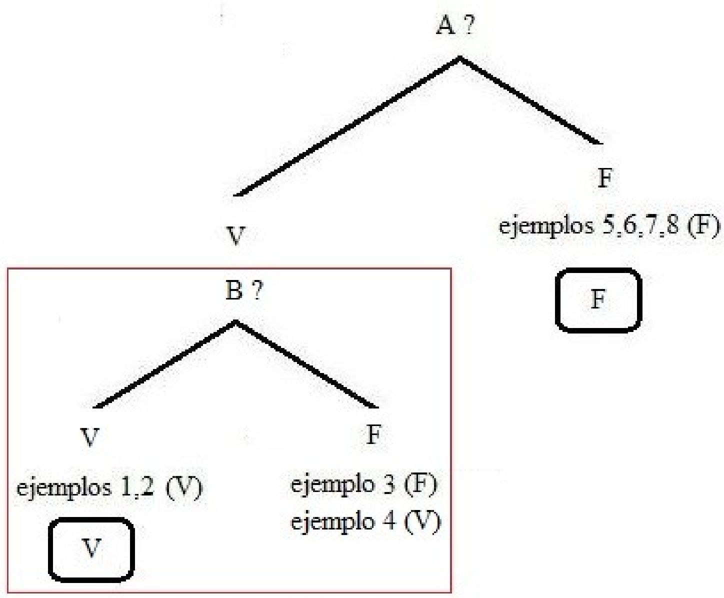
$$I(C) = \frac{2}{4}\cdot I_{CV} + \frac{2}{4}\cdot I_{CF} =$$

$$I_{CV} = -\frac{1}{2}\cdot\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\cdot\log_2\left(\frac{1}{2}\right) = 1$$

$$I_{CF} = -\frac{2}{2}\cdot\log_2\left(\frac{2}{2}\right) - \frac{0}{2}\log_2\left(\frac{0}{2}\right) = 0$$

$$\mathbf{I(C)} = \frac{2}{4}\ I_{C\,V} + \frac{2}{4}\ I_{C\,F} = \mathbf{0{,}5}$$

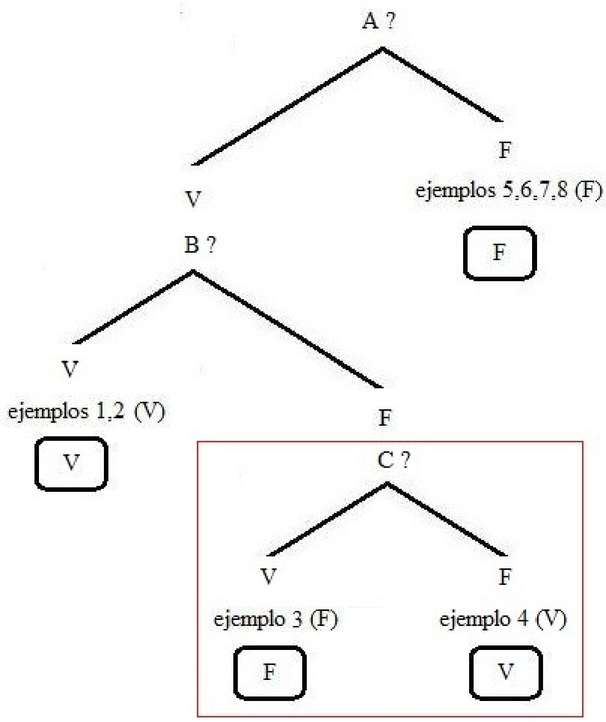
Se elegiría el atributo de menor I, pero en este caso tanto B como C tienen el mismo valor de I, por eso elegimos uno aleatoriamente, por ejemplo B. Que nos daría el siguiente árbol de decisión:



Y volvemos a calcular las ganancias de información con los atributos restantes (al quitar el Atributo B) y los ejemplos que tenían B = F.

ejemplos	**C**	¬C	B ∨ ¬C	F(A,B) = A ∧ (B ∨ ¬C)
3	V	F	F	F
4	F	V	V	V

Al quedar solo un atributo, el C, que es el que se elige (no es necesario calcular ganancia de información), el árbol de decisión sería el siguiente:


A ?
V
F
ejemplos 5,6,7,8 (F)
F
B ?
V
ejemplos 1,2 (V)
V
F
C ?
V
F
ejemplo 3 (F)
ejemplo 4 (V)
F
V

14F2

PREGUNTA 1. Valoración: 2

Describa el **algoritmo M5** (árboles de regresión) indicando:

a) Bias.
b) Cuáles son sus entradas y salidas. Así como aspectos relacionados con su representación, preprocesamiento y tratamiento.
c) Cómo le afecta el ruido: en entradas y en estructura.
d) Su complejidad espacial y temporal.
e) Control de la tarea aprendida (crítica/valoración/utilización).
f) Dependencia del conocimiento del dominio.

a) *Bias*:
Se utiliza el cálculo del menor error asociado a los atributos, para elegir cuál es el más apropiado. También se produce un bias de lenguaje , al representar modelos lineales solamente.

b) *Cuáles son sus entradas y salidas. Así como aspectos relacionados con su representación, preprocesamiento y tratamiento*.
Como entradas se tienen los ejemplos de la muestra, caracterizado por una serie de atributos, uno de ellos, numérico, identificado como el concepto clase.
La salida sería el árbol de regresión, que tiene en sus nodo intermedios una pregunta sobre los atributos y los nodos hojas con un modelo lineal para calcular la clase numerica.
No tiene tratamiento de sus entradas ni tampoco preprocesamiento.

c) *Cómo le afecta el ruido: en entradas y en estructura*.
Acepta entradas con ruido, suavizados por el análisis de errores y los modelos lineales usados para calcular la clase en los nodos hoja.
Es tolerante a fallos debido a que dispone de dos tipos de modelos en cada nodo intermedio (lineal y basado en árbol) que hace que si se elimina algún nodo o modelo de algún nodo, se pueda seguir clasificando.

d) *Complejidad espacial y temporal*.
El espacio es el ocupado por el árbol de regresión. En el peor de los casos con todos los atributos por rama, en el mejor de los casos quedría reducido al nodo raíz con un modelo lineal para todos los ejemplos y en caso intermedio, se tendría un número de nodos igual a $\sum_{j=o}^{A} i^{j}$
El tiempo depende del número de atributos y del número de ejemplos.

e) *Control de la tarea aprendida (crítica/valoración/utilización)*.
La valoración interna de lo aprendido se realiza a través del cálculo del error, que se puede hacer con un conjuntoexterno de instancias.
La forma de utilizar lo aprendido es a través del procedimiento *clasificar*.

f) *Dependencia del conocimiento del dominio*.
Solamente utiliza la definición de los atributos y sus posibles valores.

14SO – Pregunta 3

PREGUNTA 3. Valoración: 3

Una forma relativamente sencilla de construir **árboles de regresión** es sustituir el modelo lineal de cada uno de sus nodos hojas por el promediado del valor del atributo de clase de los casos pertenecientes a dicha hoja. Es decir, el árbol de regresión resultante estaría formado por un árbol en el que el valor de predicción de cada uno de sus nodos hoja correspondería a una constante. Utilice este tipo de árboles para resolver el siguiente problema.

Supongamos que una entidad bancaria quiere predecir el saldo que cada uno de sus clientes tendrá el año siguiente en función de los datos que conoce sobre sus clientes y, para ello, dispone de la información que se muestra en la tabla adjunta. ¿Cuál sería el modelo predictivo resultante? Prescinda de las etapas de poda en la construcción del árbol.

Existen dos condiciones para seguir dividiendo el árbol en un nodo:

1. Número de ejemplos en un nodo $n \geq 2$.
2. Desviación estándar de los ejemplos en un nodo $\sigma > 100$.

Muy importante: Muestre y justifique todos los pasos y resultados numéricos intermedios necesarios para la obtención del árbol.

Ejemplo	Sueldo A1	Hipoteca A2	Casado A3	Hijos A4	Saldo (Clase)
1	Medio	Media	Sí	Sí	2300
2	Medio	Alta	Sí	No	2500
3	Alto	Alta	No	Sí	4000
4	Bajo	Media	Sí	Sí	1000
5	Bajo	Baja	No	No	500
6	Medio	Baja	No	No	2400
7	Alto	Media	Sí	Sí	4100
8	Alto	Media	Sí	No	1100

La desviación estándar viene dada por $\sigma = \sqrt{\sum_i^n \frac{(x_i - \mu)^2}{n}}$ siendo μ el valor medio

Para construir un árbol de regresión en base al conjunto de los 8 ejemplos de la muestra y los 4 atributos y un árbol de regresión inicial vacío tengo que calcular el mejor atributo de dicho conjunto de ejemplos (de 1 a 8) y atributos (de A1 a A4).

El mejor atributo es aquel que proporciona el menor error ponderado, que viene dado por la fórmula $\text{Error}_{(A_i)} = \sum \left(\frac{|E_i|}{|E|}\sigma_{(A_i)}\right)$. Para calcular las desviaciones estándar ($\sigma_{(A_i)}$) hay que calcular también los valores medios μ_i de cada atributo:

$$\mu_i = \frac{\sum \textit{valores de clase de los ejemplo que tengan un valor de atributo i determinado}}{n^o \textit{ de ejemplos con un valor de atributo i determinado}}$$

$$\mu_{(A1 = Alto)} = \frac{4000 + 4100 + 1100}{3} = 3066,66$$

$\mu_{(A1 = Medio)} \quad = \quad \dfrac{2300+2500+2400}{3} = 2400$

$\mu_{(A1 = Bajo)} \quad = \quad \dfrac{1000+500}{2} = 750$

Cálculo de la desviación estándar del atributo:

$\sigma_{(Ai)} = \sqrt{\sum \dfrac{(x_i - \mu_{Ai})^2}{n}}$ = (en el libro de texto hay ejemplo que no divide entre 'n', ver explicación en fe de erratas)

$$\sigma_{(A1 = Alto)} = \sqrt{\frac{((4000-3066{,}66)^2+(4100-3066{,}66)^2+(1100-3066{,}66)^2)}{3}} = 1391{,}24$$

$$\sigma_{(A1 = Medio)} = \sqrt{\frac{((2300-2400)^2+(2500-2400)^2+(2400-2400)^2)}{3}} = 81{,}64$$

$$\sigma_{(A1 = Bajo)} = \sqrt{\frac{((1000-750)^2+(500-750)^2)}{2}} = 250$$

Cálculo del error ponderado de A1:

$$\text{Error}_{(A1)} = \sum \left(\frac{|E_{A1}|}{|E|}\sigma_{(A1)}\right)$$

$$= \frac{|E_{A1=Alto}|}{|E|}\sigma_{(A1=Alto)} + \frac{|E_{A1=Medio}|}{|E|}\sigma_{(A1=Medio)} + \frac{|E_{A1=Bajo}|}{|E|}\sigma_{(A1=Bajo)}$$

$$= \frac{3}{8}\cdot 1391{,}24 + \frac{3}{8}\cdot 81{,}64 + \frac{2}{8}\cdot 250 = 614{,}83$$

Ídem para A2:

$\mu_{(A2 = Alta)} \quad = \quad \dfrac{2500+4000}{2} = 3250$

$\mu_{(A2 = Media)} \quad = \quad \dfrac{2300+1000+4100+1100}{4} = 2125$

$\mu_{(A2 = Baja)} \quad = \quad \dfrac{500+2400}{2} = 1450$

$$\sigma_{(A2 = Alta)} = \sqrt{\frac{((2500-3250)^2+(4000-3250)^2)}{2}} = 750$$

$$\sigma_{(A2 = Media)} = \sqrt{\frac{((2300-2125)^2+(1000-2125)^2+(4100-2125)^2+(1100-2125)^2)}{4}} = 1249{,}75$$

$$\sigma_{(A2 = Baja)} = \sqrt{\frac{((500-1450)^2+(2400-1450)^2)}{2}} = 950$$

$$\text{Error}_{(A2)} = \frac{2}{8}\cdot 750 + \frac{4}{8}\cdot 1249{,}75 + \frac{2}{8}\cdot 950 = 1049{,}87$$

Ídem para A3:

$\mu_{(A3 = Si)} \quad = \quad \dfrac{2300+2500+1000+4100+1100}{5} = 2200$

$\mu_{(A3 = No)} = \dfrac{4000+500+2400}{3} = 2300$

$\sigma_{(A3 = Si)} = \sqrt{\dfrac{((2300-2200)^2+(2500-2200)^2+(1000-2200)^2+(4100-2200)^2+(1100-2200)^2)}{5}}$
$= 1127{,}83$

$\sigma_{(A3 = No)} = \sqrt{\dfrac{((4000-2300)^2+(500-2300)^2+(2400-2300)^2)}{3}} = 1430{,}62$

$Error_{(A3)} = \dfrac{5}{8}\cdot 1127{,}83 + \dfrac{3}{8}\cdot 1430{,}62 = 1241{,}4$

Ídem para A4:

$\mu_{(A4 = Si)} = \dfrac{2300+4000+1000+4100}{4} = 2850$

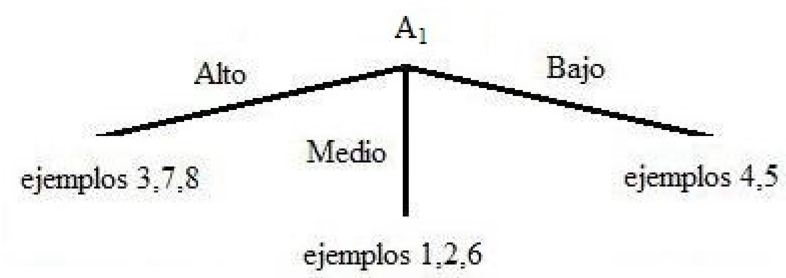
$\mu_{(A4 = No)} = \dfrac{2500+500+2400+1100}{4} = 1625$

$\sigma_{(A4 = Si)} = \sqrt{\dfrac{((2300-2850)^2+(4000-2850)^2+(1000-2850)^2+(4100-2850)^2)}{4}} = 1285{,}50$

$\sigma_{(A4 = No)} = \sqrt{\dfrac{((2500-1625)^2+(500-1625)^2+(2400-1625)^2+(1100-1625)^2)}{4}} = 852{,}70$

$Error_{(A4)} = \dfrac{4}{8}\cdot 1285{,}5 + \dfrac{4}{8}\cdot 852{,}7 = 1069{,}1$

Se elije el atributo que tenga el menor error, en este caso A1, con lo que se crea un nodo raíz en el que se distingue por este atributo y que tiene tres sucesores etiquetados con los valores del atributo: alto, medio y bajo, lo que nos da el siguiente árbol:



para cada sucesor, se comprueban las condiciones del enunciado (número de ejemplos mínimos por nodo y desviación estándar mínima)...

Para A1 = alto, se cumple la primera condición: nº de ejemplos = 3 (ejemplos 3,7,8) ≥ 2, y también la segunda, $\sigma_{(A1 = Alto)} = 1391{,}24$ que es > 100 . Por tanto este nodo puede ser expandido.

Para A1 = medio, se cumple la primera condición, nº de ejemplos = 3 (ejemplos 1,2,6) ≥ 2 pero no la segunda: $\sigma_{(A1 = Media)} = 81{,}65 < 100$. No se expande y se convierte en nodo hoja.

Para A1 = bajo, cumple la primera condición, nº de ejemplos = 2 (ejemplos 4,5) ≥ 2, y también cumple la segunda $\sigma_{(A1 = Bajo)} = 250 > 100$.

Por ello, habrá que seguir subdividiendo los valores de A1 (Sueldo) iguales a *Alto* y *Bajo*, con las siguientes llamadas recursivas:

Para el nodo de A1 (Sueldo) = Alto, llamamos a construir-árbol-regresión(ejemplos 3,7,8; atributos A2,A3,A4; nodo A1 (Sueldo) =Alto):

Ejemplo	Hipoteca A2	Casado A3	Hijos A4	Saldo (Clase)
3	Alta	No	Sí	4000
7	Media	Sí	Sí	4100
8	Media	Sí	No	1100

y buscamos el mejor atributo, aquel que tenga menor error:

$$\mu_{(A2 = Alta)} = \frac{4000}{1} = 4000$$

$$\mu_{(A2 = Media)} = \frac{4100+ 1100}{2} = 2600$$

$$\mu_{(A2 = Baja)} = 0$$

$$\sigma_{(A2 = Alta)} = \sqrt{\frac{((4000-4000)^2)}{1}} = 0$$

$$\sigma_{(A2 = Media)} = \sqrt{\frac{((4100-2600)^2+ (1100-2600)^2)}{2}} = 1729,88$$

$$\sigma_{(A2 = Baja)} = 0$$

$$\text{Error}_{(A2)} = 0 + \frac{2}{3}\cdot 1729,88 + 0 = 1153,25$$

$$\mu_{(A3 = Sí)} = \frac{4100+ 1100}{2} = 2600$$

$$\mu_{(A3 = No)} = \frac{4000}{1} = 4000$$

$$\sigma_{(A3 = Sí)} = \sqrt{\frac{((4000-4000)^2)}{1}} = 0$$

$$\sigma_{(A3 = No)} = \sqrt{\frac{((4100-2600)^2+ (1100-2600)^2)}{2}} = 1729,88$$

$$\text{Error}_{(A3)} = 0 + \frac{2}{3}\cdot 1729,88 = 1153,25$$

$$\mu_{(A4 = Sí)} = \frac{4000+ 4100}{2} = 4050$$

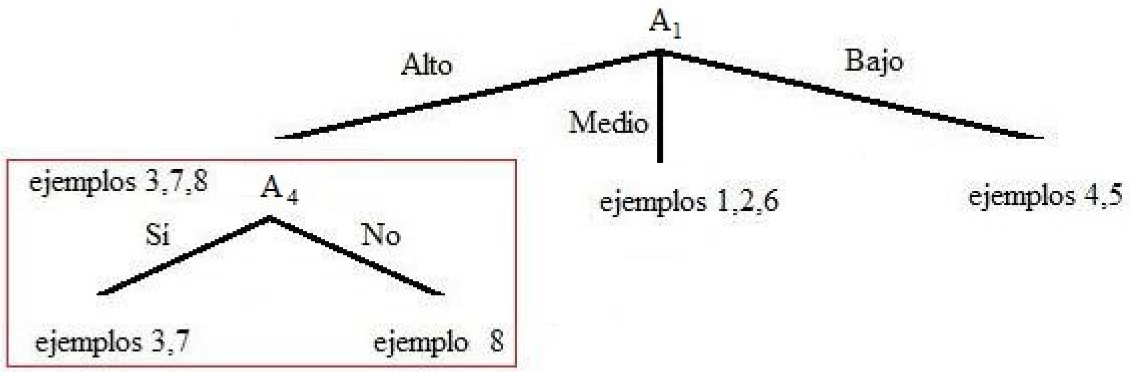
$$\mu_{(A4 = No)} = \frac{1100}{1} = 1100$$

$$\sigma_{(A4 = Sí)} = \sqrt{\frac{((4000-4050)^2+ (4100-4050)^2)}{2}} = 50$$

$$\sigma_{(A4 = No)} = \sqrt{\frac{((1100-1100)^2)}{1}} = 0$$

$$\text{Error}_{(A4)} = \frac{2}{3}\cdot 50 + 0 = 33,33$$

El atributo con menor error es A4 (Hijos), expandimos el árbol anterior y nos queda:



Antes de continuar en la rama del sucesor de A1 = Bajo, el algoritmo analiza el nuevo árbol analiza los sucesores del nuevo árbol que acabamos de general (enmarcado en rojo).

Para cada sucesor, se comprueban las condiciones del enunciado (número de ejemplos mínimos por nodo y desviación estándar mínima)...

Para el nodo con A4 = No, vemos que no puede ser expandido al no cumplir la condición de número mínimo de ejemplos por nodo, debería ser 2 y solo tiene uno. Esta condición sí la cumple el nodo con A4 = Sí, con 2 ejemplos. Vemos que $\sigma_{(A4 = Si)} = 50 > 100$, que no se cumple. Por tanto este nodo tampoco puede ser expandido y se convierte en nodo hoja.

Volvemos al nodo de A1 (Sueldo) = Bajo, y llamamos a construir-árbol-regresión(ejemplos 4,5; atributos A2,A3,A4; nodo A1 (Sueldo) =Bajo):

Ejemplo	Hipoteca A2	Casado A3	Hijos A4	Saldo (Clase)
4	Media	Sí	Sí	1000
5	Baja	No	No	500

$\mu_{(A2 = Alta)} = 0$

$\mu_{(A2 = Baja)} = \frac{500}{1} = 500$

$\mu_{(A2 = Media)} = \frac{1000}{1} = 1000$

$\sigma_{(A2 = Alta)} = 0$

$\sigma_{(A2 = Media)} = \sqrt{\frac{((1000-1000)^2)}{2}} = 0$

$\sigma_{(A2 = Baja)} = \sqrt{\frac{((500-500)^2)}{1}} = 0$

$\text{Error}_{(A2)} = 0 + 0 = 0$

$\mu_{(A3 = Si)} = \frac{1000}{1} = 1000$

$\mu_{(A3 = No)} = \frac{500}{1} = 500$

$\sigma_{(A3 = Sí)} = \sqrt{\frac{((1000-1000)^2)}{2}} = 0$

$\sigma_{(A3 = No)} = \sqrt{\frac{((500-500)^2)}{1}} = 0$

$Error_{(A3)} = 0 + 0 = 0$

$\mu_{(A4 = Sí)} = \frac{1000}{1} = 1000$

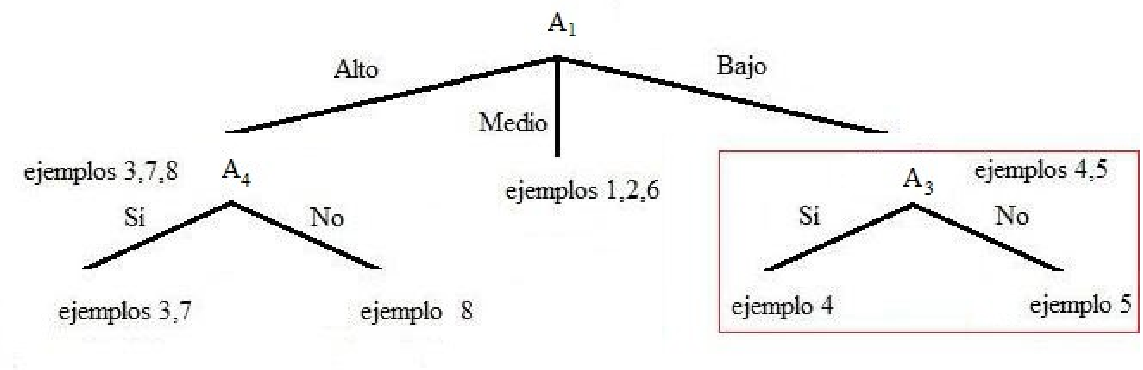
$\mu_{(A4 = No)} = \frac{500}{1} = 500$

$\sigma_{(A4 = Sí)} = \sqrt{\frac{((1000-1000)^2)}{2}} = 0$

$\sigma_{(A4 = No)} = \sqrt{\frac{((500-500)^2)}{1}} = 0$

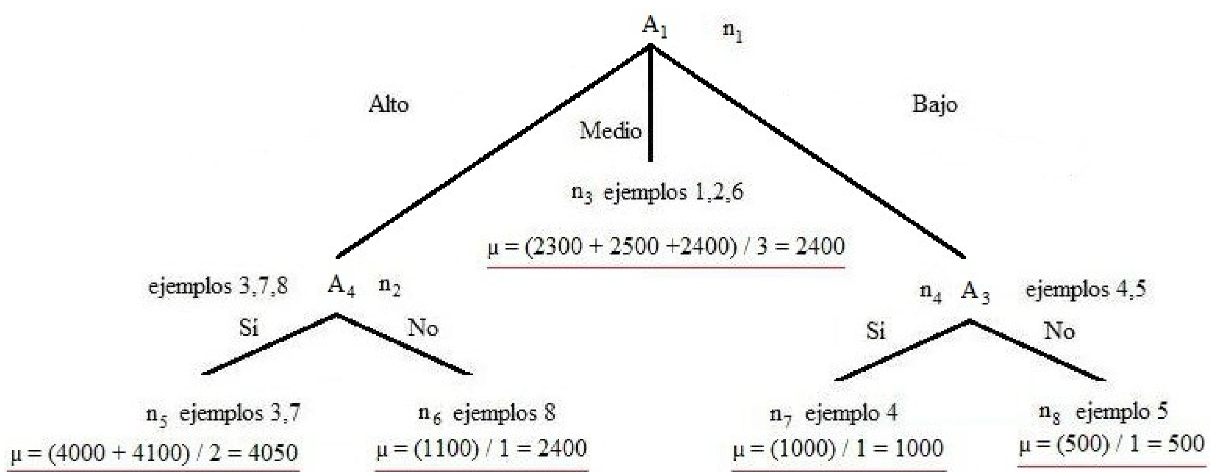
$Error_{(A4)} = 0 + 0 = 0$

Todos tienen el valor de error a 0, elegimos uno aleatoriamente, A3 (Casado) por ejemplo. Y lo expandimos:



Como ambos sucesores A3 (Casado) = Sí y A3 (Casado) = No, tienen solo un nodo, no cumplen la condición para ser expandidos y se convierten en nodos hoja.

Aquí se termina de calcular el árbol de regresión, numeramos también los nodos y se representa el modelo predictivo como el valor medio de la clase de los ejemplos de cada hoja:



16F2

PREGUNTA 3. Valoración: 2.5
Supongamos que una entidad bancaria quiere predecir el nivel de ahorro que cada uno de sus clientes tendrá el año siguiente en función de la información que dispone de cada uno de ellos (ver tabla 1). Para abordar dicha tarea, se construirá un árbol de regresión. ¿Cuál sería el modelo predictivo resultante? Prescinda de las etapas de poda en la construcción del árbol.

Para facilitar la construcción del **árbol de regresión**, en lugar de asociar un modelo lineal a cada uno de sus nodos hojas, realice un promediado del valor del atributo de clase de los casos pertenecientes a cada nodo hoja.

Existen dos condiciones para seguir dividiendo el árbol en un nodo:
1. Número de ejemplos en un nodo $n \geq 2$.
2. Desviación estándar de los ejemplos en un nodo $\sigma > 75$.

Cliente	Sueldo A1	Hipoteca A2	Casado A3	Hijos A4	Ahorro (Clase)
1	Medio	Media	Sí	Sí	3300
2	Medio	Alta	Sí	No	3500
3	Alto	Alta	No	Sí	5000
4	Bajo	Baja	Sí	No	900
5	Bajo	Baja	Sí	Sí	400
6	Medio	Baja	No	No	3400
7	Alto	Media	Sí	Sí	5100
8	Alto	Media	Sí	No	5200

Tabla 1

Nota: La desviación estándar de n muestras, $x_1,\ldots,x_n$, viene dada por $\sigma = \sqrt{\sum_i^n \frac{(x_i-\mu)^2}{n}}$, siendo μ el valor medio.

Para construir un árbol de regresión en base al conjunto de los 8 ejemplos de la muestra y los 4 atributos y un árbol de regresión inicial vacío tengo que calcular el mejor atributo de dicho conjunto de ejemplos (de 1 a 8) y atributos (de A1 a A4).
El mejor atributo es aquel que proporciona el menor error ponderado, que viene dado por la fórmula $\text{Error}_{(Ai)} = \sum \left(\frac{|E_i|}{|E|}\sigma_{(A_i)}\right)$. Para calcular las desviaciones estándar ($\sigma_{(Ai)}$) hay que calcular también los valores medios μ_i de cada atributo:

$$\mu_i = \frac{\sum \textit{valores de clase de los ejemplo que tengan un valor de atributo i determinado}}{n^o \textit{ de ejemplos con un valor de atributo i determinado}}$$

$$\mu_{(A1=Alto)} = \frac{5000+5100+5200}{3} = 5100$$

$$\mu_{(A1=Medio)} = \frac{3300+3500+3400}{3} = 3400$$

$$\mu_{(A1=Bajo)} = \frac{900+400}{2} = 650$$

Cálculo de la desviación estándar del atributo:

$\sigma_{(Ai)} = \sqrt{\sum \frac{(x_i - \mu_{Ai})^2}{n}}$ = (en el libro de texto hay ejemplo que no divide entre 'n', ver explicación en fe de erratas)

$$\sigma_{(A1\,=\,Alto)} = \sqrt{\frac{((5000-5100)^2+(5100-5100)^2+(5200-5100)^2)}{3}} = 81{,}65$$

$$\sigma_{(A1\,=\,Medio)} = \sqrt{\frac{((3300-3400)^2+(3500-3400)^2+(3400-3400)^2)}{3}} = 81{,}65$$

$$\sigma_{(A1\,=\,Bajo)} = \sqrt{\frac{((900-650)^2+(400-650)^2)}{2}} = 250$$

Cálculo del error ponderado de A1:

$$\mathrm{Error}_{(A1)} = \sum \left(\frac{|E_{A1}|}{|E|}\sigma_{(A1)}\right)$$

$$= \frac{|E_{A1=Alto}|}{|E|}\sigma_{(A1=Alto)} + \frac{|E_{A1=Medio}|}{|E|}\sigma_{(A1=Medio)} + \frac{|E_{A1=Bajo}|}{|E|}\sigma_{(A1=Bajo)}$$

$$= \frac{3}{8}\cdot 81{,}65 + \frac{3}{8}\cdot 81{,}65 + \frac{2}{8}\cdot 250 = 123{,}74$$

Ídem para A2:

$$\mu_{(A2\,=\,Alta)} = \frac{3500+5000}{2} = 4250$$

$$\mu_{(A2\,=\,Media)} = \frac{3300+5100+5200}{3} = 4533{,}33$$

$$\mu_{(A2\,=\,Baja)} = \frac{900+400+3400}{3} = 1566{,}66$$

$$\sigma_{(A2\,=\,Alta)} = \sqrt{\frac{((3500-4250)^2+(5000-4250)^2)}{2}} = 750$$

$$\sigma_{(A2\,=\,Media)} = \sqrt{\frac{((3300-4533{,}33)^2+(5100-4533{,}33)^2+(5200-4533{,}33)^2)}{2}} = 873{,}05$$

$$\sigma_{(A2\,=\,Baja)} = \sqrt{\frac{((900-1566{,}66)^2+(400-1566{,}66)^2+(3400-1566{,}66)^2)}{3}} = 1312{,}33$$

$$\mathrm{Error}_{(A2)} = \frac{2}{8}\cdot 4250 + \frac{3}{8}\cdot 4533{,}33 + \frac{3}{8}\cdot 1566{,}66 = 3350$$

Ídem para A3:

$$\mu_{(A3\,=\,Si)} = \frac{3300+3500+900+400+5100+5200}{6} = 3066{,}66$$

$$\mu_{(A3\,=\,No)} = \frac{5000+3400}{2} = 4200$$

$$\sigma_{(A3\,=\,Si)} = \sqrt{\frac{((3300-3066{,}6)^2+(3500-3066{,}66)^2+\ldots+(5200-3066{,}66)^2)}{6}} = 1858{,}91$$

$$\sigma_{(A3\,=\,No)} = \sqrt{\frac{((4000-2300)^2+(500-2300)^2+(2400-2300)^2)}{3}} = 800$$

$$\text{Error}_{(A3)} = \frac{6}{8}\cdot 1858{,}91 + \frac{2}{8}\cdot 800 = 1594{,}18$$

Ídem para A4:

$$\mu_{(A4 = Si)} = \frac{3300+5000+400+5100}{4} = 3450$$

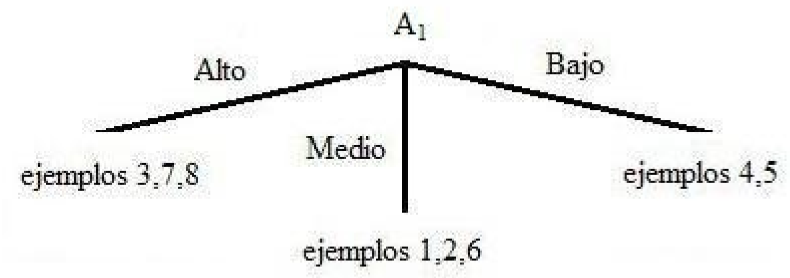
$$\mu_{(A4 = No)} = \frac{3500+900+3400+5200}{4} = 3250$$

$$\sigma_{(A4 = Si)} = \sqrt{\frac{((3300-3450)^2+(5000-3450)^2+(400-3450)^2+(5100-3450)^2)}{4}} = 1900{,}66$$

$$\sigma_{(A4 = No)} = \sqrt{\frac{((3500-3250)^2+(900-3250)^2+(3400-3250)^2+(5200-3250)^2)}{4}} = 1533{,}79$$

$$\text{Error}_{(A4)} = \frac{4}{8}\cdot 1900{,}66 + \frac{4}{8}\cdot 1533{,}79 = 1717{,}22$$

Se elije el atributo que tenga el menor error, en este caso A1, con lo que se crea un nodo raíz en el que se distingue por este atributo y que tiene tres sucesores etiquetados con los valores del atributo: alto, medio y bajo, lo que nos da el siguiente árbol:



para cada sucesor, se comprueban las condiciones del enunciado (número de ejemplos mínimos por nodo y desviación estándar mínima)...

Para A1 = Alto, nº de ejemplos = 3 (ejemplos 3,7,8) ≥ 2, y $\sigma_{(A1 = Alto)} = 81{,}65 > 75$, se cumple por tanto, y este nodo puede ser expandido.

Para A1 = Medio, nº de ejemplos = 3 (ejemplos 1,2,6) ≥ 2, y $\sigma_{(A1 = Medio)} = 81{,}65 > 75$, se cumple por tanto, y este nodo puede ser expandido.

Para A1 = Bajo, nº de ejemplos = 2 (ejemplos 4,5) ≥ 2, y $\sigma_{(A1 = Bajo)} = 250 > 75$, se cumple por tanto, y este nodo puede ser expandido.

Por ello, habrá que seguir subdividiendo los valores de A1 (Sueldo) iguales a *Alto*, *Medio* y *Bajo*, con las siguientes llamadas recursivas:

Para el nodo de A1 (Sueldo) = Alto, llamamos a construir-árbol-regresión(ejemplos 3,7,8; atributos A2,A3,A4; nodo A1 (Sueldo) =Alto):

$\sigma_{(A2)}$

Ejemplo	Hipoteca A2	Casado A3	Hijos A4	Saldo (Clase)
3	Alta	No	Sí	5000
7	Media	Sí	Sí	5100
8	Media	Sí	No	5200

y buscamos el mejor atributo, aquel que tenga menor error:

$$\mu_{(A2 = Alta)} = \frac{5000}{1} = 5000$$

$$\mu_{(A2 = Media)} = \frac{5100 + 5200}{2} = 5150$$

$$\mu_{(A2 = Baja)} = 0$$

$$\sigma_{(A2 = Alta)} = \sqrt{\frac{((5000-5000)^2)}{1}} = 0$$

$$\sigma_{(A2 = Media)} = \sqrt{\frac{((5100-5150)^2 + (5200-5150)^2)}{2}} = 50$$

$$\sigma_{(A2 = Baja)} = 0$$

$$Error_{(A2)} = 0 + \frac{2}{3} \cdot 50 + 0 = 33{,}33$$

$$\mu_{(A3 = Si)} = \frac{5100 + 5200}{2} = 5150$$

$$\mu_{(A3 = No)} = \frac{5000}{1} = 5000$$

$$\sigma_{(A3 = Si)} = \sqrt{\frac{((5100-5150)^2 + (5200-5150)^2)}{2}} = 50$$

$$\sigma_{(A3 = No)} = \sqrt{\frac{((5000-5000)^2)}{1}} = 0$$

$$Error_{(A3)} = 0 + \frac{2}{3} \cdot 50 = 33{,}33$$

$$\mu_{(A4 = Si)} = \frac{5000 + 5100}{2} = 5050$$

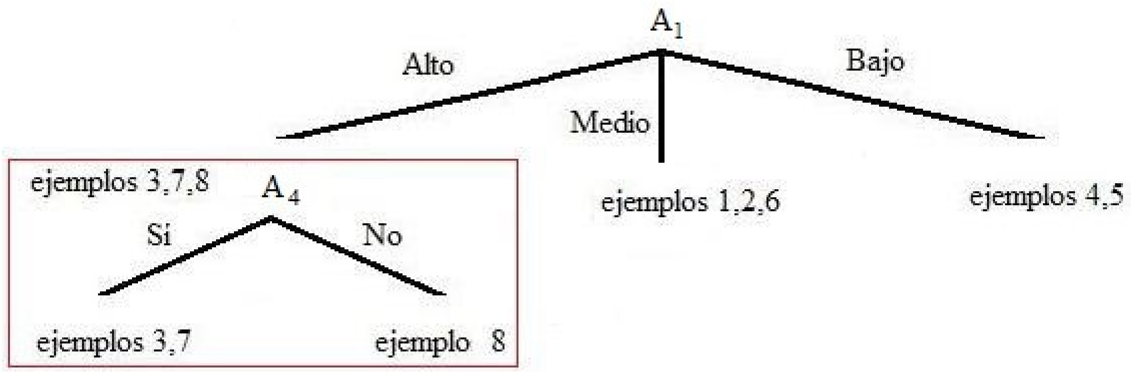
$$\mu_{(A4 = No)} = \frac{5200}{1} = 5200$$

$$\sigma_{(A4 = Si)} = \sqrt{\frac{((5000-5050)^2 + (5100-5050)^2)}{2}} = 50$$

$$\sigma_{(A4 = No)} = \sqrt{\frac{((1100-1100)^2)}{1}} = 0$$

$$Error_{(A4)} = \frac{2}{3} \cdot 50 + 0 = 33{,}33$$

El atributo con menor error es A4 (Hijos), expandimos el árbol anterior y nos queda:



Antes de continuar otras ramas de este mismo nivel (A1 = Medio y A1 = Bajo), el algoritmo analiza el nuevo árbol analiza los sucesores del nuevo árbol que acabamos de general (enmarcado en rojo).

Para cada sucesor, se comprueban las condiciones del enunciado (número de ejemplos mínimos por nodo y desviación estándar mínima)...

Para el nodo con A4 = No, vemos que no puede ser expandido al no cumplir la condición de número mínimo de ejemplos por nodo, debería ser 2 y solo tiene uno. Esta condición sí la cumple el nodo con A4 = Sí, con 2 ejemplos. Vemos que $\sigma_{(A4\,=\,Sí)} = 50 > 75$, que no se cumple. Por tanto este nodo tampoco puede ser expandido y se convierte en nodo hoja.

Volvemos al nodo de A1 (Sueldo) = Medio, y llamamos a construir-árbol-regresión(ejemplos 1,2,6; atributos A2,A3,A4; nodo A1 (Sueldo) =Medio):

Cliente	Hipoteca A2	Casado A3	Hijos A4	Ahorro (Clase)
1	Media	Sí	Sí	3300
2	Alta	Sí	No	3500
6	Baja	No	No	3400

$$\mu_{(A2\,=\,Alta)} = \frac{3500}{1} = 3500$$

$$\mu_{(A2\,=\,Media)} = \frac{3300}{1} = 3300$$

$$\mu_{(A2\,=\,Baja)} = \frac{3400}{1} = 3400$$

$$\sigma_{(A2\,=\,Alta)} = \sqrt{\frac{((3500-3500)^2)}{1}} = 0$$

$$\sigma_{(A2\,=\,Media)} = \sqrt{\frac{((3300-3300)^2)}{1}} = 0$$

$$\sigma_{(A2\,=\,Baja)} = \sqrt{\frac{((3400-3400)^2)}{1}} = 0$$

$$\text{Error}_{(A2)} = 0 + 0 + 0 = 0$$

$$\mu_{(A3 = Si)} = \frac{3300 + 3500}{2} = 3400$$

$$\mu_{(A3 = No)} = \frac{3400}{1} = 3400$$

$$\sigma_{(A3 = Si)} = \sqrt{\frac{((3300-3400)^2 + (3500-3400)^2)}{2}} = 100$$

$$\sigma_{(A3 = No)} = \sqrt{\frac{((3400-3400)^2)}{1}} = 0$$

$$\text{Error}_{(A3)} = \frac{2}{3}100 + 0 = 66{,}66$$

$$\mu_{(A4 = Si)} = \frac{3300}{1} = 3300$$

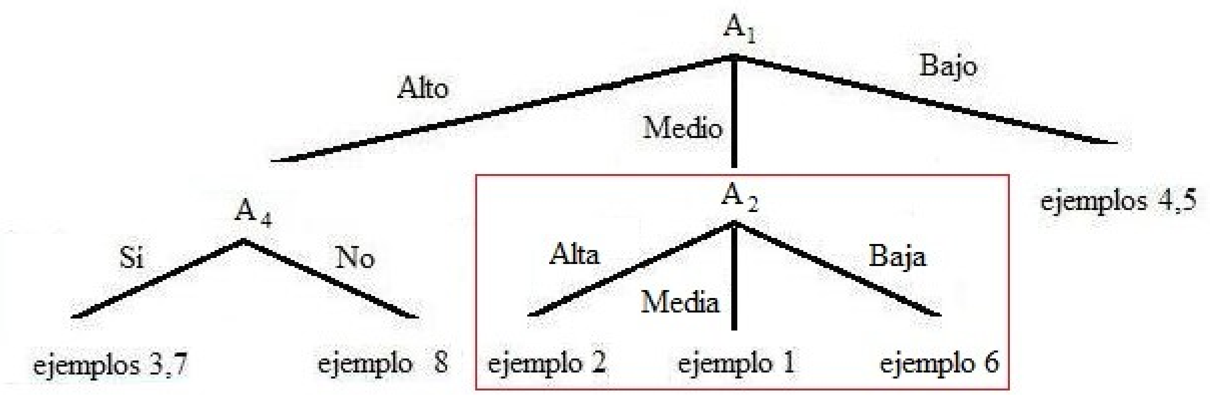
$$\mu_{(A4 = No)} = \frac{3500 + 3400}{2} = 3450$$

$$\sigma_{(A4 = Si)} = \sqrt{\frac{((3300-3300)^2)}{1}} = 0$$

$$\sigma_{(A4 = No)} = \sqrt{\frac{((3500-3450)^2 + (3400-3450)^2)}{2}} = 50$$

$$\text{Error}_{(A4)} = 0 + \frac{2}{3}50 = 33{,}33$$

El atributo A2 (Hipotecado) es el de menor error, y lo expandimos:



Del mismo modo que antes, el algoritmo analiza el nuevo árbol analiza los sucesores del nuevo árbol que acabamos de general (enmarcado en rojo) antes de continuar con la rama A1 = Bajo.

Para cada sucesor, se comprueban las condiciones del enunciado (número de ejemplos mínimos por nodo y desviación estándar mínima). Todos incumplen la condición de número de nodos mínimo, por lo que no se expanden y se transforman en nodos hoja.

Volvemos al nodo de A1 (Sueldo) = Bajo, y llamamos a construir-árbol-regresión(ejemplos 4,5; atributos A2,A3,A4; nodo A1 (Sueldo) =Bajo):

Cliente	Hipoteca A2	Casado A3	Hijos A4	Ahorro (Clase)
4	Baja	Sí	No	900
5	Baja	Sí	Sí	400

$\mu_{(A2 = Alta)} = 0$

$\mu_{(A2 = Baja)} = \dfrac{900+400}{2} = 650$

$\mu_{(A2 = Media)} = 0$

$\sigma_{(A2 = Alta)} = 0$

$\sigma_{(A2 = Media)} = 0$

$$\sigma_{(A2 = Baja)} = \sqrt{\frac{((900-650)^2+(400-650)^2)}{2}} = 250$$

$$Error_{(A2)} = 0 + 0 + \frac{2}{2}250 = 250$$

$\mu_{(A3 = Si)} = \dfrac{900+400}{2} = 650$

$\mu_{(A3 = No)} = 0$

$$\sigma_{(A3 = Si)} = \sqrt{\frac{((900-650)^2+(400-650)^2)}{2}} = 250$$

$\sigma_{(A3 = No)} = 0$

$$Error_{(A3)} = \frac{2}{2}250 + 0 = 250$$

$\mu_{(A4 = Si)} = \dfrac{400}{1} = 400$

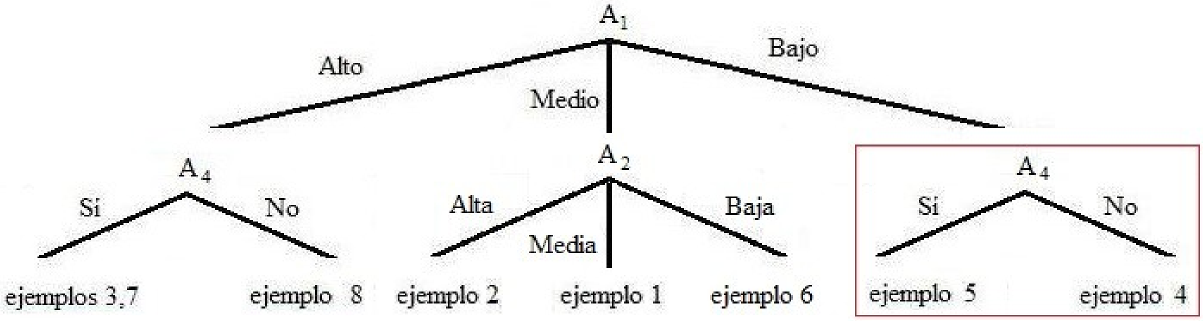
$\mu_{(A4 = No)} = \dfrac{900}{1} = 900$

$$\sigma_{(A4 = Si)} = \sqrt{\frac{((400-400)^2)}{1}} = 0$$

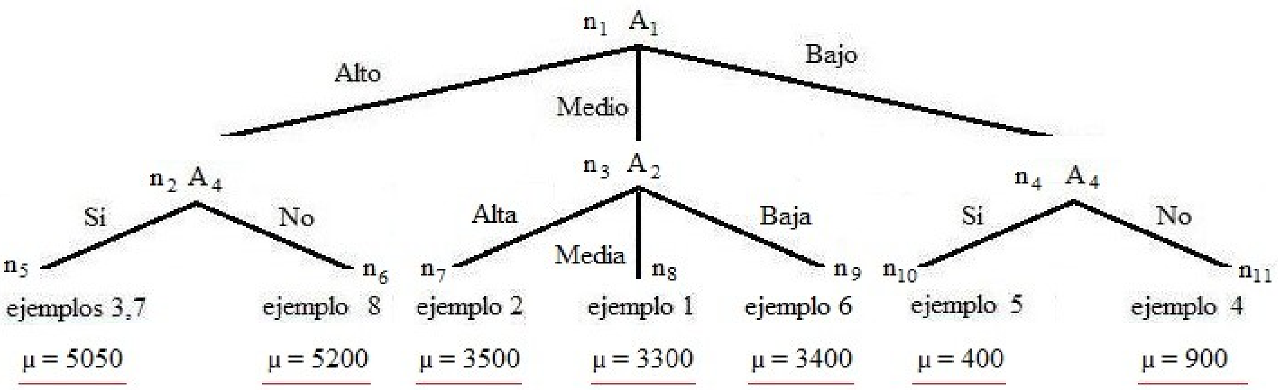
$$\sigma_{(A4 = No)} = \sqrt{\frac{((900-900)^2)}{1}} = 0$$

$$Error_{(A4)} = 0 + 0 = 0$$

El atributo de menor error es el A4 (Hijos) y lo expandimos:



Como ambos sucesores A4, tienen solo un nodo, no cumplen la condición para ser expandidos y se convierten en nodos hoja. Aquí se termina de calcular el árbol de regresión, numeramos también los nodos y se representa el modelo predictivo (valor medio de la clase de los ejemplos de las hoja):



17SO

PREGUNTA 3. Valoración: 3

Una forma relativamente sencilla de construir **árboles de regresión** es sustituir el modelo lineal de cada uno de sus nodos hojas por una constante que representa el valor medio de la clase de los ejemplos pertenecientes a dicha hoja. Utilice este tipo de árboles para resolver el siguiente problema.

Supongamos que una entidad bancaria quiere predecir el saldo que cada uno de sus clientes tendrá el año siguiente. Para ello, dispone de la información que se muestra en la tabla adjunta. ¿Cuál sería el modelo predictivo resultante? Prescinda de las etapas de poda en la construcción del árbol.

Existen dos condiciones para seguir dividiendo el árbol en un nodo:

1. Número de ejemplos en un nodo $n \geq 3$.
2. Desviación estándar de los ejemplos en un nodo $\sigma > 100$.

<u>Muy importante</u>: Muestre y justifique todos los pasos y resultados numéricos intermedios necesarios para la obtención del árbol.

Ejemplo	Sueldo A1	Nivel Gasto A2	Casado A3	Hijos A4	Saldo (Clase)
1	Bajo	Bajo	No	No	500
2	Medio	Bajo	No	No	2400
3	Alto	Medio	Sí	Sí	4100
4	Alto	Medio	Sí	No	1100
5	Medio	Medio	Sí	Sí	2300
6	Medio	Alto	Sí	No	2500
7	Alto	Alto	No	Sí	4000
8	Bajo	Medio	Sí	Sí	1000

Para construir un árbol de regresión en base al conjunto de los 8 ejemplos de la muestra y los 4 atributos y un árbol de regresión inicial vacío tengo que calcular el mejor atributo de dicho conjunto de ejemplos (de 1 a 8) y atributos (de A1 a A4).

El mejor atributo es aquel que proporciona el menor error ponderado, que viene dado por la fórmula $\text{Error}_{(Ai)} = \sum \left(\frac{|E_i|}{|E|}\sigma_{(A_i)}\right)$. Para calcular las desviaciones estándar ($\sigma_{(Ai)}$) hay que calcular también los valores medios μ_i de cada atributo:

$$\mu_i = \frac{\sum \textit{valores de clase de los ejemplo que tengan un valor de atributo i determinado}}{n^o \textit{ de ejemplos con un valor de atributo i determinado}}$$

$$\mu_{(A1 = Alto)} = \frac{4100 + 1100 + 4000}{3} = 3066{,}66$$

$$\mu_{(A1 = Medio)} = \frac{2400 + 2300 + 2500}{3} = 2400$$

$$\mu_{(A1 = Bajo)} = \frac{500 + 1000}{2} = 750$$

Cálculo de la desviación estándar del atributo:

$\sigma_{(Ai)} = \sqrt{\sum \frac{(x_i - \mu_{Ai})^2}{n}}$ = (en el libro de texto hay ejemplo que no divide entre 'n', ver explicación en fe de erratas)

$$\sigma_{(A1 = Alto)} = \sqrt{\frac{((4100-3066{,}66)^2+(1100-3066{,}66)^2+(4000-3066{,}66)^2)}{3}} = 1391{,}24$$

$$\sigma_{(A1 = Medio)} = \sqrt{\frac{((2400-2400)^2+(2300-2400)^2+(2500-2400)^2)}{3}} = 81{,}65$$

$$\sigma_{(A1 = Bajo)} = \sqrt{\frac{((500-750)^2+(100-750)^2)}{2}} = 250$$

Cálculo del error ponderado de A1:

$$\text{Error}_{(A1)} = \sum \left(\frac{|E_{A1}|}{|E|}\sigma_{(A1)}\right)$$

$$= \frac{|E_{A1=Alto}|}{|E|}\sigma_{(A1=Alto)} + \frac{|E_{A1=Medio}|}{|E|}\sigma_{(A1=Medio)} + \frac{|E_{A1=Bajo}|}{|E|}\sigma_{(A1=Bajo)}$$

$$= \frac{3}{8}\cdot 1391{,}24 + \frac{3}{8}\cdot 81{,}65 + \frac{2}{8}\cdot 250 = 614{,}83$$

Ídem para A2:

$$\mu_{(A2 = Alta)} = \frac{2500+4000}{2} = 3250$$

$$\mu_{(A2 = Media)} = \frac{4100+1100+2300+1000}{4} = 2125$$

$$\mu_{(A2 = Baja)} = \frac{500+2400}{2} = 1450$$

$$\sigma_{(A2 = Alta)} = \sqrt{\frac{((2500-3250)^2+(4000-3250)^2)}{2}} = 750$$

$$\sigma_{(A2 = Media)} = \sqrt{\frac{((4100-2125)^2+(1100-2125)^2+(2300-2125)^2+(1000-2125)^2)}{4}} = 1249{,}75$$

$$\sigma_{(A2 = Baja)} = \sqrt{\frac{((500-1450)^2+(2400-1450)^2)}{2}} = 950$$

$$\text{Error}_{(A2)} = \frac{2}{8}\cdot 750 + \frac{4}{8}\cdot 1249{,}75 + \frac{2}{8}\cdot 950 = 1049{,}87$$

Ídem para A3:

$$\mu_{(A3 = Sí)} = \frac{4100+1100+2300+2500+1000}{5} = 2200$$

$$\mu_{(A3 = No)} = \frac{500+2400+4000}{3} = 2300$$

$$\sigma_{(A3 = Sí)} = \sqrt{\frac{((4100-2200)^2+(1100-2200)^2+(2300-2200)^2+(2500-2200)^2+(1000-2200)^2)}{5}}$$

$$= 1127{,}83$$

$$\sigma_{(A3 = No)} = \sqrt{\frac{((500-2300)^2+(2400-2300)^2+(4000-2300)^2)}{3}} = 1430{,}62$$

$$\text{Error}_{(A3)} = \frac{5}{8}\cdot 1127{,}83 + \frac{3}{8}\cdot 1430{,}62 = 1241{,}38$$

Ídem para A4:

$$\mu_{(A4 = Si)} = \frac{4100+2300+4000+1000}{4} = 2850$$

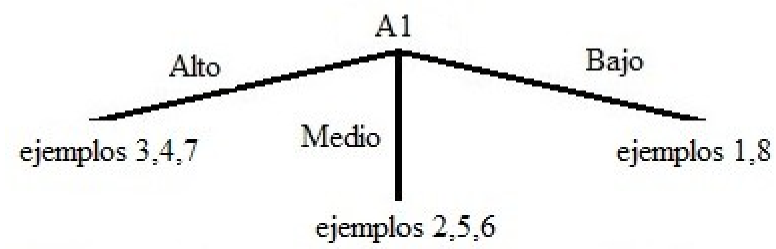
$$\mu_{(A4 = No)} = \frac{500+2400+1100+2500}{4} = 1625$$

$$\sigma_{(A4 = Si)} = \sqrt{\frac{((4100-2850)^2+(2300-2850)^2+(4000-2850)^2+(1000-2850)^2)}{4}} = 1285{,}50$$

$$\sigma_{(A4 = No)} = \sqrt{\frac{((500-1625)^2+(2400-1625)^2+(1100-1625)^2+(2500-1625)^2)}{4}} = 852{,}70$$

$$Error_{(A4)} = \frac{4}{8}\cdot 1285{,}5 + \frac{4}{8}\cdot 852{,}7 = 1069{,}1$$

Se elije el atributo que tenga el menor error, en este caso A1, con lo que se crea un nodo raíz en el que se distingue por este atributo y que tiene tres sucesores etiquetados con los valores del atributo: alto, medio y bajo, lo que nos da el siguiente árbol:



para cada sucesor, se comprueban las condiciones del enunciado (número de ejemplos mínimos por nodo y desviación estándar mínima)...

Para A1 = Alto, nº de ejemplos = 3 (ejemplos 3,4,7) ≥ 3, y $\sigma_{(A1 = Alto)} = 1391{,}24 > 100$, se cumple por tanto, y este nodo puede ser expandido.

Para A1 = Medio, nº de ejemplos = 3 (ejemplos 2,5,6) ≥ 3, y $\sigma_{(A1 = Medio)} = 81{,}65 > 100$, con lo que la segunda condición no se cumple y, por tanto, este nodo puede ser expandido y se transforma en un nodo hoja.

Para A1 = Bajo, nº de ejemplos = 2 (ejemplos 1,8) ≥ 3, que no cumple la primera condición, por lo que no se expande y se convierte en un nodo hoja.

Por ello, habrá que seguir subdividiendo el sucesor con valor de A1 (Sueldo) igual a *Alto*:

Para el nodo de A1 (Sueldo) = Alto, llamamos a construir-árbol-regresión(ejemplos 3,4,7; atributos A2,A3,A4; nodo A1 (Sueldo) =Alto):

Ejemplo	Nivel Gasto A2	Casado A3	Hijos A4	Saldo (Clase)
3	Medio	Sí	Sí	4100
4	Medio	Sí	No	1100
7	Alto	No	Sí	4000

y buscamos el mejor atributo, aquel que tenga menor error:

$$\mu_{(A2 = Alto)} = \frac{4000}{1} = 4000$$

$$\mu_{(A2 = Medio)} = \frac{4100+1100}{2} = 2600$$

$$\mu_{(A2 = Bajo)} = 0$$

$$\sigma_{(A2 = Alto)} = \sqrt{\frac{((4000-4000)^2)}{1}} = 0$$

$$\sigma_{(A2 = Medio)} = \sqrt{\frac{((4100-2600)^2+(1100-2600)^2)}{2}} = 1500$$

$$\sigma_{(A2 = Bajo)} = 0$$

$$Error_{(A2)} = 0 + \frac{2}{3}\cdot 1500 + 0 = 1000$$

$$\mu_{(A3 = Si)} = \frac{4100+1100}{2} = 2600$$

$$\mu_{(A3 = No)} = \frac{4000}{1} = 4000$$

$$\sigma_{(A3 = Si)} = \sqrt{\frac{((4100-2600)^2+(1100-2600)^2)}{2}} = 1500$$

$$\sigma_{(A3 = No)} = \sqrt{\frac{((4000-4000)^2)}{1}} = 0$$

$$Error_{(A3)} = \frac{2}{3}\cdot 1500 + 0 = 1000$$

$$\mu_{(A4 = Si)} = \frac{4100+4000}{2} = 4050$$

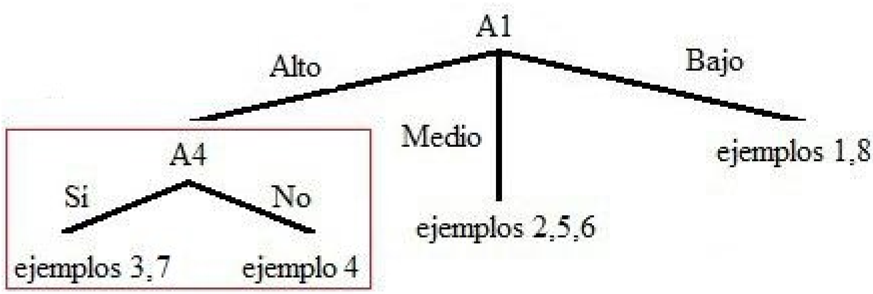
$$\mu_{(A4 = No)} = \frac{1100}{1} = 1100$$

$$\sigma_{(A4 = Si)} = \sqrt{\frac{((4100-4050)^2+(4000-4050)^2)}{2}} = 50$$

$$\sigma_{(A4 = No)} = \sqrt{\frac{((1100-1100)^2)}{1}} = 0$$

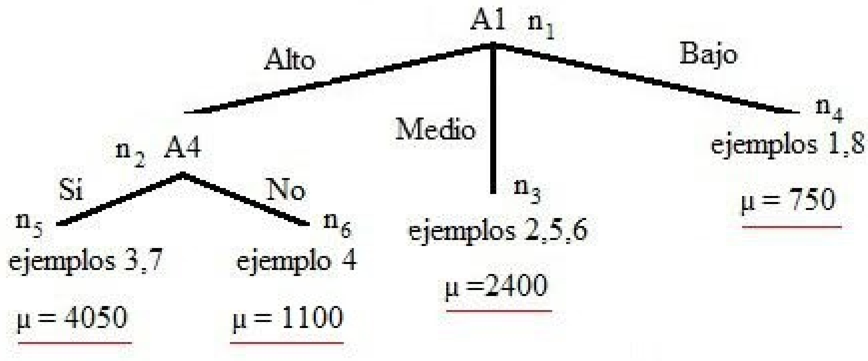
$$Error_{(A4)} = \frac{2}{3}\cdot 50 + 0 = 33,33$$

El atributo con menor error es A4 (Hijos), expandimos el árbol anterior y nos queda:



para cada sucesor, se comprueban las condiciones del enunciado (número de ejemplos mínimos por nodo y desviación estándar mínima) y vemos que ambos incumpln el número mínimo de ejemplos por nodo, por lo que no se expanden y se transforman en nodos hoja.

Aquí se termina de calcular el árbol de regresión, numeramos también los nodos y se representa el modelo predictivo como el valor medio de la clase de los ejemplos de cada hoja:



18F1

PREGUNTA 3. Valoración: 3

Una forma relativamente sencilla de construir **árboles de regresión** es sustituir el modelo lineal de cada uno de sus nodos hojas por una constante que representa el valor medio de la clase de los ejemplos pertenecientes a dicha hoja. Utilice este tipo de árboles para resolver el siguiente problema.

Supongamos que el servicio de mantenimiento de una empresa quiere predecir el tiempo de vida útil de un determinado tipo de máquina en función de la configuración de diferentes selectores **S_i**, con i=1,...,4. Para ello, dispone de la información que se muestra en la tabla 2. ¿Cuál sería el modelo predictivo resultante? Prescinda de las etapas de poda en la construcción del árbol.

Existen dos condiciones para seguir dividiendo el árbol en un nodo:

1. Número de ejemplos en un nodo $n \geq 2$.
2. Desviación estándar de los ejemplos en un nodo $\sigma > 150$.

Ejemplo	S_1	S_2	S_3	S_4	Nº Horas (Clase)
1	Bajo	Bajo	No	No	900
2	Medio	Bajo	No	No	2400
3	Alto	Medio	Sí	Sí	4100
4	Alto	Medio	Sí	No	1100
5	Medio	Medio	Sí	Sí	2300
6	Medio	Alto	Sí	No	2500
7	Alto	Alto	No	Sí	4000
8	Bajo	Medio	Sí	Sí	1000

Tabla2

Para construir un árbol de regresión en base al conjunto de los 8 ejemplos de la muestra y los 4 atributos y un árbol de regresión inicial vacío tengo que calcular el mejor atributo de dicho conjunto de ejemplos (de 1 a 8) y atributos (de S1 a S4).

El mejor atributo es aquel que proporciona el menor error ponderado, que viene dado por la fórmula $\text{Error}_{(Si)} = \sum \left(\frac{|E_i|}{|E|}\sigma_{(S_i)}\right)$. Para calcular las desviaciones estándar ($\sigma_{(Si)}$) hay que calcular también los valores medios μ_i de cada atributo:

$$\mu_i = \frac{\sum \textit{valores de clase de los ejemplo que tengan un valor de atributo i determinado}}{n^o \textit{ de ejemplos con un valor de atributo i determinado}}$$

$$\mu_{(S1 = Alto)} = \frac{4100+1100+4000}{3} = 3066{,}66$$

$$\mu_{(S1 = Medio)} = \frac{2400+2300+2500}{3} = 2400$$

$$\mu_{(S1 = Bajo)} = \frac{900+1000}{2} = 950$$

Cálculo de la desviación estándar del atributo:

$\sigma_{(Si)} = \sqrt{\sum \frac{(x_i - \mu_{Si})^2}{n}}$ = (en el libro de texto hay ejemplo que no divide entre 'n', ver explicación en fe de erratas)

$$\sigma_{(S1 = Alto)} = \sqrt{\frac{((4100-3066,66)^2 + (1100-3066,66)^2 + (4000-3066,66)^2)}{3}} = 1391,24$$

$$\sigma_{(S1 = Medio)} = \sqrt{\frac{((2400-2400)^2 + (2300-2400)^2 + (2500-2400)^2)}{3}} = 81,65$$

$$\sigma_{(S1 = Bajo)} = \sqrt{\frac{((900-950)^2 + (1000-950)^2)}{2}} = 50$$

Cálculo del error ponderado de S1:

$$\text{Error}_{(S1)} = \sum \left(\frac{|E_{S1}|}{|E|}\sigma_{(S1)}\right)$$

$$= \frac{|E_{S1=Alto}|}{|E|}\sigma_{(S1=Alto)} + \frac{|E_{S1=Medio}|}{|E|}\sigma_{(S1=Medio)} + \frac{|E_{S1=Bajo}|}{|E|}\sigma_{(S1=Bajo)}$$

$$= \frac{3}{8}\cdot 1391,24 + \frac{3}{8}\cdot 81,65 + \frac{2}{8}\cdot 50 = 564,83$$

Ídem para S2:

$$\mu_{(S2 = Alta)} = \frac{2500+4000}{2} = 3250$$

$$\mu_{(S2 = Media)} = \frac{4100+1100+2300+1000}{4} = 2125$$

$$\mu_{(S2 = Baja)} = \frac{900+2400}{2} = 1650$$

$$\sigma_{(S2 = Alta)} = \sqrt{\frac{((2500-3250)^2 + (4000-3250)^2)}{2}} = 750$$

$$\sigma_{(S2 = Media)} = \sqrt{\frac{((4100-2125)^2 + (1100-2125)^2 + (2300-2125)^2 + (1000-2125)^2)}{4}} = 1249,75$$

$$\sigma_{(S2 = Baja)} = \sqrt{\frac{((900-1650)^2 + (2400-1650)^2)}{2}} = 750$$

$$\text{Error}_{(S2)} = \frac{2}{8}\cdot 750 + \frac{4}{8}\cdot 1249,75 + \frac{2}{8}\cdot 750 = 999,87$$

Ídem para S3:

$$\mu_{(S3 = Sí)} = \frac{4100+1100+2300+2500+1000}{5} = 2200$$

$$\mu_{(S3 = No)} = \frac{900+2400+4000}{3} = 2433,33$$

$$\sigma_{(S3 = Sí)} = \sqrt{\frac{((4100-2200)^2 + (1100-2200)^2 + (2300-2200)^2 + (2500-2200)^2 + (1000-2200)^2)}{5}}$$

$$= 1127,83$$

$$\sigma_{(S3 = No)} = \sqrt{\frac{((900-2433,33)^2 + (2400-2433,33)^2 + (4000-2433,33)^2)}{3}} = 1265,79$$

$$\text{Error}_{(S3)} = \frac{5}{8} \cdot 1127{,}83 + \frac{3}{8} \cdot 1265{,}79 = 1179{,}56$$

Ídem para S4:

$$\mu_{(S4 = Sí)} = \frac{4100 + 2300 + 4000 + 1000}{4} = 2850$$

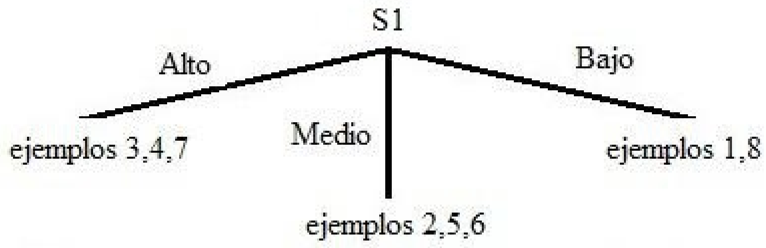
$$\mu_{(S4 = No)} = \frac{900 + 2400 + 1100 + 2500}{4} = 1725$$

$$\sigma_{(S4 = Sí)} = \sqrt{\frac{((4100-2850)^2 + (2300-2850)^2 + (4000-2850)^2 + (1000-2850)^2)}{4}} = 1285{,}50$$

$$\sigma_{(S4 = No)} = \sqrt{\frac{((900-1625)^2 + (2400-1625)^2 + (1100-1625)^2 + (2500-1625)^2)}{4}} = 729{,}30$$

$$\text{Error}_{(S4)} = \frac{4}{8} \cdot 1285{,}5 + \frac{4}{8} \cdot 729{,}30 = 1007{,}4$$

Se elije el atributo que tenga el menor error, en este caso S1, con lo que se crea un nodo raíz en el que se distingue por este atributo y que tiene tres sucesores etiquetados con los valores del atributo: alto, medio y bajo, lo que nos da el siguiente árbol:



para cada sucesor, se comprueban las condiciones del enunciado (número de ejemplos mínimos por nodo y desviación estándar mínima)...

Para S1 = Alto, nº de ejemplos = 3 (ejemplos 3,4,7) ≥ 2, y $\sigma_{(S1 = Alto)} = 1391{,}24 > 150$, se cumple por tanto, y este nodo puede ser expandido.

Para S1 = Medio, nº de ejemplos = 3 (ejemplos 2,5,6) ≥ 2, y $\sigma_{(S1 = Medio)} = 81{,}65 > 150$, con lo que la segunda condición no se cumple y, por tanto, este nodo puede ser expandido y se transforma en un nodo hoja.

Para S1 = Bajo, nº de ejemplos = 2 (ejemplos 1,8) ≥ 2, y $\sigma_{(S1 = Bajo)} = 50 > 150$, con lo que la segunda condición no se cumple y, por tanto, este nodo puede ser expandido y se transforma en un nodo hoja.

Por ello, habrá que seguir subdividiendo el sucesor con valor de S1 igual a *Alto*:

Para el nodo de S1 = Alto, llamamos a construir-árbol-regresión(ejemplos 3,4,7; atributos S2,S3,S4; nodo S1 =Alto):

Ejemplo	S_2	S_3	S_4	Nº Horas (Clase)
3	Medio	Sí	Sí	4100
4	Medio	Sí	No	1100
7	Alto	No	Sí	4000

y buscamos el mejor atributo, aquel que tenga menor error:

$$\mu_{(S2 = Alto)} = \frac{4000}{1} = 4000$$

$$\mu_{(S2 = Medio)} = \frac{4100+1100}{2} = 2600$$

$$\mu_{(S2 = Bajo)} = 0$$

$$\sigma_{(S2 = Alto)} = \sqrt{\frac{((4000-4000)^2)}{1}} = 0$$

$$\sigma_{(S2 = Medio)} = \sqrt{\frac{((4100-2600)^2+(1100-2600)^2)}{2}} = 1500$$

$$\sigma_{(S2 = Bajo)} = 0$$

$$Error_{(S2)} = 0 + \frac{2}{3}\cdot 1500 + 0 = 1000$$

$$\mu_{(S3 = Sí)} = \frac{4100+1100}{2} = 2600$$

$$\mu_{(S3 = No)} = \frac{4000}{1} = 4000$$

$$\sigma_{(S3 = Sí)} = \sqrt{\frac{((4100-2600)^2+(1100-2600)^2)}{2}} = 1500$$

$$\sigma_{(S3 = No)} = \sqrt{\frac{((4000-4000)^2)}{1}} = 0$$

$$Error_{(S3)} = \frac{2}{3}\cdot 1500 + 0 = 1000$$

$$\mu_{(S4 = Sí)} = \frac{4100+4000}{2} = 4050$$

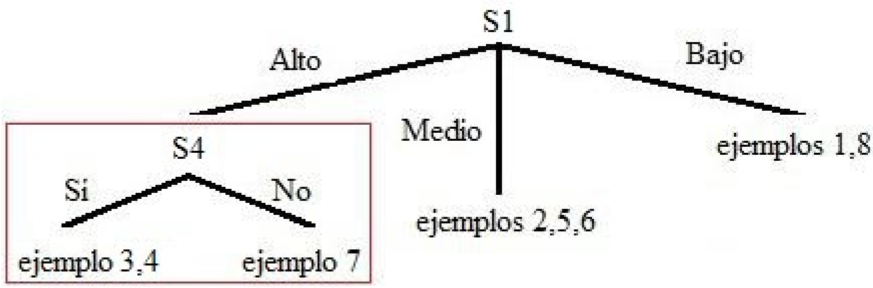
$$\mu_{(S4 = No)} = \frac{1100}{1} = 1100$$

$$\sigma_{(S4 = Sí)} = \sqrt{\frac{((4100-4050)^2+(4000-4050)^2)}{2}} = 50$$

$$\sigma_{(S4 = No)} = \sqrt{\frac{((1100-1100)^2)}{1}} = 0$$

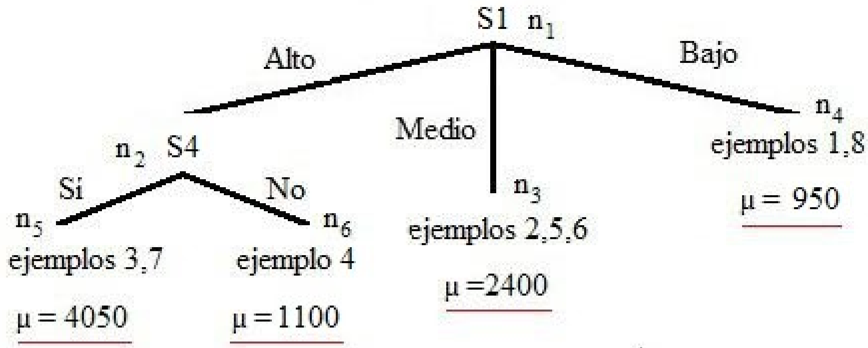
$$Error_{(S4)} = \frac{2}{3}\cdot 50 + 0 = 33,33$$

El atributo con menor error es S4, expandimos el árbol anterior y nos queda:



para cada sucesor, se comprueban las condiciones del enunciado (número de ejemplos mínimos por nodo y desviación estándar mínima) y vemos que ambos incumpln el número mínimo de ejemplos por nodo, por lo que no se expanden y se transforman en nodos hoja.

Aquí se termina de calcular el árbol de regresión, numeramos también los nodos y se representa el modelo predictivo como el valor medio de la clase de los ejemplos de cada hoja:



19SO

PREGUNTA 2. Valoración: 3

Supongamos que una empresa quiere predecir el consumo de agua que sus clientes tendrán el año siguiente en función de la información de consumo que dispone de algunos de ellos en el año actual (ver tabla 1). Para abordar dicha tarea, se construirá un **árbol de regresión**. Calcule el modelo predictivo resultante mostrando las diferentes etapas necesarias para obtenerlo. Además, tenga en cuenta la siguiente información:

(1) Prescinda de las etapas de poda en la construcción del árbol.

(2) Para facilitar la construcción del árbol de regresión, en lugar de asociar un modelo lineal a cada uno de sus nodos hojas, realice un promediado del valor del atributo de clase de los casos pertenecientes a cada nodo hoja.

(3) Existen dos condiciones para seguir dividiendo el árbol en un nodo:

- Número de ejemplos en un nodo $n \geq 2$.
- Desviación estándar de los ejemplos en un nodo $\sigma > 75$.

Cliente	A1	A2	A3	A4	Consumo
1	Media	Medio	Sí	Sí	3300
2	Alta	Medio	No	Sí	3500
3	Alta	Alto	Sí	No	5000
4	Baja	Bajo	No	Sí	900
5	Baja	Bajo	Sí	Sí	400
6	Baja	Medio	No	No	3400
7	Media	Alto	Sí	Sí	5100
8	Media	Alto	No	Sí	5200

Tabla 1

Para construir un árbol de regresión en base al conjunto de los 8 ejemplos de la muestra y los 4 atributos y un árbol de regresión inicial vacío tengo que calcular el mejor atributo de dicho conjunto de ejemplos (de 1 a 8) y atributos (de A1 a A4).

El mejor atributo es aquel que proporciona el menor error ponderado, que viene dado por la fórmula $\text{Error}_{(Ai)} = \sum \left(\frac{|E_i|}{|E|}\sigma_{(A_i)}\right)$. Para calcular las desviaciones estándar ($\sigma_{(Ai)}$) hay que calcular también los valores medios μ_i de cada atributo:

$$\mu_i = \frac{\sum valores\ de\ clase\ de\ los\ ejemplo\ que\ tengan\ un\ valor\ de\ atributo\ i\ determinado}{n^o\ de\ ejemplos\ con\ un\ valor\ de\ atributo\ i\ determinado}$$

$$\mu_{(A1=Alta)} = \frac{3500+5000}{2} = 4250$$

$$\mu_{(A1=Media)} = \frac{3300+5100+5200}{3} = 4533{,}33$$

$$\mu_{(A1=Baja)} = \frac{900+400+3400}{3} = 1566{,}66$$

Cálculo de la desviación estándar del atributo:

$\sigma_{(Ai)} = \sqrt{\sum \frac{(x_i - \mu_{Ai})^2}{n}}$ = (en el libro de texto hay ejemplo que no divide entre 'n', ver explicación en fe de erratas)

$$\sigma_{(A1 = Alta)} = \sqrt{\frac{((3500-4250)^2 + (5000-4250)^2)}{2}} = 750$$

$$\sigma_{(A1 = Media)} = \sqrt{\frac{((3300-4533{,}33)^2 + (5100-4533{,}33)^2 + (5200-4533{,}33)^2)}{3}} = 873{,}05$$

$$\sigma_{(A1 = Baja)} = \sqrt{\frac{((900-1566{,}66)^2 + (400-1566{,}66)^2 + (3400-1566{,}66)^2)}{3}} = 1312{,}33$$

Cálculo del error ponderado de A1:

$$\begin{aligned}
Error_{(A1)} &= \sum\left(\frac{|E_{A1}|}{|E|}\sigma_{(A1)}\right)\\
&= \frac{|E_{A1=Alto}|}{|E|}\sigma_{(A1=Alto)} + \frac{|E_{A1=Medio}|}{|E|}\sigma_{(A1=Medio)} + \frac{|E_{A1=Bajo}|}{|E|}\sigma_{(A1=Bajo)}\\
&= \frac{2}{8}\cdot 750 + \frac{3}{8}\cdot 873{,}05 + \frac{3}{8}\cdot 1312{,}33 = 10007{,}02
\end{aligned}$$

Ídem para A2:

$$\mu_{(A2 = Alto)} = \frac{5000+5100+5200}{3} = 5100$$

$$\mu_{(A2 = Medio)} = \frac{3300+3500+3400}{3} = 3400$$

$$\mu_{(A2 = Bajo)} = \frac{900+400}{2} = 650$$

$$\sigma_{(A2 = Alto)} = \sqrt{\frac{((5000-5100)^2 + (5100-5100)^2 + (5200-5100)^2)}{3}} = 81{,}65$$

$$\sigma_{(A2 = Medio)} = \sqrt{\frac{((3300-3400)^2 + (3500-3400)^2 + (3400-3400)^2)}{3}} = 81{,}65$$

$$\sigma_{(A2 = Bajo)} = \sqrt{\frac{((900-650)^2 + (400-650)^2)}{2}} = 250$$

$$Error_{(A2)} = \frac{3}{8}\cdot 81{,}65 + \frac{3}{8}\cdot 81{,}65 + \frac{2}{8}\cdot 250 = 123{,}74$$

Ídem para A3:

$$\mu_{(A3 = Sí)} = \frac{3300+5000+400+5100}{4} = 3450$$

$$\mu_{(A3 = No)} = \frac{3500+900+3400+5200}{4} = 3250$$

$$\sigma_{(A3 = Sí)} = \sqrt{\frac{((3300-3450)^2 + (5000-3450)^2 + (400-3450)^2 + (5100-3450)^2)}{4}} = 1900{,}66$$

$$\sigma_{(A3 = No)} = \sqrt{\frac{((3500-3250)^2 + (900-3250)^2 + (3400-3250)^2 + (5200-3250)^2)}{4}} = 1533{,}79$$

$$Error_{(A3)} = \frac{4}{8}\cdot 1900{,}66 + \frac{4}{8}\cdot 1533{,}79 = 1717{,}25$$

Ídem para A4:

$$\mu_{(A4 = Si)} = \frac{3300+3500+900+400+5100+5200}{6} = 3066{,}66$$

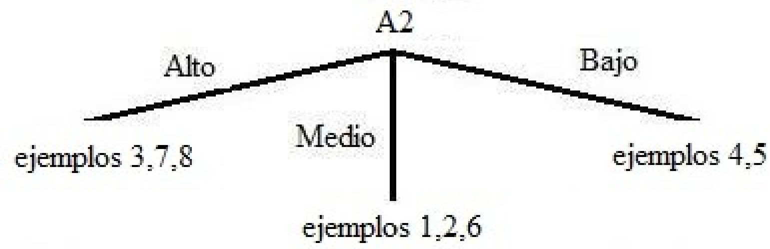
$$\mu_{(A4 = No)} = \frac{5000+3400}{2} = 4200$$

$$\sigma_{(A4 = Si)} = \sqrt{\frac{\left((3300-3066{,}66)^2+(3500-3066{,}66)^2+...+(5200-3066{,}66)^2\right)}{6}} = 1858{,}91$$

$$\sigma_{(A4 = No)} = \sqrt{\frac{\left((5000-4200)^2+(3400-4200)^2\right)}{2}} = 800$$

$$\text{Error}_{(A4)} = \frac{6}{8}\cdot 1858{,}91 + \frac{2}{8}\cdot 800 = 1594{,}18$$

Se elije el atributo que tenga el menor error, en este caso A2, con lo que se crea un nodo raíz en el que se distingue por este atributo y que tiene tres sucesores etiquetados con los valores del atributo: alto, medio y bajo, lo que nos da el siguiente árbol:



para cada sucesor, se comprueban las condiciones del enunciado (número de ejemplos mínimos por nodo y desviación estándar mínima)...

Para A2 = Alto, nº de ejemplos = 3 (ejemplos 3,7,8) ≥ 2, y $\sigma_{(A2 = Alto)} = 81{,}65 > 75$, se cumple por tanto, y este nodo puede ser expandido.

Para A2 = Medio, nº de ejemplos = 3 (ejemplos 1,2,6) ≥ 2, y $\sigma_{(A2 = Medio)} = 81{,}65 > 75$, se cumple por tanto, y este nodo puede ser expandido.

Para A2 = Bajo, nº de ejemplos = 2 (ejemplos 4,5) ≥ 2, y $\sigma_{(A2 = Bajo)} = 250 > 75$, se cumple por tanto, y este nodo puede ser expandido.

Por ello, habrá que seguir subdividiendo los sucesores con valores de A2 igual a *Alto*, *Medio* y *Bajo*:

Para el nodo de A2 = Alto, llamamos a construir-árbol-regresión(ejemplos 3,7,8; atributos A1,A2,A4; nodo A2 =Alto):

Cliente	A1	A3	A4	Consumo
3	Alta	Sí	No	5000
7	Media	Sí	Sí	5100
8	Media	No	Sí	5200

y buscamos el mejor atributo, aquel que tenga menor error:

$$\mu_{(A1 = Alta)} = \frac{5000}{1} = 5000$$

$$\mu_{(A1 = Media)} = \frac{5100+5200}{2} = 5150$$

$$\mu_{(A1 = Baja)} = 0$$

$$\sigma_{(A1 = Alta)} = \sqrt{\frac{((5000-5000)^2)}{1}} = 0$$

$$\sigma_{(A1 = Media)} = \sqrt{\frac{((5100-5150)^2+(5200-5150)^2)}{2}} = 50$$

$$\sigma_{(A1 = Baja)} = 0$$

$$\text{Error}_{(A1)} = 0 + \frac{2}{3}\cdot 50 + 0 = 33{,}33$$

$$\mu_{(A3 = Si)} = \frac{5000+5100}{2} = 5050$$

$$\mu_{(A3 = No)} = \frac{5200}{1} = 5200$$

$$\sigma_{(A3 = Si)} = \sqrt{\frac{((5000-5050)^2+(5100-5050)^2)}{2}} = 50$$

$$\sigma_{(A3 = No)} = \sqrt{\frac{((5200-5200)^2)}{1}} = 0$$

$$\text{Error}_{(A3)} = \frac{2}{3}\cdot 50 + 0 = 33{,}33$$

$$\mu_{(A4 = Si)} = \frac{5100+5200}{2} = 5050$$

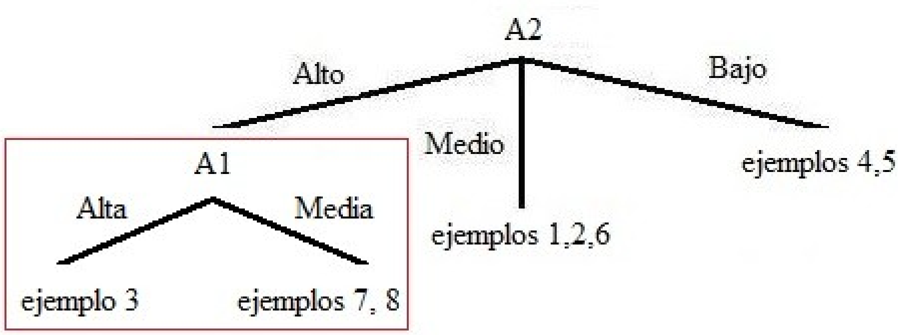
$$\mu_{(A4 = No)} = \frac{5000}{1} = 5000$$

$$\sigma_{(A4 = Si)} = \sqrt{\frac{((5100-5150)^2+(5200-5150)^2)}{2}} = 50$$

$$\sigma_{(A4 = No)} = \sqrt{\frac{((5000-5000)^2)}{1}}$$

$$\text{Error}_{(A4)} = \frac{2}{3}\cdot 50 + 0 = 33{,}33$$

Todos tienen el mismo error, elegimos uno aleatoriamente, por ejemplo A1, y tendríamos el árbol de regresión:



para cada sucesor, se comprueban las condiciones del enunciado (número de ejemplos mínimos por nodo y desviación estándar mínima)...

Para A1 = Alta, nº de ejemplos = 1 (ejemplos 3) ≥ 2, no se cumple la primera condición y, por tanto, este nodo puede ser expandido y se transforma en un nodo hoja.

Para A1 = Media, nº de ejemplos = 2 (ejemplos 7,8) ≥ 2, y $\sigma_{(A1\ =\ Media)} = 50 > 75$, con lo que la segunda condición no se cumple y, por tanto, este nodo no puede ser expandido y se transforma en un nodo hoja.

Para el nodo de A2 = Medio, llamamos a construir-árbol-regresión(ejemplos 1,2,6; atributos A1,A2,A4; nodo A2 =Medio):

Cliente	A1	A3	A4	Consumo
1	Media	Sí	Sí	3300
2	Alta	No	Sí	3500
6	Baja	No	No	3400

y buscamos el mejor atributo, aquel que tenga menor error:

$$\mu_{(A1=Alta)} = \frac{3500}{1} = 3500$$

$$\mu_{(A1=Media)} = \frac{3300}{1} = 3300$$

$$\mu_{(A1=Baja)} = \frac{3400}{1} = 3400$$

$$\sigma_{(A1=Alta)} = \sqrt{\frac{((3500-3500)^2)}{1}} = 0$$

$$\sigma_{(A1=Media)} = \sqrt{\frac{((3300-300)^2)}{1}} = 0$$

$$\sigma_{(A1=Baja)} = \sqrt{\frac{((3400-3400)^2)}{1}} = 0$$

$$\mathrm{Error}_{(A1)} = 0 + 0 + 0 = 0$$

$$\mu_{(A3 = Si)} = \frac{3300}{1} = 3300$$

$$\mu_{(A3 = No)} = \frac{3500+3400}{1} = 3450$$

$$\sigma_{(A3 = Si)} = \sqrt{\frac{((3300-3300)^2)}{1}} = 50$$

$$\sigma_{(A3 = No)} = \sqrt{\frac{((3500-3450)^2+(3400-3450)^2)}{2}} = 50$$

$$\text{Error}_{(A3)} = \frac{2}{3}\cdot 50 + 0 = 33{,}33$$

$$\mu_{(A4 = Si)} = \frac{5100+5200}{2} = 5050$$

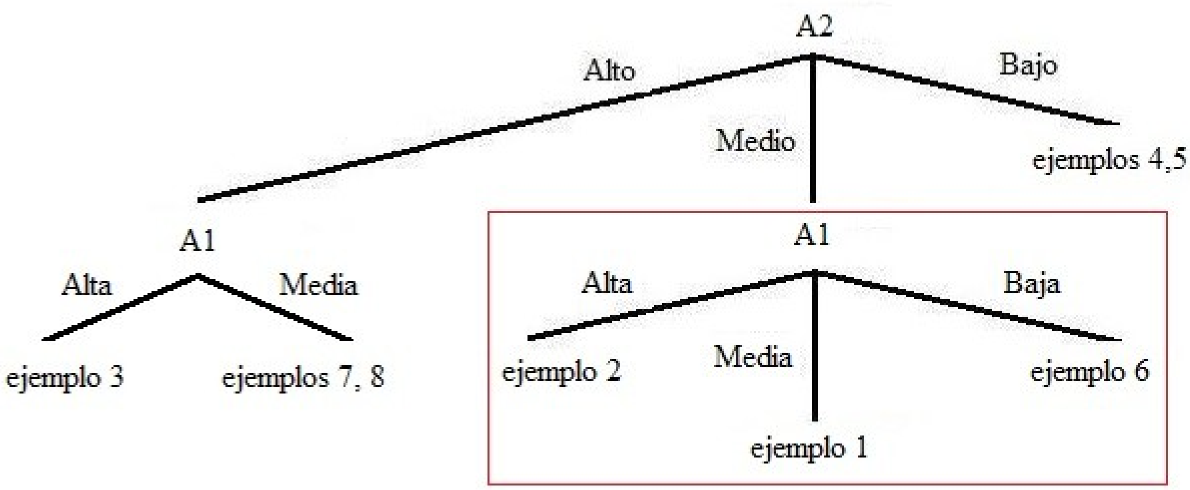
$$\mu_{(A4 = No)} = \frac{5000}{1} = 5000$$

$$\sigma_{(A4 = Si)} = \sqrt{\frac{((5100-5150)^2+(5200-5150)^2)}{2}} = 50$$

$$\sigma_{(A4 = No)} = \sqrt{\frac{((5000-5000)^2)}{1}}$$

$$\text{Error}_{(A4)} = 0 + \frac{2}{3}\cdot 50 = 33{,}33$$

El que menor error tiene es A1, por lo que lo expandimos, quedándonos el siguiente árbol de regresión:



para cada sucesor, se comprueban las condiciones del enunciado (número de ejemplos mínimos por nodo y desviación estándar mínima), pero todos tienen solo un ejemplo por nodo, por lo que no cumplen la primera condición y no se expanden, y se convierten en nodos hoja.

Para el nodo de A2 = Bajo, llamamos a construir-árbol-regresión(ejemplos 4,5; atributos A1,A2,A4; nodo A2 =Bajo):

Cliente	A1	A3	A4	Consumo
4	Baja	No	Sí	900
5	Baja	Sí	Sí	400

y buscamos el mejor atributo, aquel que tenga menor error:

$\mu_{(A1 = Alta)} = 0$

$\mu_{(A1 = Media)} = 0$

$\mu_{(A1 = Baja)} = \frac{900+400}{2} = 650$

$\sigma_{(A1 = Alto)} = 0$

$\sigma_{(A1 = Medio)} = 0$

$\sigma_{(A1 = Baja)} = \sqrt{\frac{((900-650)^2 + (400-650)^2)}{2}} = 250$

$\text{Error}_{(A1)} = 0 + 0 + \frac{2}{2} \cdot 250 = 250$

$\mu_{(A3 = Si)} = \frac{400}{1} = 400$

$\mu_{(A3 = No)} = \frac{900}{1} = 900$

$\sigma_{(A3 = Si)} = \sqrt{\frac{((400-400)^2)}{1}} = 0$

$\sigma_{(A3 = No)} = \sqrt{\frac{((900-900)^2)}{1}} = 0$

$\text{Error}_{(A3)} = 0 + 0 = 0$

$\mu_{(A4 = Si)} = \frac{900+400}{2} = 650$

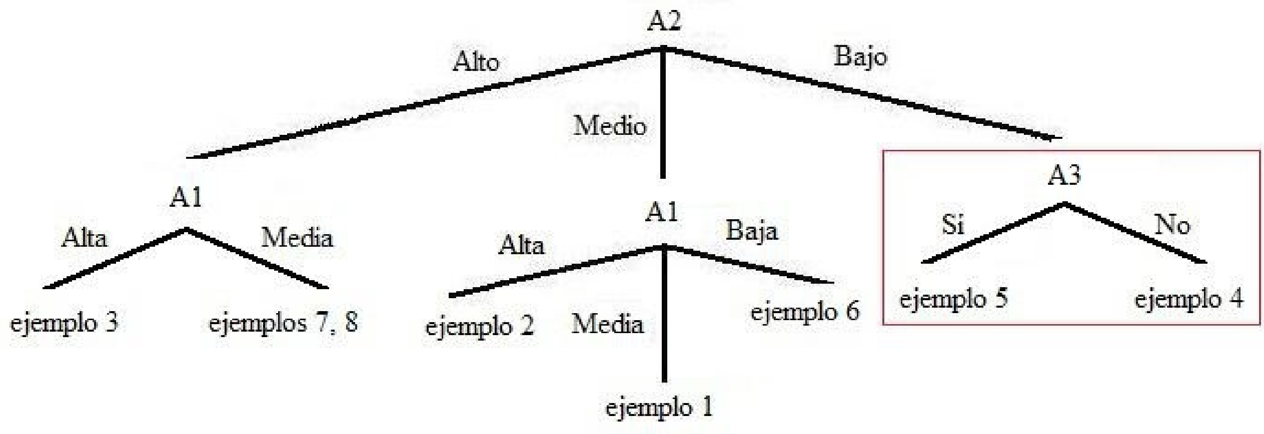
$\mu_{(A4 = No)} = 0$

$\sigma_{(A4 = Si)} = \sqrt{\frac{((900-650)^2 + (900-650)^2)}{2}} = 250$

$\sigma_{(A4 = No)} = 0$

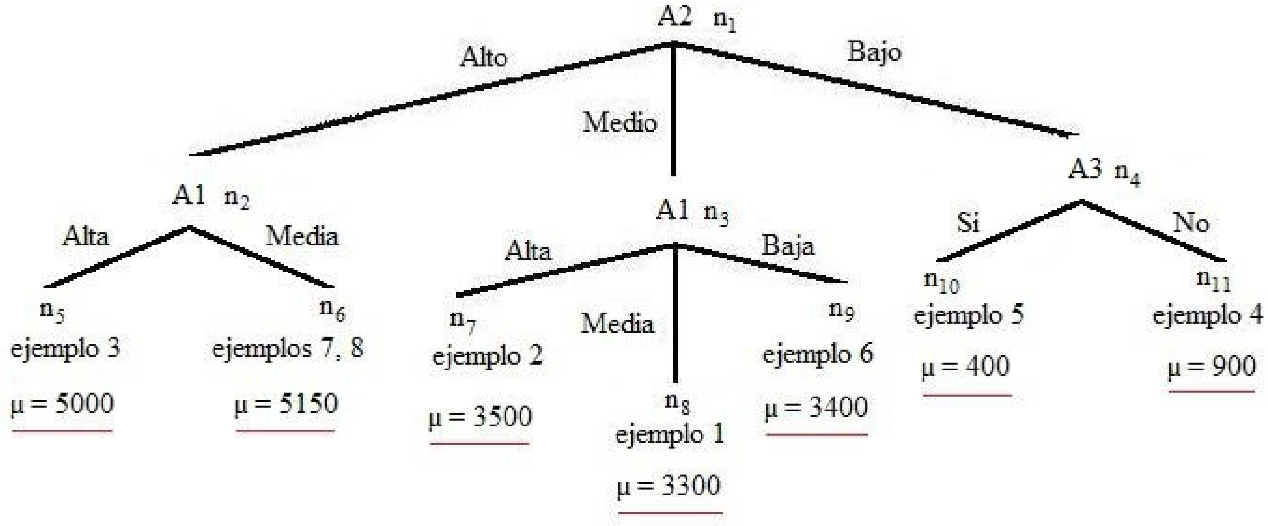
$\text{Error}_{(A4)} = \frac{2}{2} \cdot 250 + 0 = 250$

El atributo con menor error es A3, expandimos el árbol anterior y nos queda:



para cada sucesor, se comprueban las condiciones del enunciado (número de ejemplos mínimos por nodo y desviación estándar mínima), pero todos tienen solo un ejemplo por nodo, por lo que no cumplen la primera condición y no se expanden, y se convierten en nodos hoja.

Aquí se termina de calcular el árbol de regresión, numeramos también los nodos y se representa el modelo predictivo como el valor medio de la clase de los ejemplos de cada hoja:



15F2

PREGUNTA 1. Valoración: 2

En el contexto de los **mapas auto-organizados de Kohonen** conteste a las siguientes preguntas:

a) En el proceso de entrenamiento, dada una entrada, $\varepsilon=(\varepsilon_1,\dots, \varepsilon_n)$, describa cómo se determina la salida de la célula j-esima, τ_j, de la capa de competición.

b) En el proceso de entrenamiento, dada una entrada, $\varepsilon=(\varepsilon_1,\dots, \varepsilon_n)$, describa cómo se determina la célula ganadora de la capa de competición.

c) Justifique si la siguiente afirmación es verdadera o falsa: "En el proceso de entrenamiento, dada una entrada, $\varepsilon=(\varepsilon_1,\dots, \varepsilon_n)$, la fórmula para actualizar los pesos, μ_{ij}, de la célula ganadora, C_j, es: $\Delta\mu_{ij}=\alpha\cdot|\varepsilon_i - \mu_{ij}|$, donde i=1,…, n y la función |.| corresponde a la función valor absoluto"

a)

El modelo de mapas auto-organizados se aplica la función de distancia para calcular la salida de las células en la capa de competición.

$$\tau_j=d(\epsilon,\mu_j)$$

Siendo μ_j el vector de pesos correspondiente a la célula j-ésima.

La función de distancia que suele emplear es laeuclídea, obteniendo así la salida:

$$\tau_j=\sqrt{\sum_{i=1}^{n}(\epsilon-\mu_j)^2}$$

La entrada se propaga hasta la capa de competición, de forma que cada célula de la misma produce susalida

b)

Todas las células de la capa de competición, para una entrada dada, una vez calculada su salida, la comparan entre sí, para seleccionar aquella con el valor más pequeño (la menor distancia), de forma que ésta pasa a ser la célula ganadora.

c)

No es verdadera, ya que la fórmula para actualizar los pesos se calcula a partir de la siguiente regla hebbiana:

$$\Delta\mu_{ij}=\alpha_i(t)\cdot\tau_j(t)\cdot\sqrt{\sum_{i=1}^{n}(\epsilon-\mu_j)^2}$$

donde $\tau_j(t)=1$ toma el valor 1 para la célula ganadora, quedando pues:

$$\Delta\mu_{ij}=\alpha_i(t)\cdot\sqrt{\sum_{i=1}^{n}(\epsilon-\mu_j)^2}$$

16F1

PREGUNTA 1. Valoración: 2

En el contexto de las Redes neuronales auto-organizadas de Kohonen, la regla de aprendizaje tipo Hebb viene dada por:

$$\frac{d\mu_{ij}}{dt} = \alpha(t)\tau_j(t)[\varepsilon_i(t) - \mu_{ij}]$$

(a) Indique qué representan $d\mu_{ij}/dt$, μ_{ij}, $\varepsilon_i(t)$, $\alpha(t)$ y $\tau_j(t)$.
(b) Indique al menos una estrategia para definir $\alpha(t)$ (descartando la opción $\alpha(t)=K$, siendo K una constante).
(c) Si el método de aprendizaje utilizado para aprender los pesos de la red utiliza el esquema "la célula que gana se lo lleva todo", indique cómo afecta este esquema al cálculo de los parámetros indicados en el apartado (a).
(d) Si el método de aprendizaje utilizado para aprender los pesos de la red utiliza el denominado "mecanismo de vecindario", indique cómo afecta este esquema al cálculo de los parámetros indicados en el apartado (a).

a)

En la fórmula $\frac{d\mu_{ij}}{dt} = \alpha(t)\cdot\tau_j(t)\cdot[\epsilon_i(t) - \mu_{ij}]$:

$d\mu_{ij}$ /dt, es la variación en el tiempo del peso de una conexión entre las células *i* y *j*
μ_{ij}, es el peso de la conexión entre la célula i-ésima de la capa *k-1* y la célula j-ésima de la capa *k*
ϵ_i, es el valor de la componente *i* del vector de entrada y que representa el valor del atributo *i*
α(t), es la tasa de aprendizaje en un determinado instante *t*
τ_i, es la salida de la célula *i* de la capa de competición

b)

Se puede definir decrementano su valor una cantidad constante pequeña β tras cada ciclo completo de todos los patrones de aprendizaje, en la forma $\alpha(t+1) = \alpha(t) - \beta$
Se puede definir decrementando su valor siguiendo un esquema logarítmico. En ese caso, el decremento es muy elevado en las primeras iteraciones y se va reduciendo paulatinamente hasta que alcanza valores muy pequeños, con asíntota en el cero.

c)

$\frac{d\mu_{ij}}{dt} = \alpha(t)\cdot(\epsilon_i(t) - \mu_{ij}(t))$ si c_i es la célula ganadora, ó 0 en caso contrario

d)

$\frac{d\mu_{ik}}{dt} = \frac{\alpha(t)}{d(c_k, c_j)}\cdot(\epsilon_i(t) - \mu_{ij}(t))$ si c_i es la célula ganadora y $d(c_k, c_j) < \theta$ ó 0 en caso contrario

17F2

PREGUNTA 1. Valoración: 2.5

En el contexto de los **mapas auto-organizados de Kohonen** conteste a las siguientes preguntas:

a) En el proceso de entrenamiento, dada una entrada, $\varepsilon=(\varepsilon_1,\ldots, \varepsilon_n)$, describa cómo se calcula la salida de la célula j-esima, τ_j, de la capa de competición (antes de elegir la célula ganadora).

b) En el proceso de entrenamiento, dada una entrada, $\varepsilon=(\varepsilon_1,\ldots, \varepsilon_n)$, describa cómo se determina la célula ganadora de la capa de competición.

c) Justifique si la siguiente afirmación es verdadera o falsa: "En el proceso de entrenamiento (sin aplicar mecanismo de vecindario), dada una entrada, $\varepsilon=(\varepsilon_1,\ldots, \varepsilon_n)$, la fórmula para actualizar los pesos, μ_{ij}, de la célula ganadora, C_j, es: $\Delta\mu_{ij}=\alpha(t)\cdot(\varepsilon_i(t)- \mu_{ij}(t))^2$, donde i=1,…, n"

d) En el método de Kohonen, la tasa de aprendizaje, α, no es constante sino que varía con el tiempo, $\alpha(t)$. Describa los dos esquemas más usados para hacer variar el valor de dicha tasa.

e) Indique la expresión que permite actualizar los pesos, μ_{ij}, de la célula ganadora, C_j, cuando en el proceso de entrenamiento se incorpora el mecanismo de vecindario.

a)

La salida de una célula j (τ_j) en el modelo de Kohonen se podría calcular a través de la distancia euclídea:

$$\tau_j = d(\varepsilon, \mu_j) = \sqrt{\sum_{i=1}^{n} (\varepsilon_i - \mu_{ij})^2}$$

donde ε representa la entrada a la red y μ el peso de las conexiones entre neuronas.

b)

Todas las células de la capa de competición, para una entrada dada, una vez calculada su salida, la comparan entre sí, para seleccionar aquella con el valor más pequeño (la menor distancia), de forma que esta pasa a ser la célula ganadora.

c)

No exactamente la fórmula se calcula a partir de la regla hebbiana siguiente:

$$\Delta\mu_{ij} = \alpha(t)\cdot\tau_j(t)\cdot(\epsilon_i(t)-\mu_{ij}(t))$$

donde $\tau_j(t)$ toma el valor 1 (célula ganadora), quedando pues:

$$\Delta\mu_{ij} = \alpha(t)\cdot(\epsilon_i(t)-\mu_{ij}(t))$$

Sobra pues el "elevado al cuadrado".

d)

Los dos esquemas más usados para hacer variar la tasa de aprendizaje α son:

- Decrementando una cantidad constante pequeña β tras cada ciclo completo de todos los patrones de aprendizaje.
- Decrementando el valor siguiendo un esquema logarítmico, siendo éste muy elevado en las primeras iteraciones reduciéndose paulatinamente hasta que alcanza valores muy pequeños con una asíntota en el cero.

e)

La expresión que permite actualizar los pesos de la célula ganadora cuando en el proceso de entrenamiento se incorpora el mecanismo de vecindario sería:

$$\frac{d\mu_{ik}}{dt} = \begin{cases} \frac{\alpha(t)}{d(c_k, c_j)} \left(\epsilon_i(t) - \mu_{ij}(t)\right) & si\ célula\ c_i\ ganadora\ y\ d(c_k, c_j) < \theta \\ 0 & en\ caso\ contrario \end{cases}$$

donde $\boldsymbol{\alpha}$ es la tasa de aprendizaje, $\boldsymbol{d}$ la distancia de una célula a su vecina, $\boldsymbol{\mu}$ los pesos de las conexiones y $\boldsymbol{\theta}$ es el radio dentro del cual se considera vecinas a otras neuronas.

18F2 – Pregunta 1.d

PREGUNTA 1. Valoración: 2
Responda a las siguientes cuestiones:

a) Tras ejecutar el **algoritmo de eliminación de candidatos**, el valor del conjunto G={(x_1,A_{21},x_3)} y del conjunto S={(A_{11},A_{21},A_{31})}, donde A_{ij} corresponde al valor *j*-ésimo del atributo *i*-ésimo. Indique el conjunto de hipótesis que conforma el espacio de versiones (EV).

b) Justifique si en una **Red Neuronal Artificial (RNA),** que usa el algoritmo de retropropagación del error para aprender los pesos, se podría utilizar la función escalón como función de activación.

c) Muestre la función que expresa la salida, s_j, de una neurona perteneciente a una capa oculta de una **RNA**. Identifique todos los elementos expresados en dicha función.

d) En las **redes neuronales auto-organizadas de Kohonen**, los pesos de la célula ganadora, C_j, se actualizan de acuerdo a la siguiente regla hebbiana:

$$\frac{d\mu_{ij}}{dt} = \alpha(t)\tau_j(t)\left[\epsilon_i(t) - \mu_{ij}(t)\right]$$

Indique una función válida para actualizar el valor de la tasa de aprendizaje.

d)

Una función válida para actualizar el valor de la tasa de aprendizaje podría ser:

$\alpha\,(t + 1) = \alpha\,(t) - \beta$

que responde al esquema donde se decrementa el valor de la tasa de aprendizaje α en una cantidad pequeña β tras cada ciclo completo de todos los patrones de aprendizaje.

<u>PREGUNTA 1</u>. Valoración: 1

Describa los siguientes aspectos del **algoritmo de los mapas de auto-organización de Kohonen**:

a) *Bias*.

b) Cómo se comparte ante el ruido: en entradas y en estructura.

a) *Bias*:

- <u>del lenguaje</u>: es solamente aplicable a valores numéricos, lo que limita el espacio de hipótesis.
- <u>heurístico</u>: la distancia de cada ejemplo a las neuronas, que marcarán la pertenencia a una u otra categoría.

b) *Cómo se comporta el ruido*:

- <u>en entradas</u>: no se ven considerablemente afectados, ya que la red es capaz de eliminar el ruido tanto en aprendizaje como el de la fase de clasificación de nuevos ejemplos.
- <u>en estructura</u>: la eliminación de una célula en la capa de competición supondrá, en el peor de los casos, la desaparición de una de las categorías existentes, pero el modelo podría seguir clasificando los ejemplos de la categoría eliminada en en la categoría más parecida de las existentes en la red. La eliminación de células en la entrada puede clasificar ejemplos incorrectamente en el peor de los casos. SI la célula no es muy crítica, su desaparición no afectará mucho al proceso de clasificación. Lo mismo ocurre con la desaparición de conexiones.

15F1, 18F1, 18SR y 19F2

PREGUNTA 1. Valoración: 2

Describa el **algoritmo de retropropagación** (Redes Neuronales Artificiales) indicando:

a) Bias.
b) Cómo se comparte ante el ruido: en entradas y en estructura.
c) Complejidad: espacial y temporal.
d) Control de la tarea aprendida (Crítica/utilidad/valoración)
e) Dependencia del conocimiento del dominio.

A alto nivel, lo que realiza el aprendizaje en una RNA se puede resumir como:

1. Se introduce un nuevo ejemplo de entrenamiento en las entradas, del cual también se sabe la salida que debe producir.
2. Para aquellas células cuyas entradas sólo estén compuestas por las propias entradas de la red, se calcula su salida.
3. Se calcula la salida de aquellas células de las que se dispongan de todos sus valores de sus entradas.
4. Se repite el paso 3 hasta que todas las células hayan producido su salida
5. Se compara la salida producida por la red con la que debería haberse producido
6. En función de la coparación anterior, realizar el ajuste de todos los pesos de la RNA
7. Si todavía quedan ejemplos de entrenamiento, volver al paso 1.

Las **bias** de este algoritmo están en:
a) <u>la función de error elegida</u>, en la que todos los pasos del algoritmo están dirigidos a minimizar el error producido por la red para todos los ejemplos del conjunto de aprendizaje.
b) <u>el lenguaje de descripción empleado</u>, ya que solo los ejemplos con valores de atributos estríctamente numéricos pueden ser tratados.
c) <u>el espacio de hipótesis</u>, dado por la función matemática no lineal de las entradas y solo podrán ser aprendidas aquelllas hipótesis que se puedan describir en estos términos.
d) <u>el procedimiento del descenso del gradiente</u>, que guía la regla de aprendizaje.
e) <u>el conjunto de aprendizaje</u>, que debe ser representativo del problema a resolver, en el que todas las posibilidade relevantes estén representadas, ya que también ello guiará el aprendizaje.
f) <u>la inicialización inicial de los pesos</u>, aleatorio, y que condicionará todo el proceso.

El **tratamiento del ruido** se puede decribir:
<u>En entradas</u>: si el conjunto de ejemplos es muy grande, pequeñas modificaciones de los mismos no afectarán significativamente al resultado final del aprendizaje. Es reforzado por el hecho de que la red solo busca aproximaciones, no resultados exactos.
<u>En estructura</u>: las RNA usa esquemas de almacenamiento distribuidos, en lo que destaca la redundancia de la información, que permite un correcto funcionamiento de la red , aún cuando el sistema haya podido sufrir una destrucción parcial.

La **complejidad**:
El <u>espacio</u> ocupado por el sistema es fijo, dependiendo de la arquitectura de la red. Para una de n capas con m_i células en la capa i, existe un número c de conexiones igual a $\sum_{i=2}^{n} m_{i-1} \cdot m_i$
El <u>tiempo</u> depende del número de iteraciones i hasta que se alcanza el criterio de convergencia, y será $T = i \cdot p \cdot c \cdot (tc + tp)$, con p como el nº de ejemplos, c el número de conexiones, tc es el tiempo necesario para actualizar una conexión y tp es el tiempo en propagar la señal por una conexión.

El **control de la tarea aprendida**:
La valoración de la tarea aprendida se realiza a través de un conjunto de ejemplos de validación (de los que se conoce su salida), que se van introduciendo en el red y obteniendo las salidas de la red. De esa forma se calcula el error total del sistema como la media de los error cometido en cada ejemplo.
Este método es **independiente del conocimiento del dominio**.

17F1

PREGUNTA 1. Valoración: 2

En el contexto de las **redes neuronales artificiales (RNA)** conteste a las siguientes preguntas:

a) Justifique si la siguiente afirmación es verdadera o falsa: "El cálculo de la salida de la célula j-ésima de la capa k se calcula como:

$$s_j^k = \sum_{i=1}^{n(c_{k-1})} s_i^{k-1} W_{ij}^k$$

donde $n(c_{k-1})$ es el número de células de la capa k-1 y W_{ij}^k es el peso de la conexión que une la célula i-ésima de la capa k-1, cuya salida es s_i^{k-1}, con la célula cuya salida se está calculando".

b) Justifique si la siguiente afirmación es verdadera o falsa: "La actualización de los pesos de la red neuronal se realiza de acuerdo a la siguiente fórmula:

$$\omega_{ij}^k(t+1) = \omega_{ij}^k(t) + \mu\, s_i^{k-1}$$

donde $\omega_{ij}^k(t+1)$ y $\omega_{ij}^k(t)$ son los pesos asociados a la conexión que une la célula i-ésima de la capa k-1 con la célula j-ésima de la capa k en los instantes de tiempo $t+1$ y t, respectivamente; s_i^{k-1} es la salida de la célula i-ésima de la capa k-1; y μ es lo que se denomina tasa de aprendizaje".

c) Describa el tratamiento del ruido (en entradas y en estructura) en RNAs.

a)

Falsa. La salida de cualquier célula no solamente depende de las salidas de las células de la capa inmediatamente superior que están conectada con ella y ponderadas con los pesos de las conexiones, como se indica en la igualdad, sino que también hay que aplicarle una función no lineal de tipo umbral, conocida como función de activación.

b)

Falsa. La actualización de los pesos no dependen solo del valor inicial del peso de las conexiones que unan dos células, de la salida de las células de la capa anterior que se conectan a otra célula y de la ganancia o tasa de aprendizaje, sino también de un valor de incremento que depende de la derivada del error asociado a la célula.

c)

El **tratamiento del ruido** se puede decribir:

En entradas: si el conjunto de ejemplos es muy grande, pequeñas modificaciones de los mismos no afectarán significativamente al resultado final del aprendizaje. Es reforzado por el hecho de que la red solo busca aproximaciones, no resultados exactos.

En estructura: las RNA usa esquemas de almacenamiento distribuidos, en lo que destaca la redundancia de la información, que permite un correcto funcionamiento de la red , aún cuando el sistema haya podido sufrir una destrucción parcial.

18F2 – Pregunta 1.b, 1.c

PREGUNTA 1. Valoración: 2
Responda a las siguientes cuestiones:

a) Tras ejecutar el **algoritmo de eliminación de candidatos**, el valor del conjunto G={(x_1,A_{21},x_3)} y del conjunto S={(A_{11},A_{21},A_{31})}, donde A_{ij} corresponde al valor *j*-ésimo del atributo *i*-ésimo. Indique el conjunto de hipótesis que conforma el espacio de versiones (EV).

b) Justifique si en una **Red Neuronal Artificial (RNA),** que usa el algoritmo de retropropagación del error para aprender los pesos, se podría utilizar la función escalón como función de activación.

c) Muestre la función que expresa la salida, s_j, de una neurona perteneciente a una capa oculta de una **RNA**. Identifique todos los elementos expresados en dicha función.

d) En las **redes neuronales auto-organizadas de Kohonen**, los pesos de la célula ganadora, C_j, se actualizan de acuerdo a la siguiente regla hebbiana:

$$\frac{d\mu_{ij}}{dt} = \alpha(t)\tau_j(t)\left[\epsilon_i(t) - \mu_{ij}(t)\right]$$

Indique una función válida para actualizar el valor de la tasa de aprendizaje.

b)

No podría usar la función escalón ya que la función de activación debe ser derivable en todo el dominio y la función escalón no es derivable, al no ser una función continua.

c)

La salida s_j de una neurona perteneciente a una capa oculta de una RNA viene dada por:

$$s_j = F\left(\sum_{i=0}^{n} s_i \cdot W_{ij}\right)$$, donde:

i	varía de cero a n indicando las células conectadas a j
s_j	salida de la célula i-ésima de la capa anterior
W_{ij}	Peso de la conexión entre la célula i y j
F(x)	función no lineal de tipo umbral

14F1

PREGUNTA 2. Valoración: 4

Se quiere construir un **clasificador Naïve-Bayes** para discriminar si un paciente tiene una determinada patología a partir del resultado de pruebas clínicas (presión arterial, número de leucocitos) y de la existencia o no de un determinado síntoma (dolor de cabeza). Para aprender el clasificador, se utilizarán los datos de la tabla 1. Se pide:

(a) Calcular las diferentes tablas de probabilidad y distribuciones de normalidad necesarias para construir el clasificador.

(b) Clasifique un paciente, con valores [PA=Alta, NL=9.0, DC=sí], como patológico o no, de acuerdo al clasificador construido.

Muy importante: Deberá describir y justificar la forma de obtener cada una de las probabilidades.

Paciente	Presión Arterial	Número Leucocitos (mill/mm^3)	Dolor de cabeza	**PATOLOGÍA**
1	Alta	8.1	No	**sí**
2	Alta	2.1	No	**no**
3	Normal	8.2	No	**no**
4	Alta	13.0	No	**sí**
5	Alta	14.1	Sí	**sí**
6	Normal	7.5	No	**no**
7	Normal	4.0	No	**no**
8	Normal	2.1	Sí	**no**
9	Alta	3.1	Sí	**no**

Tabla 1

a)

Probabilidades a priori:

Valor de atributo clase (Patología)	$P(C_k)$, k = 1,2
C_1(Sí)	3/9 = 0,33
C_2(No)	6/9 = 0,66

Probabilidades condicionadas (atributos discretos):

Presión Arterial	$P(PA_{Alta} \mid C_k)$, k = 1,2	$P(PA_{Normal} \mid C_k)$, k = 1,2
C_1(Sí)	$3/3 = 1 \rightarrow 0{,}95^*$	$0/3 = 0 \rightarrow 0{,}05^*$
C_2(No)	2/6 = 0,33	4/6 = 0,66

*Corrección de la estimación de la probabilidad

Dolor de Cabeza	$P(DC_{Sí} \mid C_k)$, k = 1,2	$P(DC_{No} \mid C_k)$, k = 1,2
C_1(Sí)	1/3 = 0,33	2/3 = 0,66
C_2(No)	2/6 = 0,33	4/6 = 0,66

Probabilidades condicionadas (atributos contínuos):

Número de Leucocitos (NL) / C_k = Sí

Media, $\mu = \frac{\sum x_i}{n} = \frac{8,1+13,0+14,1}{3} = 11,73$

Desviación típica, $\sigma = \sqrt{\frac{\left(\sum (x_i-\mu)^2\right)}{(n-1)}} = \sqrt{\frac{\left((8,1-11,73)^2+\ldots+(14,1-11,73)^2\right)}{2}} = 3,19$

Número de Leucocitos (NL) / C_k = No

Media, $\mu = \frac{\sum x_i}{n} = \frac{2,1+8,2+7,5+4,0+2,1+3,1}{6} = 4,5$

Desviación típica, $\sigma = \sqrt{\frac{\left(\sum (x_i-\mu)^2\right)}{(n-1)}} = \sqrt{\frac{\left((2,1-4,5)^2+(8,2-4,5)^2+\ldots+(3,1-4,5)^2\right)}{5}} = 2,70$

Número de Leucocitos (NL). Distribución de normalidad:

NL	**μ**	**σ**	**$1/\sqrt{(2\pi\sigma^2)}$**	**$1/(2\sigma^2)$**	**$g(x,\mu,\sigma) = (1/\sqrt{(2\pi\sigma^2)}) \cdot \exp[-1/(2\sigma^2)\cdot(x-\mu)^2]$**
C_1(Sí)	11.73	3.19	0.12506	0.04913	$0,12506 \cdot \exp[-0,04913 \cdot (x-11,73)^2]$
C_2(No)	4.5	2.7	0.14775	0,0685	$0,14775 \cdot \exp[-0,0685 \cdot (x-4,5)^2]$

b)

Clasificación de paciente X = [PA = Alta, NL = 9.0, DC = Sí]

Por teorema de Bayes, $P(X) = \max (P(X \mid C_{sí}), P(X \mid C_{no}))$:

$P(X \mid C_{Sí}) = P(PA_{Alta} \mid C_{sí}) \cdot P(NL_{9.0} \mid C_{sí}) \cdot P(DC_{Sí} \mid C_{sí}) \cdot P(C_{sí})$
$= 0.95 \cdot 0.12506 \cdot \exp[-0.04913 \cdot (9.0-11.73)^2] \cdot 0.33 \cdot 0.33$
$= 0.00896$

$P(X \mid C_{No}) = P(PA_{Alta} \mid C_{No}) \cdot P(NL_{9.0} \mid C_{No}) \cdot P(DC_{Sí} \mid C_{No}) \cdot P(C_{No}) =$
$= 0.33 \cdot 0,14775 \cdot \exp[-0,0685 \cdot (9.0-4,5)^2] \cdot 0.33 \cdot 0.66 =$
$= 0.00265$

$P(X) = \max (P(X \mid C_{sí}), P(X \mid C_{no})) = \max (0.00896, 0.00265) = 0.00896 = P(C_{sí})$

El paciente X se clasifica como normal.

14F2 – Pregunta 2

PREGUNTA 2. Valoración: 4

Se quiere construir un **clasificador Naïve-Bayes** para discriminar si un paciente tiene una determinada patología a partir del resultado de pruebas clínicas (presión arterial, recuento de leucocitos) y de la existencia o no de un determinado síntoma (dolor de cabeza). Para aprender el clasificador, se utilizarán los datos de la tabla 1. Se pide:

(a) Calcular las diferentes tablas de probabilidad y distribuciones de normalidad necesarias para construir el clasificador.

(b) Clasifique un paciente, con valores [PA=120, RL=Normal, DC=sí], como patológico o normal, de acuerdo al clasificador construido.

Muy importante: Deberá describir y justificar la forma de obtener cada una de las probabilidades.

Paciente	Presión Arterial (mm de Hg)	Recuento de Leucocitos	Dolor de cabeza	**PATOLOGÍA**
1	100	Bajo	sí	**no**
2	125	Bajo	sí	**no**
3	145	Normal	no	**sí**
4	135	Bajo	no	**no**
5	100	Normal	no	**no**
6	140	Alto	no	**sí**
7	138	Alto	sí	**sí**
8	110	Normal	no	**no**
9	105	Bajo	no	**no**

Tabla 1

a)

Probabilidades a priori:

Valor de atributo clase (Patología)	$P(C_k)$, k = 1,2
C_1(Sí)	3/9 = 0,33
C_2(No)	6/9 = 0,66

Probabilidades condicionadas (atributos discretos):

Recuento Leucocitos	$P(RL_{Alto} \mid C_k)$, k = 1,2	$P(RL_{Normal} \mid C_k)$, k = 1,2	$P(RL_{Bajo} \mid C_k)$, k = 1,2
C_1(Sí)	2/3 = 0,66 → 0,66 · (1-0,05) = 0,63333	1/3 = 0,33 → 0,33 · (1-0,05) = 0,31666	0/3 = 0 → 0,05
C_2(No)	0/6 = 0 → 0,05	2/6 → 0,33 0,33 · (1-0,05) = 0,31666	4/6 → 0,66 0,66 · (1-0,05) = 0,63333

Se ha normalizado C_k para que $P(RL_{Alto} \mid C_k) + P(RL_{Normal} \mid C_k) + P(RL_{Bajo} \mid C_k)$ sea 1

Dolor de Cabeza	$P(DC_{Sí} \mid C_k)$, k = 1,2	$P(DC_{No} \mid C_k)$, k = 1,2
C_1(Sí)	1/3 → 0,33	2/3 → 0,66
C_2(No)	2/6 → 0,33	4/6 → 0,66

Probabilidades condicionadas (atributos contínuos):

Presión Arterial (PA) / C_k = Sí

$$\text{Media, } \mu = \frac{\sum x_i}{n} = \frac{145+140+138}{3} = 141$$

$$\text{Desviación típica, } \sigma = \sqrt{\frac{\left(\sum (x_i-\mu)^2\right)}{(n-1)}} = \sqrt{\frac{\left((145-141)^2+(140-141)^2+(138-141)^2\right)}{2}} = 3{,}6055$$

Presión Arterial (PA) / C_k = No

$$\text{Media, } \mu = \frac{\sum x_i}{n} = \frac{100+125+135+100+110+105}{6} = 112{,}5$$

$$\text{Desviación típica, } \sigma = \sqrt{\frac{\left(\sum (x_i-\mu)^2\right)}{(n-1)}} = \sqrt{\frac{\left((100-112{,}5)^2+(125-112{,}5)^2+\ldots\right)}{5}} = 14{,}4048$$

Presión Arterial (PA). Distribución de normalidad:

PA	μ	σ	$1/\sqrt{(2\pi\sigma^2)}$	$1/(2\sigma^2)$	$g(x,\mu,\sigma) = (1/\sqrt{(2\pi\sigma^2)}) \cdot \exp[-1/(2\sigma^2)\cdot(x-\mu)^2]$
C_1(Sí)	141	3.6055	0.11064	0.03846	$0{,}11064 \cdot \exp[-0{,}03846 \cdot (x-141)^2]$
C_2(No)	112.5	14.4048	0.02769	0.00241	$0.02769 \cdot \exp[-0{,}002409 \cdot (x-112{,}5)^2]$

b)

Clasificación de paciente X = [PA = 120, RL = Normal, DC = Sí]

Por teorema de Bayes, $P(X) = \max (P(X \mid C_{sí}), P(X \mid C_{no}))$:

$P(X \mid C_{Sí})$ $= P(PA_{120} \mid C_{sí}) \cdot P(RL_{Normal} \mid C_{sí}) \cdot P(DC_{Sí} \mid C_{sí}) \cdot P(C_{sí}) =$
$= 0.11064 \cdot \exp[-0.03846 \cdot (120-141)^2] \cdot 0{,}31666 \cdot 0.33 \cdot 0.33 =$
$= 1.6425 \cdot 10^{-10}$

$P(X \mid C_{No})$ $= P(PA_{120} \mid C_{No}) \cdot P(RL_{Normal} \mid C_{No}) \cdot P(DC_{Sí} \mid C_{no}) \cdot P(C_{No}) =$
$= 0.02769 \cdot \exp[-0{,}002409 \cdot (120-112{,}5)^2] \cdot 0{,}31666 \cdot 0.33 \cdot 0.66 =$
$= 1.6676 \cdot 10^{-10}$

$P(X) = \max (P(X \mid C_{si}), P(X \mid C_{no})) = \max (1.6425 \cdot 10^{-10}, 1.6676 \cdot 10^{-10}) = 1.6676 \cdot 10^{-10} = P(C_{No})$
El paciente X se clasifica como normal.

15F2

PREGUNTA 4. Valoración: 3

Se quiere construir un **clasificador Naïve-Bayes** que permita establecer la concesión o no de un préstamo hipotecario a los clientes de un banco. Para ello, se dispone de una serie de indicadores como los mostrados en la tabla inferior. Se pide:

(a) Calcular las diferentes tablas de probabilidad y distribuciones de normalidad necesarias para construir el clasificador.

(b) Clasifique un cliente, con valores [TT=funcionario, C=sí, IA=90.0,], de acuerdo al clasificador construido.

Cliente	Tipo de trabajador	Casado	Ingresos Anuales (miles euros)	**Conceder hipoteca**
1	Funcionario	No	80.1	**sí**
2	Funcionario	No	20.1	**no**
3	Autónomo	No	80.2	**no**
4	Funcionario	No	130.0	**sí**
5	Funcionario	Sí	140.1	**sí**
6	Autónomo	No	70.5	**no**
7	Autónomo	No	40.0	**no**
8	Autónomo	Sí	20.1	**no**
9	Funcionario	Sí	30.1	**no**

a)

Probabilidades a priori:

Valor de atributo clase (Conceder Hipoteca)	$P(C_k)$, k = 1,2
C_1(Sí)	3/9 = 0,33
C_2(No)	6/9 = 0,66

Probabilidades condicionadas (atributos discretos):

Tipo Trabajador	$P(TT_{Funcionario} \mid C_k)$, k = 1,2	$P(TT_{Autónomo} \mid C_k)$, k = 1,2
C_1(Sí)	3/3 = 1 → 0,95*	0/3 = 0 → 0,05*
C_2(No)	2/6 = 0,33	4/6 = 0,66

*Corrección de la estimación de la probabilidad

Casado	$P(CAS_{Sí} \mid C_k)$, k = 1,2	$P(CAS_{No} \mid C_k)$, k = 1,2
C_1(Sí)	1/3 = 0,33	2/3 = 0,66
C_2(No)	2/6 = 0,33	4/6 = 0,66

Probabilidades condicionadas (atributos contínuos):

Ingresos Anuales (IA) / C_k = Sí

Media, $\mu = \frac{\sum x_i}{n} = \frac{80{,}1+130{,}0+140{,}1}{3} = 116{,}73$

Desviación típica, $\sigma = \sqrt{\frac{(\sum(x_i-\mu)^2)}{(n-1)}} = \sqrt{\frac{((80{,}1-116{,}73)^2+...+(140-116{,}73)^2)}{2}} = 32{,}1248$

Ingresos Anuales (IA) / C_k = No

Media, $\mu = \frac{\sum x_i}{n} = \frac{20{,}1+80{,}2+70{,}5+40{,}0+20{,}1+30{,}1}{6} = 43{,}5$

Desviación típica, $\sigma = \sqrt{\frac{(\sum(x_i-\mu)^2)}{(n-1)}} = \sqrt{\frac{((20{,}1-43{,}5)^2+...+(30{,}1-43{,}5)^2)}{5}} = 25{,}9338$

Ingresos anuales (IA). Distribución de normalidad:

IA	μ	σ	1/√(2πσ²)	1/(2σ²)	g(x,μ,σ) = (1/√(2πσ²)) · exp[-1/(2σ²)·(x-μ)²]
C_1(Sí)	116.73	32.1248	0.01463	0.00048	0,01463 · exp[-0,00048 · (x-116,73)²]
C_2(No)	43.5	25.9338	0.01538	0.00074	0.01538 · exp[-0,00074 · (x-43,5)²]

b)

Clasificación de paciente X = [TT = Funcionario, CAS = Sí, IA = 90.0]

Por teorema de Bayes, P(X) = max (P(X | $C_{sí}$),P(X | C_{no})):

P(X | $C_{Sí}$) = P($TT_{Funcionario}$ | $C_{sí}$) · P($CAS_{Sí}$ | $C_{sí}$) · P($IA_{90,0}$ | $C_{sí}$) · P($C_{sí}$) =
= 0.95 · 0.33 · 0.01463 · exp[-0.00048 · (90,0-116,73)²] · 0,33 =
= 1.0741 · 10^{-3}

P(X | C_{No}) = P($TT_{Funcionario}$ | C_{No}) · P($CAS_{Sí}$ | C_{No}) · P($IA_{90,0}$ | C_{no}) · P(C_{No}) =
= 0.33 · 0.33 · 0.01538 · exp[-0,00074 · (90,0-43,5)²] · 0.66 =
= 2.2316 · 10^{-4}

P(X) = max (P(X | C_{si}),P(X | C_{no})) = max (1.0741 · 10^{-3}, 2.2316 · 10^{-4}) = 1.0741 · 10^{-3} = P($C_{Sí}$)

El cliente X se clasifica como apto para conceder préstamo.

15SO

PREGUNTA 1. Valoración: 2

Dentro del contexto de los **clasificadores NaiveBayes** describir y justificar:

(a) Bias.

(b) Cómo les afecta el ruido en entradas y en estructura.

(c) Su complejidad espacial y temporal.

a) *Bias*

El uso del Teorema de Bayes y el cambio de representación que realiza esta técnica de información sibólica y numérica a información numérica.

b) *Cómo les afecta el ruido en entradas y en estructura*

En entradas: la estadística suaviza el ruido que se produzca en los datos de entrada

En estructura: el cambio de una probabilidad en un elemento de las matrices puede afecar a la clasificación final, aunque al verse multiplicado por otros términos numéricos, estos pueden hacer que el resultado no sea tan diferente como si no hubiera fallado el elemento.

c) *Su complejidad espacial y temporal*

Espacio: Siendo A_D : nº atributos discretos, A_C : nº atributos contínuos, K el nº de clases y V el nº máximo de valores para los atributos discretos, se definen 3 matrices que ocupan K, $A_D \cdot V \cdot K$ y $A_C \cdot 2 \cdot K$. Por tanto el espacio total será $K + A_D \cdot V \cdot K + A_C \cdot 2 \cdot K$

Tiempo:

El cálculo de la matriz de clase es proporcional a N + K (N: nº ejemplos)

El de las matrices de probabilidades de atributos discretos son proporcionales a $A_D \cdot V \cdot K \cdot N$

El de las matrices de probabilidades de atributos contínuos son proporcionales a $A_C \cdot K \cdot N$

Por tanto, el tiempo total será proporcional a la suma de los anteriores.

16F1

PREGUNTA 3. Valoración: 3

Se quiere construir un **clasificador Naïve-Bayes** que permita establecer la concesión o no de una ayuda estatal a los habitantes de una comunidad autónoma en función de una seria de indicadores I_i, con i=1,2, 3. Conocida la información de la tabla 2, se pide:

(a) Calcular las diferentes tablas de probabilidad y distribuciones de normalidad necesarias para construir el clasificador.

(b) Clasifique un cliente, con valores [I1=0.90, I2=A, I3=sí], de acuerdo al clasificador construido.

Habitante	I_1	I_2	I_3	Conceder ayuda
1	0.81	A	No	sí
2	0.21	A	No	no
3	0.82	B	No	no
4	1.30	A	No	sí
5	1.41	A	Sí	sí
6	0.75	B	No	no
7	0.40	B	No	no
8	0.21	B	Sí	no
9	0.31	A	Sí	no

Tabla 2

a)

Probabilidades a priori:

Valor de atributo clase (Patología)	$P(C_k)$, k = 1,2
C_1(Sí)	3/9 = 0,33
C_2(No)	6/9 = 0,66

Probabilidades condicionadas (atributos discretos):

I_2	$P(I_2 = A \mid C_k)$, k = 1,2	$P(I_2 = B \mid C_k)$, k = 1,2
C_1(Sí)	$3/3 = 1 \rightarrow 0{,}95^*$	$0/3 = 0 \rightarrow 0{,}05^*$
C_2(No)	$2/6 \rightarrow 0{,}33$	$4/6 \rightarrow 0{,}66$

*Corrección de la estimación de la probabilidad

I_3	$P(I_3 = Sí \mid C_k)$, k = 1,2	$P(I_3 = No \mid C_k)$, k = 1,2
C_1(Sí)	$1/3 \rightarrow 0{,}33$	$2/3 \rightarrow 0{,}66$
C_2(No)	$2/6 \rightarrow 0{,}33$	$4/6 \rightarrow 0{,}66$

Probabilidades condicionadas (atributos contínuos):

I_1 / C_k = Sí

Media, $\mu = \frac{\sum x_i}{n} = \frac{0{,}81+1{,}30+1{,}41}{3} = 1{,}17$

Desviación típica, $\sigma = \sqrt{\frac{\left(\sum (x_i-\mu)^2\right)}{(n-1)}} = \sqrt{\frac{\left((0{,}81-1{,}17)^2+...+(1{,}41-1{,}17)^2\right)}{2}} = 0{,}3194$

I_1 / C_k = No

Media, $\mu = \frac{\sum x_i}{n} = \frac{0{,}21+0{,}82+0{,}75+0{,}40+0{,}21+0{,}31}{6} = 0{,}45$

Desviación típica, $\sigma = \sqrt{\frac{\left(\sum (x_i-\mu)^2\right)}{(n-1)}} = \sqrt{\frac{\left((0{,}21-0{,}45)^2+...+(0{,}31-0{,}45)^2\right)}{5}} = 0{,}2698$

I_1 . Distribución de normalidad:

I_1	**μ**	**σ**	**$1/\sqrt{(2\pi\sigma^2)}$**	**$1/(2\sigma^2)$**	**$g(x,\mu,\sigma) = (1/\sqrt{(2\pi\sigma^2)}) \cdot \exp[-1/(2\sigma^2)\cdot(x-\mu)^2]$**
C_1(Sí)	1.17	0.3194	1.24903	4.90117	$1{,}24903 \cdot \exp[-4{,}90117 \cdot (x-1{,}17)^2]$
C_2(No)	0.45	0.2698	1.47865	6.86888	$1.47865 \cdot \exp[-6{,}86888 \cdot (x-0{,}45)^2]$

b)

Clasificación de paciente X = [I_1 = 0.90, I_2 = A, I_3 = Sí]

Por teorema de Bayes, P(X) = max (P(X | C_{si}),P(X | C_{no})):

P(X | C_{Si}) = P(I_1 = 0.90 | C_{si}) · P(I_2 = A | C_{si}) · P(I_3 = Sí | C_{si}) · P(C_{si}) =
= $1{,}24903 \cdot \exp[-4{,}90117 \cdot (0{,}90-1{,}17)^2] \cdot 0.95 \cdot 0.33 \cdot 0{,}33$ =
= 0.09039

P(X | C_{No}) = P(I_1 = 0.90 | C_{No}) · P(I_2 = A | C_{No}) · P(I_3 = Sí | C_{no}) · P(C_{No}) =
= $1.47865 \cdot \exp[-6{,}86888 \cdot (0{,}90-0{,}45)^2] \cdot 0.33 \cdot 0.33 \cdot 0.66$ =
= 0.02644

P(X) = max (P(X | C_{si}),P(X | C_{no})) = max (0.09039, 0.02644) = 0.09039 = P(C_{Si})

El cliente X se clasifica como apto para conceder ayuda.

16F2

PREGUNTA 4. Valoración: 3

Se quiere construir un **clasificador Naïve-Bayes** para discriminar si un paciente tiene una determinada patología a partir del resultado de dos pruebas clínicas (PC_1 y PC_2) y de la existencia o no de un determinado síntoma (S_1). Para aprender el clasificador, se utilizarán los datos de la tabla 2. Se pide:

(a) Calcular las diferentes tablas de probabilidad y distribuciones de normalidad necesarias para construir el clasificador.

(b) Clasifique un paciente, con valores [PC_1=Alta, PC_2=9.0, S_1=sí], como patológico o no, de acuerdo al clasificador construido.

Paciente	S_1	PC_1	PC_2	**PATOLOGÍA**
1	No	Alta	2.1	**no**
2	No	Normal	8.2	**no**
3	No	Normal	7.5	**no**
4	No	Normal	4.0	**no**
5	Sí	Normal	2.1	**no**
6	Sí	Alta	3.1	**no**
7	No	Alta	8.1	**sí**
8	No	Alta	13.0	**sí**
9	Sí	Alta	14.1	**sí**

Tabla 2

a)

Probabilidades a priori:

Valor de atributo clase (Patología)	$P(C_k)$, k = 1,2
C_1(Sí)	3/9 = 0,33
C_2(No)	6/9 = 0,66

Probabilidades condicionadas (atributos discretos):

S_1	$P(S_1 = \text{Sí} \mid C_k)$, k = 1,2	$P(S_1 = \text{No} \mid C_k)$, k = 1,2
C_1(Sí)	1/3 = 0,33	2/3 = 0,66
C_2(No)	2/6 = 0,33	4/6 = 0,66

PC_1	$P(PC_1 = \text{Alta} \mid C_k)$, k = 1,2	$P(PC_1 = \text{Normal} \mid C_k)$, k = 1,2
C_1(Sí)	3/3 = 1 → 0,95*	0/3 = 0 → 0,05*
C_2(No)	2/6 = 0,33	4/6 = 0,66

*Corrección de la estimación de la probabilidad

Probabilidades condicionadas (atributos contínuos):

PC_2 / C_k = Sí

Media, $\mu = \frac{\sum x_i}{n} = \frac{8{,}1+13{,}0+14{,}1}{3} = 11{,}73$

Desviación típica, $\sigma = \sqrt{\frac{\left(\sum (x_i-\mu)^2\right)}{(n-1)}} = \sqrt{\frac{\left((8{,}1-11{,}73)^2+\ldots+(14{,}1-11{,}73)^2\right)}{2}} = 3{,}19$

PC_2 / C_k = No

Media, $\mu = \frac{\sum x_i}{n} = \frac{2{,}1+8{,}2+7{,}5+4{,}0+2{,}1+3{,}1}{6} = 4{,}5$

Desviación típica, $\sigma = \sqrt{\frac{\left(\sum (x_i-\mu)^2\right)}{(n-1)}} = \sqrt{\frac{\left((2{,}1-4{,}5)^2+\ldots+(3{,}1-4{,}5)^2\right)}{5}} = 2{,}69$

PC_2 . Distribución de normalidad:

PC_2	μ	σ	$1/\sqrt{(2\pi\sigma^2)}$	$1/(2\sigma^2)$	$g(x,\mu,\sigma) = (1/\sqrt{(2\pi\sigma^2)}) \cdot \exp[-1/(2\sigma^2)\cdot(x-\mu)^2]$
C_1(Sí)	11.73	3.19	0.12506	0.04913	$0{,}12506 \cdot \exp[-0{,}04913 \cdot (x-11{,}73)^2]$
C_2(No)	4.5	2.69	0.14830	0.06909	$0{,}14830 \cdot \exp[-0{,}06909 \cdot (x-4{,}5)^2]$

b)

Clasificación de paciente X = [PC_1 = Alta, PC_2 = 9.0, S_1 = Sí]

Por teorema de Bayes, $P(X) = \max(P(X \mid C_{si}), P(X \mid C_{no}))$:

$P(X \mid C_{Si})$ $= P(PC_1 = \text{Alta} \mid C_{si}) \cdot P(PC_2 = 9{,}0 \mid C_{si}) \cdot P(S_1 = \text{Sí} \mid C_{si}) \cdot P(C_{si}) =$
$= 0.95 \cdot 0{,}12506 \cdot \exp[-0{,}04913 \cdot (9{,}0-11{,}73)^2] \cdot 0.33 \cdot 0{,}33 =$
$= 0.07173$

$P(X \mid C_{No})$ $= P(PC_1 = \text{Alta} \mid C_{No}) \cdot P(PC_2 = 9{,}0 \mid C_{No}) \cdot P(S_1 = \text{Sí} \mid C_{no}) \cdot P(C_{No}) =$
$= 0.33 \cdot 0{,}14830 \cdot \exp[-0{,}06909 \cdot (9{,}0-4{,}5)^2] \cdot 0.33 \cdot 0.66 =$
$= 0.01774$

$P(X) = \max(P(X \mid C_{si}), P(X \mid C_{no})) = \max(0.07173, 0.01774) = 0.07173 = P(C_{Si})$

El paciente X se clasifica como patológico.

17SR – Pregunta 1 y 19SR – Pregunta 2

PREGUNTA 1. Valoración: 3

Se quiere construir un **clasificador Naive-Bayes** para discriminar si un paciente tiene una determinada patología a partir del resultado de dos pruebas clínicas (PC_1 y PC_2) y de la existencia o no de un determinado síntoma (S1). Para aprender el clasificador, se utilizarán los datos de la tabla 1. Se pide:

(a) Calcular las diferentes tablas de probabilidad y distribuciones de normalidad necesarias para construir el clasificador.

(b) Clasifique un paciente, con valores [PC_1=120, PC_2=Normal, S_1=sí], como patológico o normal, de acuerdo al clasificador construido.

Paciente	PC_1	PC_2	S_1	**PATOLOGÍA**
1	100	Bajo	sí	**no**
2	125	Bajo	sí	**no**
3	145	Normal	no	**sí**
4	135	Bajo	no	**no**
5	100	Normal	no	**no**
6	140	Alto	no	**sí**
7	138	Alto	sí	**sí**
8	110	Normal	no	**no**
9	105	Bajo	no	**no**

Tabla 1

a)

Probabilidades a priori:

Valor de atributo clase (Patología)	$P(C_k)$, k = 1,2
C_1(Sí)	3/9 = 0,33
C_2(No)	6/9 = 0,66

Probabilidades condicionadas (atributos discretos):

PC_2	$P(PC_2 = \text{Alto} \mid C_k)$, k = 1,2	$P(PC_2 = \text{Normal} \mid C_k)$, k = 1,2	$P(PC_2 = \text{Bajo} \mid C_k)$, k = 1,2
C_1(Sí)	2/3 = 0,66	1/3 = 0,66	0/3 = 0 → 0,05*
C_2(No)	0/6 = 0 → 0,05*	2/6 = 0,33	4/6 = 0,66

*Corrección de la estimación de la probabilidad

S_1	$P(S_1 = \text{Sí} \mid C_k)$, k = 1,2	$P(S_1 = \text{No} \mid C_k)$, k = 1,2
C_1(Sí)	1/3 = 0,33	2/3 = 0,66
C_2(No)	2/6 = 0,33	4/6 = 0,66

Probabilidades condicionadas (atributos contínuos):

PC_1 / C_k = Sí

Media, $\mu = \frac{\sum x_i}{n} = \frac{145+140+138}{3} = 141$

Desviación típica, $\sigma = \sqrt{\frac{(\sum (x_i-\mu)^2)}{(n-1)}} = \sqrt{\frac{((145-141)^2+(140-141)^2+(138-141)^2)}{2}} = 3{,}60$

PC_1 / C_k = No

Media, $\mu = \frac{\sum x_i}{n} = \frac{100+125+135+100+110+105}{6} = 112{,}5$

Desviación típica, $\sigma = \sqrt{\frac{(\sum (x_i-\mu)^2)}{(n-1)}} = \sqrt{\frac{((100-4{,}5)^2+\ldots+(105-4{,}5)^2)}{5}} = 14{,}40$

PC_1 . Distribución de normalidad:

PC_1	μ	σ	$1/\sqrt{(2\pi\sigma^2)}$	$1/(2\sigma^2)$	$g(x,\mu,\sigma) = (1/\sqrt{(2\pi\sigma^2)}) \cdot \exp[-1/(2\sigma^2)\cdot(x-\mu)^2]$
C_1(Sí)	141	3.60	0.11081	0.03858	$0{,}11081 \cdot \exp[-0{,}03858 \cdot (x-141)^2]$
C_2(No)	112.5	14.40	0.02770	0.00241	$0{,}02770 \cdot \exp[-0{,}00241 \cdot (x-112{,}5)^2]$

b)

Clasificación de paciente X = [PC_1 = 120, PC_2 = Normal, S_1 = Sí]

Por teorema de Bayes, P(X) = max (P(X | C_{si}),P(X | C_{no})):

P(X | C_{Si}) = P(PC_1 = 120 | C_{si}) · P(PC_2 = Normal | C_{si}) · P(S_1 = Sí | C_{si}) · P(C_{si}) =
= $0{,}11081 \cdot \exp[-0{,}03858 \cdot (120-141)^2]$ · 0.66 · 0.33 · 0.33 =
= $3.252 \cdot 10^{-10}$

P(X | C_{No}) = P(PC_1 = 120 | C_{No}) · P(PC_2 = Normal | C_{No}) · P(S_1 = Sí | C_{no}) · P(C_{No}) =
= $0{,}02770 \cdot \exp[-0{,}00241 \cdot (120-112{,}5)^2]$ · 0.33 · 0.33 · 0.66 =
= $1.738 \cdot 10^{-10}$

P(X) = max (P(X | C_{si}),P(X | C_{no})) = max ($3.252 \cdot 10^{-10}$, $1.738 \cdot 10^{-10}$) = $3.252 \cdot 10^{-10}$ = P(C_{Si})

El paciente X se clasifica como patológico.

17F2

PREGUNTA 3. Valoración: 2.5

Se quiere construir un **clasificador Naïve-Bayes** para que el departamento de riesgo crediticio de un Banco permita automatizar el proceso de decisión relacionado con la concesión o no de préstamos. Para ello se dispone del conjunto de datos mostrado en la tabla 2. Se pide:

(a) Calcular las diferentes tablas de probabilidad y funciones de densidad de probabilidad necesarias para construir el clasificador.

(b) De acuerdo al clasificador construido, determine la concesión o no de un préstamo a un cliente del que se conoce la siguiente información: [F=No, IA=40, C=sí].

Muy importante: Deberá describir y justificar la forma de obtener cada una de las probabilidades.

Cliente	**Funcionario**	**Ingresos Anuales (miles de euros)**	**Casado**	**CONCEDER PRÉSTAMO**
1	Sí	81	No	**sí**
2	Sí	21	No	**no**
3	No	82	No	**no**
4	Sí	130	No	**sí**
5	Sí	141	Sí	**sí**
6	No	75	No	**no**
7	No	40	No	**no**
8	No	21	Sí	**no**
9	Sí	31	Sí	**no**

Tabla 2

a)

Probabilidades a priori

Valor de atributo clase (Conceder Préstamo)	$P(C_k)$, k = 1,2
C_1(Sí)	3/9 = 0,33
C_2(No)	6/9 = 0,66

Probabilidades condicionadas (atributos dicretos)

Funcionario	$P(F = \text{Sí} \mid C_k)$, k = 1,2	$P(F = \text{No} \mid C_k)$, k = 1,2
C_1(Sí)	3/3 = 1 → 0,95*	0/3 = 0 → 0,05*
C_2(No)	2/6 = 0,33	4/6 = 0,66

*Corrección de la estimación de la probabilidad

Casado	P(CAS = Sí | C_k), k = 1,2	P(CAS = No | C_k), k = 1,2
C_1(Sí)	1/3 = 0,33	2/3 = 0,66
C_2(No)	2/6 = 0,33	4/6 = 0,66

Probabilidades condicionadas (atributos contínuos)

IA / C_k = Sí

$$\text{Media, } \mu = \frac{\sum x_i}{n} = \frac{81+130+141}{3} = 117{,}33$$

$$\text{Desviación típica, } \sigma = \sqrt{\frac{\left(\sum (x_i-\mu)^2\right)}{(n-1)}} = \sqrt{\frac{\left((81-117{,}33)^2+\ldots+(141-117{,}33)^2\right)}{2}} = 31{,}94$$

IA / C_k = No

$$\text{Media, } \mu = \frac{\sum x_i}{n} = \frac{21+82+75+40+21+31}{6} = 45$$

$$\text{Desviación típica, } \sigma = \sqrt{\frac{\left(\sum (x_i-\mu)^2\right)}{(n-1)}} = \sqrt{\frac{\left((100-4{,}5)^2+\ldots+(105-4{,}5)^2\right)}{5}} = 26{,}98$$

IA . Distribución de normalidad:

IA	**μ**	**σ**	**$1/\sqrt{(2\pi\sigma^2)}$**	**$1/(2\sigma^2)$**	**$g(x,\mu,\sigma) = (1/\sqrt{(2\pi\sigma^2)}) \cdot \exp[-1/(2\sigma^2)\cdot(x-\mu)^2]$**
C_1(Sí)	117,33	31,94	0,01249	0,00049	$0{,}01249 \cdot \exp[-0{,}00049 \cdot (x-117{,}33)^2]$
C_2(No)	45	26,98	0,01478	0,00068	$0{,}00068 \cdot \exp[-0{,}01478 \cdot (x-45)^2]$

b)

Clasificación de paciente X = [F = No, IA = 40, CAS = Sí]

Por teorema de Bayes, $P(X) = \max(P(X \mid C_{si}), P(X \mid C_{no}))$:

$P(X \mid C_{Si}) = P(F = No \mid C_{si}) \cdot P(IA = 40 \mid C_{si}) \cdot P(CAS = Sí \mid C_{si}) \cdot P(C_{si}) =$
$= 0.05 \cdot 0{,}01249 \cdot \exp[-0{,}00049 \cdot (40-117{,}33)^2] \cdot 0.33 \cdot 0.33 =$
$= 3{,}630 \cdot 10^{-6}$

$P(X \mid C_{No}) = P(F = No \mid C_{No}) \cdot P(IA = 40 \mid C_{No}) \cdot P(CAS = Sí \mid C_{no}) \cdot P(C_{No}) =$
$= 0.66 \cdot 0{,}01478 \cdot \exp[-0{,}00068 \cdot (40-45)^2] \cdot 0.33 \cdot 0.66 =$
$= 2{,}088 \cdot 10^{-3}$

$P(X) = \max(P(X \mid C_{si}), P(X \mid C_{no})) = \max(3.630 \cdot 10^{-6}, 1.468 \cdot 10^{-3}) = 1.468 \cdot 10^{-3} = P(C_{No})$

El objeto X se clasifica como no apto para concederle un préstamo.

17SO

PREGUNTA 1. Valoración: 3

Se quiere construir un **clasificador Naïve-Bayes** que permita establecer la concesión o no de un préstamo a los clientes de un banco en función de una seria de indicadores I_i, con i=1,2,3, mostrados en la siguiente tabla. Se pide calcular las diferentes tablas de probabilidad y funciones de densidad de probabilidad necesarias para construir el clasificador bayesiano.

Cliente	I_1	I_2	I_3	Conceder préstamo
1	B	No	0.75	no
2	B	No	0.40	no
3	B	Sí	0.21	no
4	A	Sí	0.31	no
5	A	No	0.81	sí
6	A	No	0.21	no
7	B	No	0.82	no
8	A	No	1.30	sí
9	A	Sí	1.41	sí

Probabilidades a priori

Valor de atributo clase (Conceder Préstamo)	$P(C_k)$, k = 1,2
C_1(Sí)	3/9 = 0,33
C_2(No)	6/9 = 0,66

Probabilidades condicionadas (atributos dicretos)

I_1	$P(I_1 = A \mid C_k)$, k = 1,2	$P(I_1 = B \mid C_k)$, k = 1,2
C_1(Sí)	3/3 = 1 → 0,95*	0/3 = 0 → 0,05*
C_2(No)	2/6 = 0,33	4/6 = 0,66

*Corrección de la estimación de la probabilidad

I_2	$P(I_2 = Sí \mid C_k)$, k = 1,2	$P(I_2 = No \mid C_k)$, k = 1,2
C_1(Sí)	1/3 = 0,33	2/3 = 0,66
C_2(No)	2/6 = 0,33	4/6 = 0,66

Probabilidades condicionadas (atributos contínuos)

I_3 / C_k = Sí

Media, μ = $\frac{\sum x_i}{n} = \frac{0{,}81+1{,}30+1{,}41}{3}$ = 1,173

Desviación típica, σ = $\sqrt{\frac{(\sum (x_i-\mu)^2)}{(n-1)}} = \sqrt{\frac{((0{,}81-1{,}173)^2+...+(1{,}41-1173)^2)}{2}}$ = 0,3194

I_3 / C_k = No

Media, μ = $\frac{\sum x_i}{n} = \frac{0{,}75+0{,}40+0{,}21+0{,}31+0{,}21+0{,}82}{6}$ = 0,45

Desviación típica, σ = $\sqrt{\frac{(\sum (x_i-\mu)^2)}{(n-1)}} = \sqrt{\frac{((0{,}75-0{,}45)^2+...+(0{,}82-0{,}45)^2)}{5}}$ = 0,2698

I_3 . Distribución de normalidad:

I_3	μ	σ	1/√(2πσ²)	1/(2σ²)	g(x,μ,σ) = (1/√(2πσ²)) · exp[-1/(2σ²)·(x-μ)²]
C_1(Sí)	1,173	0,3194	1,249	4,9011	1,249 · exp[-4,9011 · (x-1,173)²]
C_2(No)	0,45	0,2698	1,4794	6,8688	1,4794 · exp[-6,8688 · (x-0,45)²]

18F1

PREGUNTA 4. Valoración: 3

Se quiere construir un **clasificador Naive-Bayes** para discriminar si un paciente tiene una determinada patología a partir del resultado de una prueba clínica (PC) y de la existencia o no de dos determinados síntoma (S1 y S2). Para aprender el clasificador, se utilizarán los datos de la tabla 3. Se pide:

(a) Calcular las diferentes tablas de probabilidad y distribuciones de normalidad necesarias para construir el clasificador.

(b) Clasifique un paciente, con valores [PC=120, S_1=No, S_2=Sí], como patológico o normal, de acuerdo al clasificador construido.

Paciente	PC	S_1	S_2	**PATOLOGÍA**
1	100	Sí	No	**No**
2	125	Sí	No	**No**
3	145	No	Sí	**Sí**
4	135	No	No	**No**
5	100	No	No	**No**
6	140	No	Sí	**Sí**
7	138	Sí	Sí	**Sí**
8	110	No	Sí	**No**
9	105	No	No	**No**

Tabla 3

a)

Probabilidades a priori

Valor de atributo clase (Patología)	$P(C_k)$, k = 1,2
C_1(Sí)	3/9 = 0,33
C_2(No)	6/9 = 0,66

Probabilidades condicionadas (atributos discretos)

S_1	$P(S_1$ = Sí$\mid C_k)$, k = 1,2	$P(S_1$ = No$\mid C_k)$, k = 1,2
C_1(Sí)	1/3 = 0,33	2/3 = 0,66
C_2(No)	2/6 = 0,33	4/6 = 0,66

S_2	$P(S_2$ = Sí$\mid C_k)$, k = 1,2	$P(S_2$ = No$\mid C_k)$, k = 1,2
C_1(Sí)	3/3 = 1 → 0,95*	0/3 = 0 → 0,05*
C_2(No)	1/6 = 0,16	5/6 = 0,83

*Corrección de la estimación de la probabilidad

Probabilidades condicionadas (atributos continuos)

PC / C_k = Sí

Media, $\mu = \dfrac{\sum x_i}{n} = \dfrac{145+140+138}{3} = 141$

Desviación típica, $\sigma = \sqrt{\dfrac{\left(\sum (x_i-\mu)^2\right)}{(n-1)}} = \sqrt{\dfrac{\left((145-141)^2+(140-141)^2+(138-141)^2\right)}{2}} = 3{,}60$

PC / C_k = No

Media, $\mu = \dfrac{\sum x_i}{n} = \dfrac{100+125+135+100+110+105}{6} = 112{,}5$

Desviación típica, $\sigma = \sqrt{\dfrac{\left(\sum (x_i-\mu)^2\right)}{(n-1)}} = \sqrt{\dfrac{\left((100-112{,}5)^2+\ldots+(105-112{,}5)^2\right)}{5}} = 14{,}40$

PC. Distribución de normalidad:

PC	μ	σ	$1/\sqrt{(2\pi\sigma^2)}$	$1/(2\sigma^2)$	$g(x,\mu,\sigma) = (1/\sqrt{(2\pi\sigma^2)}) \cdot \exp[-1/(2\sigma^2)\cdot(x-\mu)^2]$
C_1(Sí)	141	3,60	0,1108	0,0385	$0{,}1108 \cdot \exp[-0{,}0385 \cdot (x-141)^2]$
C_2(No)	112,5	14,40	0,0277	0,0024	$0{,}0277 \cdot \exp[-0{,}0024 \cdot (x-112{,}5)^2]$

b)

Clasificación de paciente X = [PC = 120, S_1 = No, S_2 = Sí]

P(X | $C_{Sí}$) = P(PC = 120 | $C_{Sí}$) · P(S_1 = No | $C_{Sí}$) · P(S_2 = Sí | $C_{Sí}$) · P($C_{Sí}$)
= $0{,}1108 \cdot \exp[-0{,}0385 \cdot (120-141)^2] \cdot 0{,}66 \cdot 0{,}95 \cdot 0{,}33$
= $9.697 \cdot 10^{-10}$

P(X | C_{No}) = P(PC = 120 | C_{No}) · P(S_1 = No | C_{No}) · P(S_2 = Sí | C_{No}) · P(C_{No})
= $0{,}0277 \cdot \exp[-0{,}0024 \cdot (120-112{,}5)^2] \cdot 0{,}66 \cdot 0{,}16 \cdot 0{,}66$
= $1.686 \cdot 10^{-3}$

P(X) = max (P(X | C_{si}),P(X | C_{no})) = max ($9.697 \cdot 10^{-10}$, $1.686 \cdot 10^{-3}$) = $1.686 \cdot 10^{-3}$ = P(C_{No})

El paciente X se clasifica como normal.

18SO

PREGUNTA 4. Valoración: 3

Se quiere construir un **clasificador Naive-Bayes** para discriminar si un paciente tiene una determinada patología a partir del resultado de una prueba clínica (PC1) y de la existencia o no de un determinado síntoma (S1). Para aprender el clasificador, se utilizarán los datos de la tabla 2. Se pide:

(a) Calcular las diferentes tablas de probabilidad y distribuciones de normalidad necesarias para construir el clasificador.

(b) Clasifique un paciente, con valores [PC1=16, S1=Sí], como patológico o normal, de acuerdo al clasificador construido.

Paciente	PC1	S1	PATOLOGÍA
1	10	Sí	**Sí**
2	12	Sí	**Sí**
3	14	No	**No**
4	13	Sí	**Sí**
5	10	Sí	**Sí**
6	14	No	**No**
7	13	No	**No**
8	11	No	**Sí**
9	10	Sí	**Sí**

Tabla 2

a)

Probabilidades a priori

Valor de atributo clase (Patología)	$P(C_k)$, k = 1,2
C_1(Sí)	6/9 = 0,66
C_2(No)	3/9 = 0,33

Probabilidades condicionadas (atributos discretos)

S_1	$P(S_1 = \text{Sí} \mid C_k)$, k = 1,2	$P(S_1 = \text{No} \mid C_k)$, k = 1,2
C_1(Sí)	5/6 = 0,83	1/6 = 0,16
C_2(No)	$0/3 = 0 \rightarrow 0{,}05^*$	$3/3 = 1 \rightarrow 0{,}95^*$

*Corrección de la estimación de la probabilidad

Probabilidades condicionadas (atributos continuos)

PC1 / C_k = Sí

Media, $\mu = \frac{\sum x_i}{n} = \frac{10+12+13+10+11+10}{6} = 11$

Desviación típica, $\sigma = \sqrt{\frac{\left(\sum (x_i-\mu)^2\right)}{(n-1)}} = \sqrt{\frac{\left((10-11)^2+(12-11)^2+...+(10-11)^2\right)}{2}} = 1{,}26$

PC1 / C_k = No

Media, $\mu = \frac{\sum x_i}{n} = \frac{14+14+13}{3} = 13{,}66$

Desviación típica, $\sigma = \sqrt{\frac{\left(\sum (x_i-\mu)^2\right)}{(n-1)}} = \sqrt{\frac{\left((14-13{,}66)^2+...+(13-13{,}66)^2\right)}{2}} = 0{,}57$

PC1. Distribución de normalidad:

PC1	**μ**	**σ**	**$1/\sqrt{(2\pi\sigma^2)}$**	**$1/(2\sigma^2)$**	**$g(x,\mu,\sigma) = (1/\sqrt{(2\pi\sigma^2)}) \cdot \exp[-1/(2\sigma^2)\cdot(x-\mu)^2]$**
C_1(Sí)	11	1,26	0,31662	0,31494	$0{,}31662 \cdot \exp[-0{,}31494 \cdot (x-11)^2]$
C_2(No)	13,66	0,57	0,69989	1,53893	$0{,}69989 \cdot \exp[-1{,}53893 \cdot (x-13{,}66)^2]$

b)

Clasificación de paciente X = [PC1 = 16, S_1 = Sí]

P(X | C_{Si}) = P(PC = 16 | C_{Si}) · P(S_1 = Sí| C_{Si}) · P(C_{Si})
= $0{,}31662 \cdot \exp[-0{,}31494 \cdot (16-11)^2] \cdot 0{,}83 \cdot 0{,}66$
= $6.603 \cdot 10^{-5}$

P(X | C_{No}) = P(PC = 16 | C_{No}) · P(S_1 = Sí | C_{No}) · P(C_{No})
= $0{,}69989 \cdot \exp[-1{,}53893 \cdot (16-13{,}66)^2] \cdot 0{,}05 \cdot 0{,}33$
= $2.528 \cdot 10^{-6}$

P(X) = max (P(X | C_{si}),P(X | C_{no})) = max ($6.603 \cdot 10^{-5}$, $2.528 \cdot 10^{-6}$) = $6.603 \cdot 10^{-5}$ = P(C_{Si})

El paciente X se clasifica como patológico.

19F1

PREGUNTA 3. Valoración: 3

Se quiere construir un **clasificador Naive-Bayes** para discriminar si un paciente tiene una determinada patología a partir del resultado de una prueba clínica (PC) y de la existencia o no de dos determinados síntoma (S_1 y S_2). Para aprender el clasificador, se utilizarán los datos de la tabla 2. Se pide:

(a) Calcular las diferentes tablas de probabilidad y distribuciones de normalidad necesarias para construir el clasificador.

(b) Clasifique un paciente, con valores [PC=150, S_1=No, S_2=No], como patológico o normal, de acuerdo al clasificador construido.

Paciente	**PC**	**S_1**	**S_2**	**PATOLOGÍA**
1	100	Sí	No	**Sí**
2	125	Sí	No	**Sí**
3	145	No	Sí	**No**
4	135	No	No	**Sí**
5	100	No	No	**Sí**
6	140	No	Sí	**No**
7	138	Sí	Sí	**No**
8	110	No	Sí	**Sí**
9	105	No	No	**Sí**

Tabla 2

a)

Probabilidades a priori

Valor de atributo clase (Patología)	$P(C_k)$, k = 1,2
C_1(Sí)	6/9 = 0,66
C_2(No)	3/9 = 0,33

Probabilidades condicionadas (atributos discretos)

S_1	$P(S_1 = Sí \mid C_k)$, k = 1,2	$P(S_1 = No \mid C_k)$, k = 1,2
C_1(Sí)	2/6 = 0,33	4/6 = 0,66
C_2(No)	1/3 = 0,33	2/3 = 0,66

S_2	$P(S_2 = Sí \mid C_k)$, k = 1,2	$P(S_2 = No \mid C_k)$, k = 1,2
C_1(Sí)	1/6 = 0,16	5/6 = 0,83
C_2(No)	3/3 = 1 → 0,95*	0/3 = 0 → 0,05*

*Corrección de la estimación de la probabilidad

Probabilidades condicionadas (atributos continuos)

PC / C_k = Sí

Media, $\mu = \dfrac{\sum x_i}{n} = \dfrac{100+125+135+100+110+105}{6} = 112{,}5$

Desviación típica, $\sigma = \sqrt{\dfrac{\left(\sum (x_i-\mu)^2\right)}{(n-1)}} = \sqrt{\dfrac{((100-112{,}5)^2+\ldots+(105-112{,}5)^2)}{5}} = 14{,}40$

PC / C_k = No

Media, $\mu = \dfrac{\sum x_i}{n} = \dfrac{145+140+138}{3} = 141$

Desviación típica, $\sigma = \sqrt{\dfrac{\left(\sum (x_i-\mu)^2\right)}{(n-1)}} = \sqrt{\dfrac{((145-141)^2+(140-141)^2+(138-141)^2)}{2}} = 3{,}60$

PC. Distribución de normalidad:

PC	**μ**	**σ**	**$1/\sqrt{(2\pi\sigma^2)}$**	**$1/(2\sigma^2)$**	**$g(x,\mu,\sigma) = (1/\sqrt{(2\pi\sigma^2)}) \cdot \exp[-1/(2\sigma^2)\cdot(x-\mu)^2]$**
C_1(Sí)	112,5	14,40	0,0277	0,0024	$0{,}0277 \cdot \exp[-0{,}0024 \cdot (x-112{,}5)^2]$
C_2(No)	141	3,60	0,1108	0,0385	$0{,}1108 \cdot \exp[-0{,}0385 \cdot (x-141)^2]$

b)

Clasificación de paciente X = [PC = 150, S_1 = No, S_2 = No]

$P(X \mid C_{Sí})$ $= P(PC = 150 \mid C_{Sí}) \cdot P(S_1 = No \mid C_{Sí}) \cdot P(S_2 = No \mid C_{Sí}) \cdot P(C_{Sí})$
$= 0{,}0277 \cdot \exp[-0{,}0024 \cdot (150-112{,}5)^2] \cdot 0{,}66 \cdot 0{,}83 \cdot 0{,}66$
$= 3.426 \cdot 10^{-4}$

$P(X \mid C_{No})$ $= P(PC = 150 \mid C_{No}) \cdot P(S_1 = No \mid C_{No}) \cdot P(S_2 = No \mid C_{No}) \cdot P(C_{No})$
$= 0{,}1108 \cdot \exp[-0{,}0385 \cdot (150-141)^2] \cdot 0{,}66 \cdot 0{,}05 \cdot 0{,}33$
$= 5.336 \cdot 10^{-5}$

$P(X) = \max(P(X \mid C_{si}), P(X \mid C_{no})) = \max(3.426 \cdot 10^{-4}, 5.336 \cdot 10^{-5}) = 3.426 \cdot 10^{-4} = P(C_{Sí})$

El paciente X se clasifica como patológico.

19F2

PREGUNTA 2. Valoración: 3

Se quiere construir un **clasificador Naive-Bayes** para discriminar si un paciente tiene una determinada patología a partir del resultado de una prueba clínica (PC) y de la existencia o no de dos determinados síntoma (S_1 y S_2). Para aprender el clasificador, se utilizarán los datos de la tabla 1. Se pide:

(a) Calcular las diferentes tablas de probabilidad y distribuciones de normalidad necesarias para construir el clasificador.

(b) Clasifique un paciente, con valores [PC=9, S_1=No, S_2=No], como patológico o normal, de acuerdo al clasificador construido.

Paciente	PC	S_1	S_2	PATOLOGÍA
1	10	No	No	**No**
2	14	No	Sí	**Sí**
3	13	Sí	Sí	**Sí**
4	11	No	Sí	**No**
5	10	No	No	**No**
6	10	Sí	No	**No**
7	12	Sí	No	**No**
8	14	No	Sí	**Sí**
9	13	No	No	**No**

Tabla 1

a)

Probabilidades a priori

Valor de atributo clase (Patología)	$P(C_k)$, k = 1,2
C_1(Sí)	3/9 = 0,33
C_2(No)	6/9 = 0,66

Probabilidades condicionadas (atributos discretos)

S_1	$P(S_1 = Sí \mid C_k)$, k = 1,2	$P(S_1 = No \mid C_k)$, k = 1,2
C_1(Sí)	1/3 = 0,33	2/3 = 0,66
C_2(No)	2/6 = 0,33	4/6 = 0,66

S_2	$P(S_2 = Sí \mid C_k)$, k = 1,2	$P(S_2 = No \mid C_k)$, k = 1,2
C_1(Sí)	3/3 = 1 → 0,95*	0/3 = 0 → 0,05*
C_2(No)	1/6 = 0,16	5/6 = 0,83

*Corrección de la estimación de la probabilidad

Probabilidades condicionadas (atributos continuos)

PC / C_k = Sí

Media, $\mu = \frac{\sum x_i}{n} = \frac{14+13+14}{3} = 13{,}66$

Desviación típica, $\sigma = \sqrt{\frac{\left(\sum (x_i-\mu)^2\right)}{(n-1)}} = \sqrt{\frac{((14-13{,}66)^2+\ldots+(14-13{,}66)^2)}{2}} = 0{,}57$

PC / C_k = No

Media, $\mu = \frac{\sum x_i}{n} = \frac{10+11+10+10+12+13}{6} = 11$

Desviación típica, $\sigma = \sqrt{\frac{\left(\sum (x_i-\mu)^2\right)}{(n-1)}} = \sqrt{\frac{((10-11)^2+(11-11)^2+\ldots+(13-11)^2)}{5}} = 1{,}26$

PC. Distribución de normalidad:

PC	μ	σ	$1/\sqrt{(2\pi\sigma^2)}$	$1/(2\sigma^2)$	$g(x,\mu,\sigma) = (1/\sqrt{(2\pi\sigma^2)}) \cdot \exp[-1/(2\sigma^2)\cdot(x-\mu)^2]$
C_1(Sí)	13,66	0,57	0,69989	1,53893	$0{,}69989 \cdot \exp[-1{,}53893 \cdot (x-13{,}66)^2]$
C_2(No)	11	1,26	0,31662	0,31494	$0{,}31662 \cdot \exp[-0{,}31494 \cdot (x-11)^2]$

b)

Clasificación de paciente X = [PC = 9, S_1 = No, S_2 = No]

$P(X \mid C_{Si})$ = $P(\,PC = 9 \mid C_{Si}) \cdot P(\,S_1 = No \mid C_{Si}) \cdot P(\,S_2 = No \mid C_{Si}) \cdot P(C_{Si})$
= $0{,}69989 \cdot \exp[-1{,}53893 \cdot (9-13{,}66)^2] \cdot 0{,}66 \cdot 0{,}05 \cdot 0{,}33$
= $2.335 \cdot 10^{-17}$

$P(X \mid C_{No})$ = $P(\,PC = 9 \mid C_{No}) \cdot P(\,S_1 = No \mid C_{No}) \cdot P(\,S_2 = No \mid C_{No}) \cdot P(C_{No})$
= $0{,}31662 \cdot \exp[-0{,}31494 \cdot (9-11)^2] \cdot 0{,}66 \cdot 0{,}83 \cdot 0{,}66$
= 0.0324

$P(X) = \max(P(X \mid C_{si}), P(X \mid C_{no})) = \max(2.335 \cdot 10^{-17}, 0.0324) = 0.0324 = P(C_{No})$

El paciente X se clasifica como normal.

19SO

PREGUNTA 4. Valoración: 3

Se quiere construir un **clasificador Naive-Bayes** para determinar a qué clase pertenece un objeto a partir del valor de un atributo numérico (N1) y dos atributos discretos (D1 y D2). Para aprender el clasificador, se utilizarán los datos de la tabla 2. Se pide:

(a) Calcular las diferentes tablas de probabilidad y distribuciones de normalidad necesarias para construir el clasificador.

(b) Clasifique un objeto de clasificación desconocida y valores de atributos [N1=75, D1=No, D2=No] a la clase correspondiente.

Objeto	N1	D1	D2	CLASE
1	70	No	Sí	**A**
2	69	No	No	**A**
3	55	No	Sí	**B**
4	52	Sí	Sí	**B**
5	50	Sí	No	**B**
6	62	Sí	No	**B**
7	72	No	Sí	**A**
8	64	Sí	Sí	**B**
9	50	Sí	No	**B**

Tabla 2

a)

Probabilidades a priori

Valor atributo clase	$P(C_k)$, k =1,2
C_1(A)	3/9
C_2(B)	6/9

Probabilidades condicionadas (atributos discretos)

D1	$P(D1 = Sí \mid C_k)$, k =1,2	$P(D1 = No \mid C_k)$, k =1,2
C_1(A)	$0/3 = 0 \rightarrow 0{,}05^*$	$3/3 = 1 \rightarrow 0{,}95^*$
C_2(B)	5/6 = 0,83	1/6 = 0,16

*Corrección de la estimación de la probabilidad

D2	$P(D2 = Sí \mid C_k)$, k =1,2	$P(D2 = No \mid C_k)$, k =1,2
C_1(A)	2/3 = 0,66	1/3 = 0,33
C_2(B)	3/6 = 0,5	3/6 = 0,5

Probabilidades condicionadas (atributos continuos)

N1 / C_k = A

$$\text{Media, } \mu = \frac{\sum x_i}{n} = \frac{70+69+72}{3} = 70{,}333$$

$$\text{Desviación típica, } \sigma = \sqrt{\frac{\left(\sum (x_i-\mu)^2\right)}{(n-1)}} = \sqrt{\frac{\left((70-70{,}333)^2+\ldots+(72-70{,}333)^2\right)}{2}} = 1{,}527$$

N1 / C_k = B

$$\text{Media, } \mu = \frac{\sum x_i}{n} = \frac{55+52+50+62+64+50}{6} = 55{,}5$$

$$\text{Desviación típica, } \sigma = \sqrt{\frac{\left(\sum (x_i-\mu)^2\right)}{(n-1)}} = \sqrt{\frac{\left((55-55{,}5)^2+\ldots+(50-55{,}5)^2\right)}{5}} = 6{,}123$$

PC. Distribución de normalidad:

N1	**μ**	**σ**	**$1/\sqrt{(2\pi\sigma^2)}$**	**$1/(2\sigma^2)$**	**$g(x,\mu,\sigma) = (1/\sqrt{(2\pi\sigma^2)}) \cdot \exp[-1/(2\sigma^2)\cdot(x-\mu)^2]$**
C_1(A)	70,333	1,527	0,26125	0,21443	$0{,}26125 \cdot \exp[-0{,}21443 \cdot (x-70{,}333)^2]$
C_2(B)	55,5	6,123	0,06515	0,01333	$0{,}06515 \cdot \exp[-0{,}01333 \cdot (x-55{,}5)^2]$

b)

Clasificación de objeto X = [N1 = 75, D1 = No, D2 = No]

$P(X \mid C_A) = P(N1 = 75 \mid C_A) \cdot P(D1 = No \mid C_A) \cdot P(D2 = No \mid C_A) \cdot P(C_A)$
$= 0{,}26125 \cdot \exp[-0{,}21443 \cdot (75-70{,}333)^2] \cdot 0{,}95 \cdot 0{,}33 \cdot 0{,}33$
$= 2.531 \cdot 10^{-4}$

$P(X \mid C_B) = P(N1 = 75 \mid C_B) \cdot P(D1 = No \mid C_B) \cdot P(D2 = No \mid C_B) \cdot P(C_B)$
$= 0{,}06515 \cdot \exp[-0{,}01333 \cdot (75-55{,}5)^2] \cdot 0{,}16 \cdot 0{,}5 \cdot 0{,}66$
$= 2.163 \cdot 10^{-5}$

$P(X) = \max(P(X \mid C_A), P(X \mid C_B)) = \max(2{,}531 \cdot 10^{-4}, 2{,}163 \cdot 10^{-5}) = 2{,}531 \cdot 10^{-4} = P(C_A)$

El objeto X se clasifica como A.

14SR

PREGUNTA 1. Valoración: 2.5

Describir el **algoritmo Q-learning** (aprendizaje por refuerzo) indicando:

a) La tarea que realiza en términos de búsqueda: conjunto de estados, conjunto de operadores, estado inicial y final, heurística.
b) Bias.
c) Cuáles son sus entradas y salidas. Así como aspectos relacionados con su representación, preprocesamiento y tratamiento.
d) Cómo le afecta el ruido: en entradas y en estructura.
e) Su complejidad espacial y temporal.
f) Control de la tarea aprendida (crítica/valoración/utilización).
g) Dependencia del conocimiento del dominio.

a)

La tarea en términos de búsqueda se puede resumir como:

- Conjunto de estados: cada uno formado por una matriz Q(s,a) con tantas filas como estados haya y tantas columnas como acciones, con valores generalmente reales en cada posición de la matriz.
- Conjunto de operadores: solo uno, '*modificar valor de la posición*'.
- Estado inicial: la matriz Q(s,a) se inicializa a 0 normalmente.
- Estado final: la condición de finalización puede ser ejecutar *n* ciclos de aprendizaje o comprobar que la matriz (estado) no varñia significativamente entre 2 ciclos de algoritmo. El objetivo: encontrar una política de acción óptima.
- Heurística: modificación del valor a través de la fórmula:

$$Q_t(s,a)=\alpha \cdot [R(s,a)+\gamma \cdot máx_{b\in A}(Q_{t-1}(s',b))]+(1-\alpha)\cdot Q_t - 1(s,a)$$

donde γ (0<γ <1) representa el descuento de la importancia que tiene el refuerzo en el tiempo y α permite que funcione en entornos deterministas (haciendo α=1 en la fórmula).

b)

El principal bias de esta técnica es de representación, dado que asume que el número de estados y acciones es finito y no muy grande. Si no se cumpliera esta restricción la tabla Q podría hacerse incluso infinita.

c)

- Representación de las entradas: instancias descritas en términos de 2 estados (valores discretos), una acción (valor discreto) y un refuerzo (valor contínuo).
- Tratamiento de las entradas: Incremental e iterativo.
- Preprocesamiento de las entradas: no.
- Representación de las salidas: matriz numérica.

d)

En entradas: la fórmula no determinista lo suaviza a través de la ponderción de α

En estructura: los cambios de un refuerzo en un elemento de la matriz tampoco afecta mucho al resultado final.

e)

Complejidad espacial: el espacio requerido es de $|S| \times |A|$ posiciones de la matriz, que puede suponer un problema cuando se trata en los que los estados o las acciones se representan con descripciones numéricas contínuas (haciendo las matrices de dimensiones infinitas).

Complejidad temporal: es proporcional al número de instancias de entrenamiento y al número de ciclos que se desee ejecutar.

f)

Crítica/ utilidad/ valoración: la crítica surge cuando aparece un nuevo ejemplo y hay que volver a recalcular la matriz de refuerzos.

Utilización de lo aprendido: para utilizar lo aprendido, teniendo ya la matriz Q y estando en el estado s, el sistema tendría que realizar aquella acción que maximice el refuerzo esperado en el tiempo, seleccionando la acción *a* de acuerdo a la política $\pi(s)$ tal que:

$$\pi(s) = \arg \max_a [Q(S,a)]$$

Es decir, elegiendo aquella acción con la que se espera obtener un máximo refuezo en el futuro.

g)

La dependencia del dominio está relacionada con la representación de los estados y delas acciones, así con cómo se defina el refuerzo.