ALGORITMO ID3

- Desarrollado por J. Ross Quinlan en 1983.
- ID3 significa *Induction Decision Trees*.
- Pertenece a la familia TDIDT (Top-Down Induction of Decision Trees).

Objetivo

- Construir un árbol de decisión que explique cada instancia de la secuencia de entrada de la manera más compacta posible a partir de una tabla de inducción.

- En cada momento elige el mejor atributo dependiendo de una determinada heurística.
- Determinar las variables que portan información relevante para la solución del problema
- Establecer la secuencia dentro del árbol de decisión.

Inconveniente

- Favorece indirectamente a aquellos atributos con muchos valores, los cuales no tienen que ser los más útiles.
- Genera árboles de decisión a partir de ejemplos de partida.

- Conflictos en la base de conocimientos donde diferentes soluciones se alcanzan con variables con los mismos valores asociados.
- Manejo discreto de los valores de las variables (rangos para discretizar una variable continua).
- Generación de grandes árboles de decisión que no representan garantía de reglas eficientes.
- Aplicables sólo a problemas de clasificación y diagnóstico.
- La generación de conclusiones intermedias se encuentra implícito en la generación de reglas a partir de la tabla de inducción (no se aprecia discriminación en el árbol de decisión).

Características

- Crear un árbol de decisión como un método para aproximar una función objetivo de valores **discretos**, que es resistente al ruido en los datos y que es capaz de hallar o aprender de una disyunción de expresiones.
- El resultado puede expresarse como un conjunto de reglas ***Si-entonces***.
- Intenta encontrar el árbol más sencillo que separa mejor los ejemplos.
- Es recursivo.
- No se realiza *"backtracking"*.
- Utiliza la entropía.

Estructura

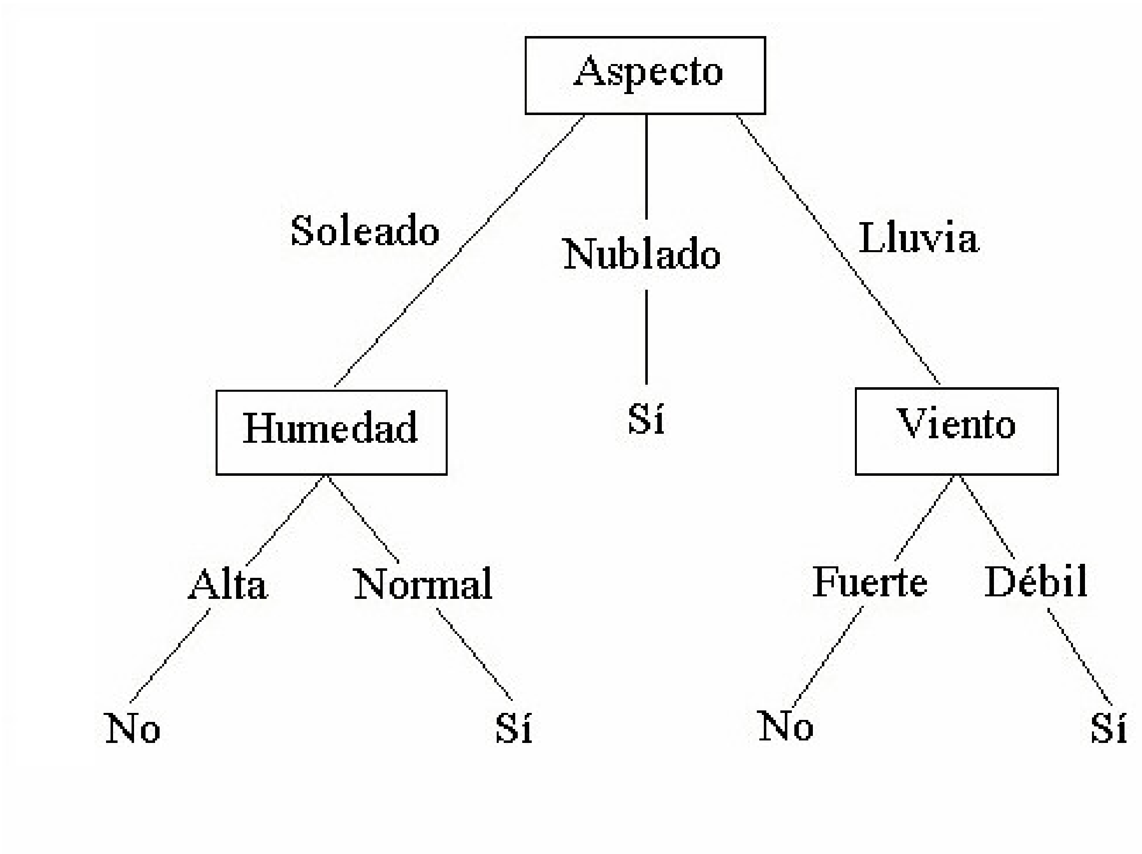
Los árboles de decisión están formados por:

- ***Nodos***: Nombres o identificadores de los atributos.
- ***Ramas***: Posibles valores del atributo asociado al nodo.
- ***Hojas***: Conjuntos ya clasificados de ejemplos y etiquetados con el nombre de una clase.

Un ejemplo de árbol de decisión es el siguiente:

Datos

- **Atributos:** Son los factores que influencian la clasificación o decisión.
- La selección de atributos debe basarse en el conocimiento acumulado por la experiencia.
- En este algoritmo cada atributo forma un nodo intermedio en un árbol cuyas hojas o nodos terminales son las clases o decisiones.
- **Clase:** Posibles valores de solución
- **Ejemplos:** Es el conjunto de combinaciones de atributos dados.
- Dado el conjunto de ejemplos, el ID3 selecciona el atributo que subdivide los ejemplos *de la mejor manera*.



- **Entropía:** Es la medida de la incertidumbre que hay en un sistema. Es decir, ante una determinada situación, la probabilidad de que ocurra cada uno de los posibles resultados.

♦ La función de entropía más usada es la binaria. Su expresión es con logaritmos base 2:

$$I(p,n) = -(p/(p+n))*LOG_2(p/(p+n))-(n/(p+n))*LOG_2(n/(p+n))$$

Recuerde:

Si $10^x=1{,}000$ entonces $x=3$

Es decir, si … $10^x=y$ entonces $x=\log_{10}(y)$

Si $2^x=y$ entonces $x=\log_2(y)$

Para calcular el logaritmo base 2:

Si $2^x=y$ entonces …

$$x=\log_{10}(y) \;/\; \log_{10}(2)$$

Aplicado a la fórmula de entropía…

$$\log_{10}(p/p+n) \ / \ \log_{10}(2)$$

Cálculo del total de la entropía de los atributos:

$$E(A)=((p1+n1)*I(p1,n1)+(p2+n2)*I(p2,n2)+\ldots+(pv+nv)*I(pv,nv))/(p+n)$$

Interpretación de la entropía:

- Un ejemplo de la entropía binaria podría ser sacar una bola de color blanco o negro de una bolsa.
- Si en la bolsa hay 3 bolas blancas y 3 negras el resultado es completamente

desconocido, es decir la incertidumbre es máxima, es decir la entropía es 1.

- Si en la bolsa hay 6 bolas negras el resultado es conocido de antemano, luego la incertidumbre no existe, y la entropía es 0.

♦ **Ganancia:** Es la diferencia entre la entropía de un nodo y la de uno de sus descendientes.

♦ En el fondo no es más que una heurística, que como veremos nos servirá para la elección del mejor atributo en cada nodo.

Ganancia(A) = I(p,n) - E(A)

- Un buen criterio parece ser escoger el atributo que *gana* la mayor información.
- ID3 examina todos los atributos y escoge el de máxima ganancia, forma la ramificación y usa el mismo proceso recursivamente para formar sub-árboles a partir de los *v* nodos generados

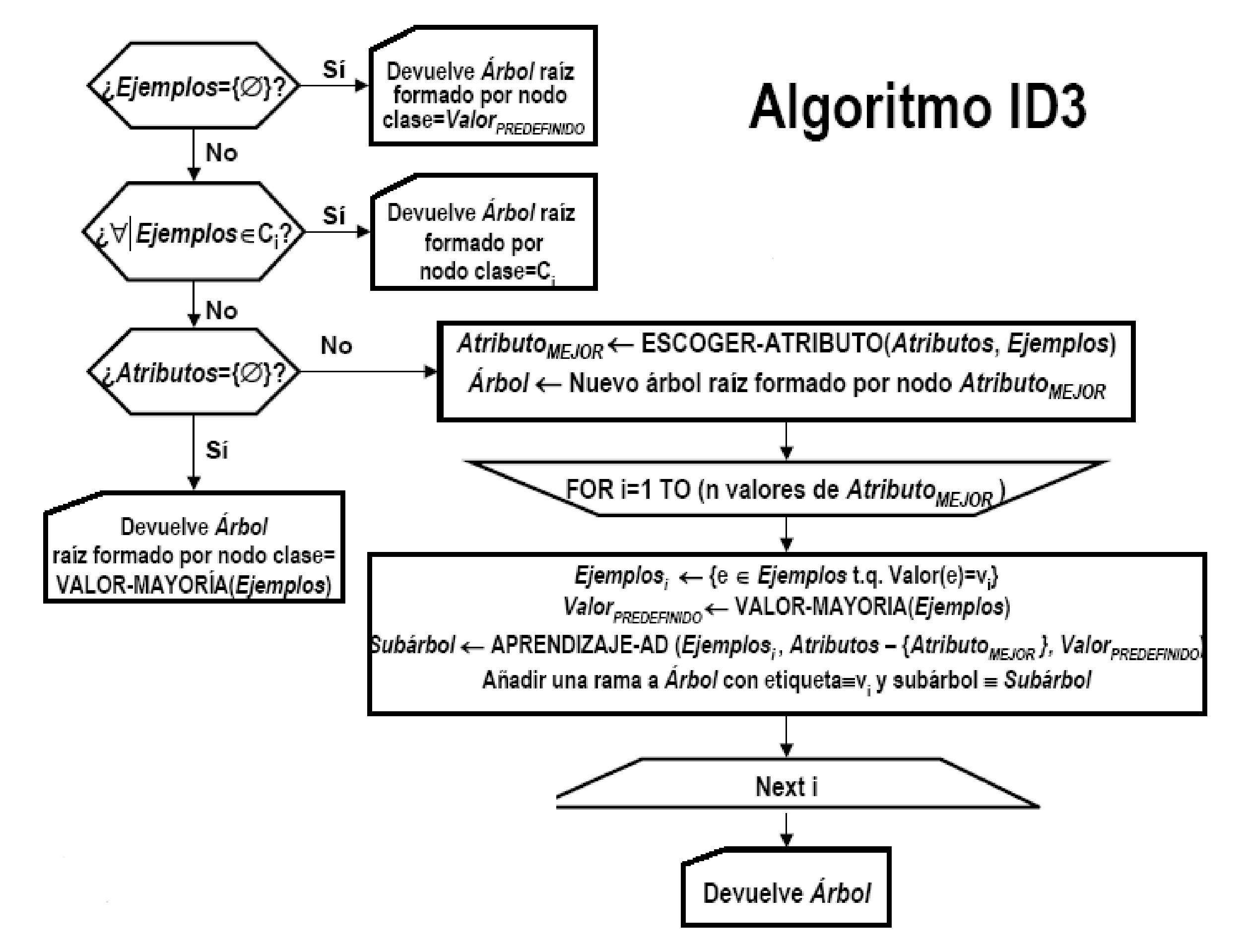
Procedimiento

- El árbol de decisión se recorre desde la raíz y tanto en ella como en cada uno de los demás nodos se decide cuál rama tomar basándonos en el valor de algún atributo del ejemplar que se esté clasificando, hasta llegar a un nodo terminal (hoja), que corresponde a la clase en que queda clasificado el ejemplar.
- Los árboles de decisión se adaptan especialmente bien a ciertos tipos de problemas. Básicamente, los casos para

los que son apropiados son aquellos en los que:

- Los ejemplos pueden ser descritos como pares valor-atributo.
- La función objetivo toma valores discretos.
- Podemos tomar hipótesis con disyunciones.
- Posible existencia de ruido en el conjunto de entrenamiento.
- Los valores de algunos atributos en los ejemplos del conjunto de entrenamiento pueden ser desconocidos.

Diagrama de flujo



Ejemplo

Una persona se dispone a jugar golf y tomará su decisión de asistir considerando los siguientes factores climatológicos:

Caso #	General	Temper.	Humedad	Viento	Clase
1	asoleado	caliente	alta	no	N
2	asoleado	caliente	alta	si	N
3	nublado	caliente	alta	no	P
4	lluvioso	templada	alta	no	P
5	lluvioso	fría	normal	no	P
6	lluvioso	fría	normal	si	N
7	nublado	fría	normal	si	P
8	asoleado	templada	alta	no	N
9	asoleado	fría	normal	no	P
10	lluvioso	templada	normal	no	P
11	asoleado	templada	normal	si	P
12	nublado	templada	alta	si	P
13	nublado	caliente	normal	no	P
14	lluvioso	templada	alta	si	N

- Nueve objetos son clase P y cinco son clase N
- La información requerida para la clasificación es:

$$I(p,n) = -(9/14)*\log_2(9/14) - (5/14)*\log_2(5/14) = 0.940 \text{ bits}$$

- Considerando el atributo ***General***, con sus tres valores (v=3):
  Para el primer valor, hay 5 objetos que lo tienen, 2 clase P y 3 clase N, entonces (Soleado):

p1 = 2 , n1 = 3 , I(p1,n1) = 0.971

- Análogamente, para el segundo valor posible de A (Nublado):

p2 = 4 , n2 = 0 , I(p2,n2) = 0

Y para el tercer valor de A (Lluvioso):

p3 = 3 , n3 = 2 , I(p3,n3) = 0.971

Por lo tanto el requisito de información esperada, después de revisar este atributo es:

E (General) =
(5*I(p1,n1) + 4*I(p2,n2) + 5*I(p3,n3))/14 = 0.694

Y la Ganancia de este atributo es:

Ganancia (General) =
0.940 - E(General) = 0.246

Y el mismo procedimiento aplicado a los otros tres atributos da:

Atributo Temperatura

Caliente

p1 = 2 , n1 =2 , I(p1,n1) = 1

$$I(p,n) = -(2/4)*\log_2(2/4) - (2/4)*\log_2(2/4) = 1$$

Templada

$p1 = 4$, $n1 = 2$, $I(p1,n1) = 0.918$

I(p,n) =
- (4/6)* $\log_2$(4/6) - (2/6)*$\log_2$ (2/)= .918

Fría

$p1 = 3$, $n1 = 1$, $I(p1,n1) = 0.811$

I(p,n) =
- (3/4)*$\log_2$ (3/4) - (1/4)* $\log_2$(1/4)= .811

E (Temperatura) =
(4*I(p1,n1) + 6*I(p2,n2) + 4*I(p3,n3))/14

E (Temperatura) =
(4*(1) + 6*(.918) + 4*(.811))/14

E (Temperatura) = 0.911

Ganancia (Temperatura) =
0.940 - E(Temperatura) = 0.029

Ganancia (Temperatura) =
0.940 -.911=.029

Atributo Humedad

Alta

p1 = 3 , n1 =4 , I(p1,n1) = .985

I(p,n) =
$$-(3/7)*\log_2(3/7) - (4/7)*\log_2(4/7)= .985$$

Normal

p1 = 6 , n1 = 1 , I(p1,n1) = 0.591

I(p,n) =
$$-(6/7)*\log_2(6/7) - (1/7)*\log_2(1/7)= .591$$

E (Humedad) =
(7*I(p1,n1) + 7*I(p2,n2))/14

E (Humedad) =
(7*(.985) + 7*(.591))/14

E (Humedad) = 0.788

Ganancia (Humedad) =
0.940 - E(Humedad) = 0.151

Ganancia (Humedad)= .940-
.788=.151

Atributo Viento

Si

p1 = 3 , n1 =3 , I(p1,n1) = 1

I(p,n) =
$-(3/6)* \log_2(3/6) - (3/6)*\log_2(3/6)=$
1

No

p1 = 6 , n1 = 2 , I(p1,n1) = 0.811

I(p,n) =
$-(6/8)* \log_2(6/8) - (2/8)*\log_2(2/8)=$
.811

E (Viento) =
(6*I(p1,n1) +8*I(p2,n2))/14

E (Viento) = (6*(1) + 8*(.811))/14

E (Viento) = 0.892

Ganancia (Viento)=
0.940 - E(Viento) = 0.151

Ganancia (Viento)= .940-.892=.048

Ganancia (Temperatura) = 0.029
Ganancia (Humedad) = 0.151
Ganancia (Viento) = 0.048

Otro ejemplo

Analizar los factores que tienen mayor impacto en diferentes *fondos de inversión*:

- *Tasas de interés*
- Cantidad de *efectivo* en Japón, Europa Occidental y los Estados Unidos
- El nivel de *tensión* internacional (p. ejem. posibles operaciones militares, incidentes de terrorismo, etc.

Basándose en la lista de atributos, se obtienen los siguientes datos históricos:

Tipo de fondo de inversión	Tasas de interés	Efectivo	Tensión	Valor del fondo de inversión (Clase)
Electrónica	Alta	Alta	Media	Media
Electrónica	Baja	Alta	Media	Alta
Electrónica	Media	Baja	Alta	Baja
Acciones de valores	Alta	Alta	Media	Alta
Acciones de valores	Baja	Alta	Media	Media
Acciones de valores	Media	Baja	Alta	Media
Bienes raíces	Alta	Alta	Media	Baja
Bienes raíces	Baja	Alta	Media	Alta
Bienes raíces	Media	Baja	Alta	Baja

- Note que la última columna representa la conclusión que buscará el algoritmo, donde se identifican tres posibilidades de inversión: *alta*, *media* y *baja*.
- El siguiente paso es desarrollar el árbol de decisión obtenido de dicha tabla (aplicando el cálculo de entropía de cada atributo)

- La fórmula de entropía para cualquier atributo A_k es:

$$H(C|A_k)=\sum_{j=1}^{M_k} p(a_{k,j})\left[-\sum_{i=1}^{N} p(c_i \mid a_{k,j}) \log_2 p(c_i|a_{k,j})\right]$$

donde:

$H(C\|A_k)$	Entropía de la propiedad de clasificación del atributo A_k
$p(a_{k,j})$	Probabilidad de que el atributo *k* esté en el valor *j*
$p(c_i \| a_{k,j})$	Probabilidad de la clase c_i cuando la variable *k* está en el valor *j*
M_k	Número total de valores para el atributo A_k; $j=1,2..\ M_k$
N	Número total de clases $i=1,2..N$
K	Número total de atributos; $k=1,2…K$

De la tabla se obtiene:

- Cuatro atributos (*Tipo de fondo de inversión*, *Tasas de interés*, *Efectivo*, *Tensión*), esto es K=4

- Tres clases (Valor del fondo de inversión: *alta*, *media* o *baja*); esto es $N=3$
- Tres valores para el atributo *Tipo de fondo de inversión* (Electrónica, Acciones de valores o Bienes raíces); esto es $M_1=3$
- Tres valores para el atributo *Tasas de interés* (Alta, Media o Baja); esto es $M_2=3$
- Dos valores para el atributo *Efectivo* (Alta o Baja); esto es $M_3=2$
- Dos valores para el atributo *Tensión* (Alta o Media); esto es $M_4=2$

El siguiente paso es calcular la entropía para cada atributo, p. ejem. para calcular la entropía del atributo *Efectivo*:

$p(a_{3,1})$ = probabilidad que el *Efectivo* sea Alto = 6/9 (el Efectivo es Alto en 6 de 9 veces)

$p(a_{3,2})$ = probabilidad que el *Efectivo* sea Bajo = 3/9

$p(c_1 \mid a_{3,1})$ = probabilidad de que un valor de la clase sea Alto cuando el *Efectivo* es Alto = 3/6

$p(c_2 \mid a_{3,1})$ = probabilidad de que un valor de la clase sea Media cuando el *Efectivo* es Alto = 2/6

$p(c_3 \mid a_{3,1})$ = probabilidad de que un valor de la clase sea Baja cuando el *Efectivo* es Alto = 1/6

$p(c_1 \mid a_{3,2})$ = probabilidad de que un valor de la clase sea Alto cuando el *Efectivo* es Baja = 0/3

$p(c_2 \mid a_{3,2})$ = probabilidad de que un valor de la clase sea Media cuando el *Efectivo* es Baja = 1/3

$p(c_3 \mid a_{3,2})$ = probabilidad de que un valor de la clase sea Baja cuando el *Efectivo* es Baja = 2/3

Se sustituyen los valores en la fórmula:

$$H(C|A_k)=\sum_{j=1}^{M_k} p(a_{k,j})\left[-\sum_{i=1}^{N} p(c_i \mid a_{k,j}) \log_2 p(c_i|a_{k,j})\right]$$

y se obtiene:

H(C|*Efectivo*)=6/9 [-3/6 $\log_2$ (3/6) – (2/6) $\log_2$(2/6) – (1/6) $\log_2$ (1/6)] + 3/9 [-0/3 $\log_2$ (0/3) – (1/3) $\log_2$(1/3) – (2/3) $\log_2$ (2/3)] = 1.2787

De manera similar se obtiene:

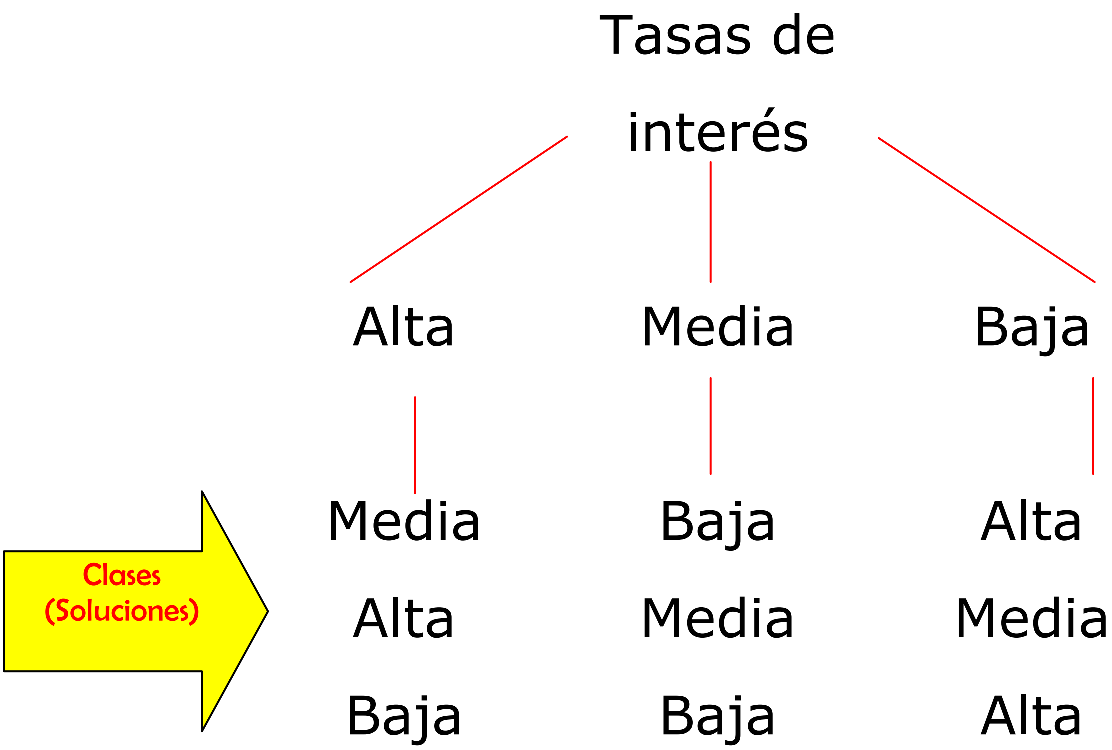
H(C|*Tasas de interés*) = 1.140333

H(C|*Tensión*) = 1.2787

H(C|*Tipo de fondo de inversión*) = 1.140333

Existe un empate en la entropía más baja entre la *Tasas de interés* y los *Tipos de fondo de inversión*.

- Si se rompe este empate en forma arbitraria y se construye un árbol de decisión teniendo como raíz la *Tasa de interés*, se presenta una clasificación inconclusa, ya que cada opción de dicho atributo conduce a múltiples soluciones de las clases.



- Una vez ramificado el árbol en el atributo *Tasa de interés*, se elimina dicha columna en la tabla original y se generan 3 sub-tablas (una cuando la *Tasa de interés* es Alta, otra cuando es Media y otra cuando es Baja):

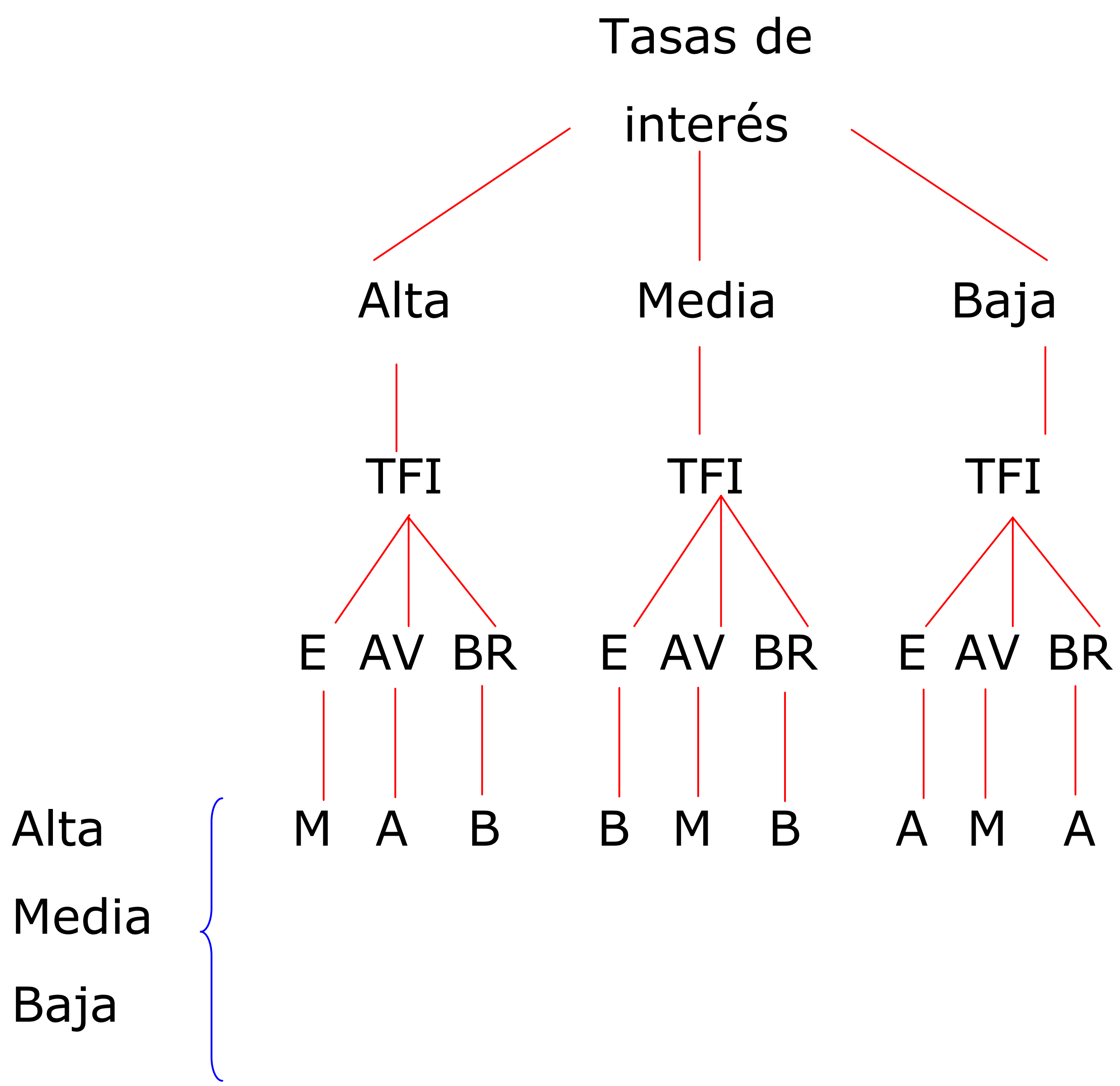
Clasificación cuando la Tasa de Interés es Alta:

Tipo de fondo de inversión	Efectivo	Tensión	Valor del fondo de inversión (Clase)
Electrónica	Alta	Media	Media
Acciones de valores	Alta	Media	Alta
Bienes raíces	Alta	Media	Baja

- Una vez más es necesario calcular la entropía para cada atributo (*Tipo de fondo de inversión*, *Efectivo* y *Tensión*)

- Como resultado se obtiene que el atributo *Tipo de fondo de inversión* tiene la menor entropía (igual a cero).

- Al ramificar el árbol considerando el *Tipo de fondo de inversión,* se obtiene:



De ahí se obtiene que cada rama conduce a una solución única.

Por último, se describen las reglas de producción que forman el Sistema Experto:

Regla 1: IF **tasa_interes** = **alta**
AND **tipo_fondo_inversion** = **electronica**
THEN ***valor***=**Media**

Regla 2: IF **tasa_interes** = **alta**
AND **tipo_fondo_inversion** = **acciones_valores**
THEN ***valor***=**Alta**

Regla 3: IF **tasa_interes** = **alta**
AND **tipo_fondo_inversion** = **bienes_raices**
THEN ***valor***=**Baja**

Regla 4: IF **tasa_interes** = **media**
AND **tipo_fondo_inversion** = **electronica**
THEN ***valor***=**Baja**

Regla 5: IF **tasa_interes** = **media**
AND **tipo_fondo_inversion** = **acciones_valores**
THEN ***valor***=**Media**

Regla 6: IF **tasa_interes** = **media**
AND **tipo_fondo_inversion** = **bienes_raices**
THEN ***valor***=**Baja**

Regla 7: IF **tasa_interes** = **baja**
AND **tipo_fondo_inversion** = **electronica**
THEN ***valor***=**Alta**

Regla 8: IF **tasa_interes** = **baja**

AND **tipo_fondo_inversion** = **acciones_valores**
THEN ***valor***=**Media**

Regla 9: IF **tasa_interes** = **baja**
AND **tipo_fondo_inversion** = **bienes_raices**
THEN ***valor***=**Alta**

También se pueden combinar ciertas reglas:

Regla 5/8: IF **tipo_fondo_inversion** = **acciones_valores**
AND **tasa_interes** = **media**
OR **tasa_interes** = **baja**
THEN ***valor***=**Media**

Regla 3/6: IF **tipo_fondo_inversion** = **bienes_raices**
AND **tasa_interes** = **alta**
OR **tasa_interes** = **media**
THEN ***valor***=**Baja**

Sin embargo, algunos paquetes de desarrollo de sistemas expertos no soportan premisas con cláusulas disyuntivas.

Otro ejemplo

Paciente	Presión Arterial	Urea en Sangre	Gota	Hipotiroidismo	Administrar Tratamiento
1	Alta	Alta	Si	No	No
2	Alta	Alta	Si	Si	No
3	Normal	Alta	Si	No	Si
4	Baja	Normal	Si	No	Si
5	Baja	Baja	No	No	Si
6	Baja	Baja	No	Si	No
7	Normal	Baja	No	Si	Si
8	Alta	Normal	Si	No	No
9	Alta	Baja	No	No	Si
10	Baja	Normal	No	No	Si
11	Alta	Normal	No	Si	Si
12	Normal	Normal	Si	Si	Si
13	Normal	Alta	No	No	Si
14	Baja	Normal	Si	Si	No

$$I(p,n) = -(9/14)*\log_2(9/14) - (5/14)*\log_2(5/14) = 0.940 \text{ bits}$$

- Considerando el atributo ***Presión Arterial*** , con sus tres valores (v=3): Para el primer valor, hay 5 objetos que lo tienen, 2 clase Si y 3 clase No, entonces (Alta):

p1 = 2 , n1 = 3 , I(p1,n1) = 0.971

- Análogamente, para el segundo valor posible de A (Normal):

p2 = 4 , n2 = 0 , I(p2,n2) = 0

Y para el tercer valor de A (Baja):

p3 = 3 , n3 = 2 , I(p3,n3) = 0.971

Por lo tanto el requisito de información esperada, después de revisar este atributo es:

E (Presion Arterial) =

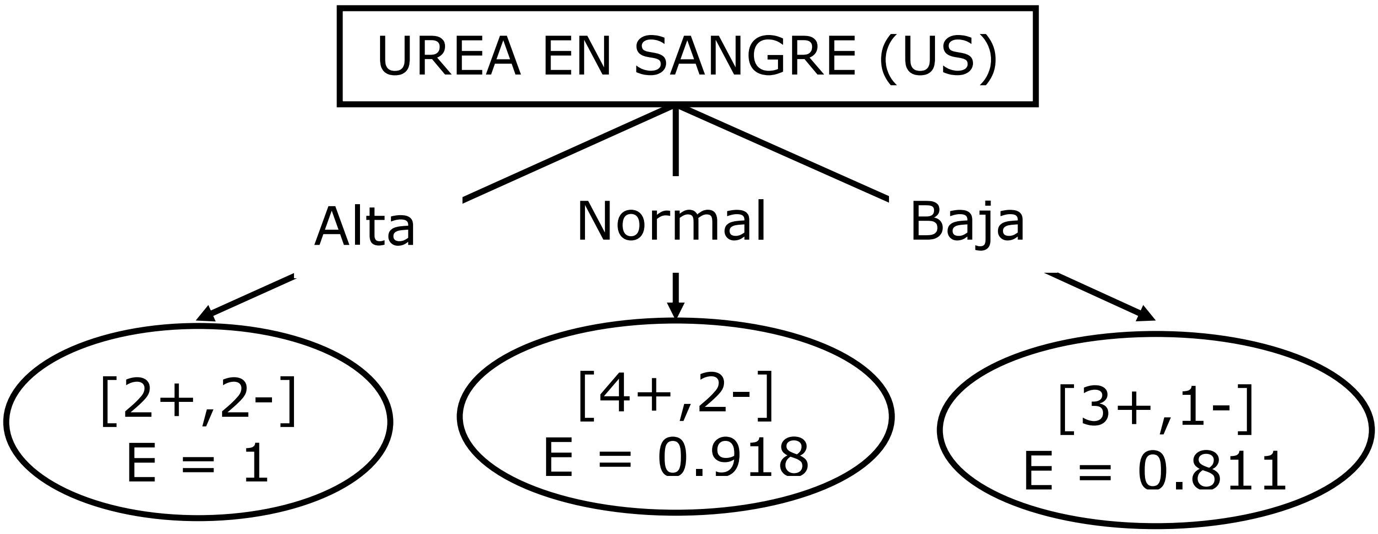
(5*I(p1,n1) + 4*I(p2,n2) + 5*I(p3,n3))/14 = 0.694

Y la Ganancia de este atributo es:

Ganancia (Presion Arterial) = 0.940 - E(Presion Arterial) = 0.246

Y el mismo procedimiento aplicado a los otros tres atributos da:

Atributo Urea en Sangre



Alta

p1 = 2 , n1 =2 , I(p1,n1) = 1

I(p,n) =

$$-(2/4)*\log_2(2/4) - (2/4)*\log_2(2/4)= 1$$

Normal

p1 = 4 , n1 = 2 , I(p1,n1) = 0.918

$$I(p,n) = - (4/6)* \log_2(4/6) - (2/6)*\log_2(2/6)= 0.918$$

Baja

p1 = 3, n1 = 1 , I(p1,n1) = 0.811

$$I(p,n) = - (3/4)*\log_2 (3/4) - (1/4)* \log_2(1/4)= 0.811$$

E (Urea en Sangre) = (4*I(p1,n1) + 6*I(p2,n2) + 4*I(p3,n3))/14

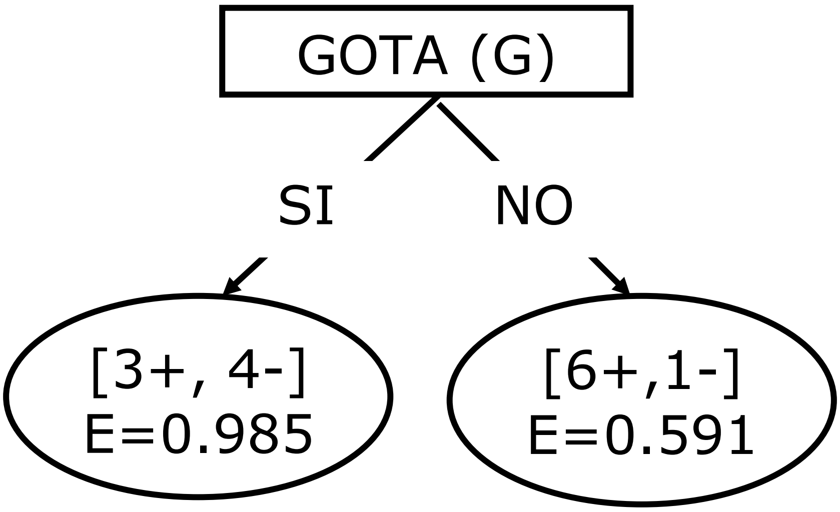
E (Urea en Sangre) = (4*(1) + 6*(0.918) + 4*(0.811))/14

E (Urea en Sangre) = 0.911

Ganancia (Urea en Sangre) = 0.940 - E(Urea en Sangre) = 0.571

Ganancia (Urea en Sangre) = 0.940 -0.911=0.029

Atributo Gota



Si

$p1 = 3$, $n1 = 4$, $I(p1,n1) = 0.985$

$$I(p,n) = -(3/7) * \log_2 (3/7) - (4/7) * \log_2 (4/7) = 0.985$$

No

$p1 = 6$, $n1 = 1$, $I(p1,n1) = 0.591$

I(p,n) =
-(6/7)* log2 (6/7) – (1/7)* log2 (1/7) = 0.591

E (Gota) =
(7*I(p1,n1) + 7*I(p2,n2))/14

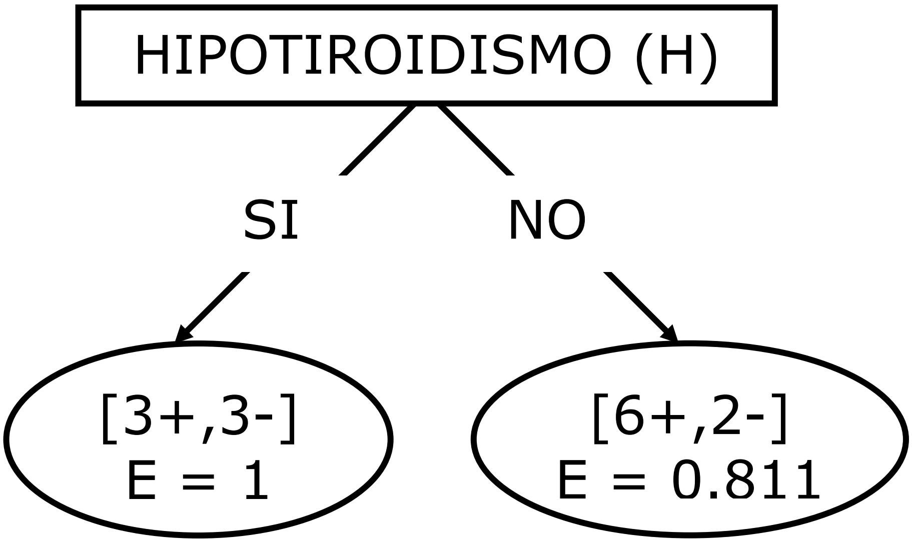
E (Gota) =
(7*(0.985) + 7*(0.591))/5

E (Gota) = 0.788

Ganancia (Gota) =
0.940 - E(Gota) = 0.152

Ganancia (Gota) =
0.940 – 0.788 = 0.152

Atributo Hipotiroidismo



Si

$p1 = 3$, $n1 = 3$, $I(p1,n1) = 1$

$$\mathbf{I(p,n) = -(3/6)*\log_2(3/6) - (3/6)*\log_2(3/6) = 1}$$

No

$p1 = 6$, $n1 = 2$, $I(p1,n1) = 0.811$

$$\mathbf{I(p,n) = -(6/8)*\log_2(6/8) - (2/8)*\log_2(2/8) = 0.811}$$

E (Hipotiroidismo) =
(6*I(p1,n1) + 8*I(p2,n2))/14

E (Hipotiroidismo) =
(3*(1) + 2*(0.811))/14

E (H) = 0.892

Ganancia (Hipotiroidismo) =
0.940 - E(Hipotiroidismo) = 0.048

Ganancia (Hipotiroidismo) =
0.940 -0.892=0.048

Presión Arterial

Normal

Pcte.	Urea	Gota	Hipot.	"TO"
3	Alta	Si	No	Sí
7	Baja	No	Sí	Sí
12	Normal	Sí	Sí	Sí
13	Alta	No	No	Sí

Alta

Pcte.	Urea	Gota	Hipot.	"TO"
1	Alta	Si	No	No
2	Alta	Sí	Sí	No
8	Normal	Sí	No	No
9	Baja	No	No	Sí
11	Normal	No	Sí	Sí

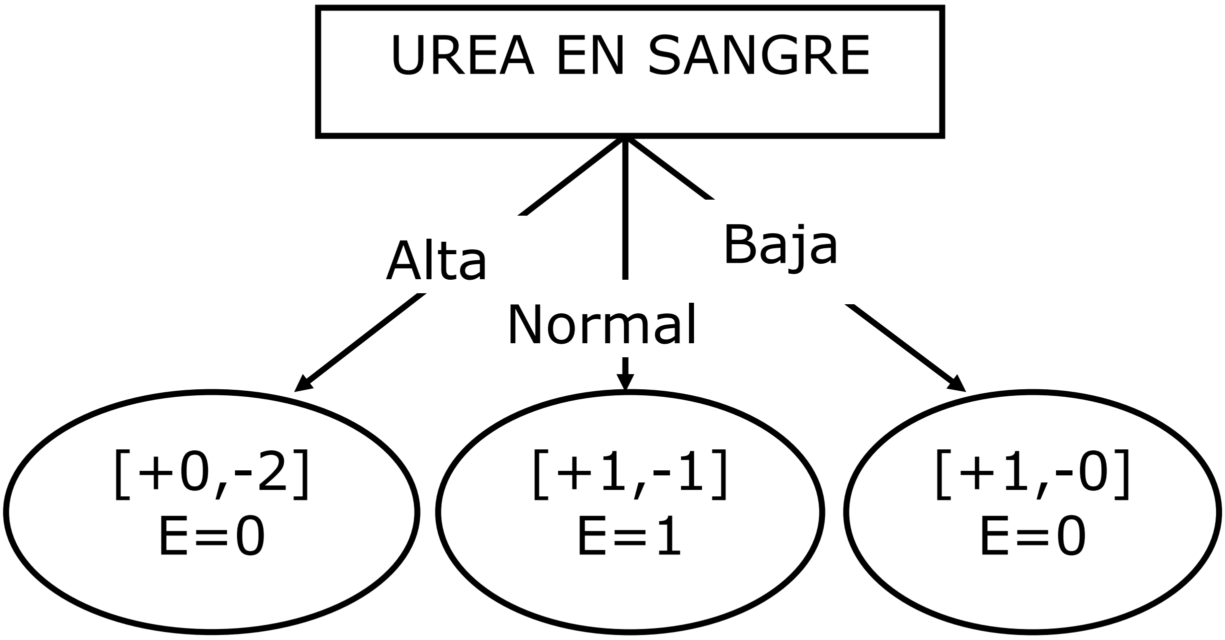
Baja

Pcte.	Urea	Gota	Hipot.	"TO"
4	Normal	Si	No	Sí
5	Baja	No	No	Sí
6	Baja	No	Sí	No
10	Normal	No	No	Sí
14	Normal	Sí	Sí	No

Presión Arterial Alta

$$I(p,n) = -(2/5)*\log_2(2/5) - (3/5)*\log_2(3/5) = 0.971 \text{ bits}$$

Atributo Urea en Sangre



Alta

$p1 = 0$, $n1 = 2$, $I(p1,n1) = 0$

$$\mathbf{I(p,n) = -(0/2)*\log_2(0/2) - (2/2)*\log_2(2/2) = 0}$$

Normal

$$p1 = 1 , n1 = 1 , I(p1,n1) = 1$$

$$I(p,n) = - (1/2)* \log_2(1/2) - (1/2)*\log_2 (1/2) = 1$$

Baja

$$p1 = 1, n1 = 0 , I(p1,n1) = 0$$

$$I(p,n) = - (1/1)*\log_2 (1/1) - (0/1)* \log_2(0/1) = 0$$

E (Urea en sangre) = (4*I(p1,n1) + 6*I(p2,n2) + 4*I(p3,n3))/5

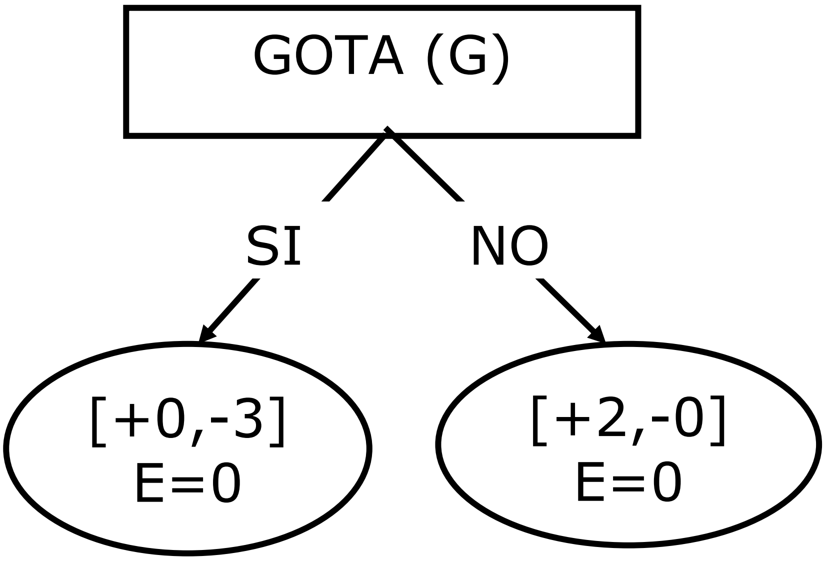
E (Urea en sangre) = (2*(0) + 2*(1) + 1*(0))/5

E (Urea en sangre) = 0.4

Ganancia (Urea en sangre) = 0.971 - E(Urea en sangre) = 0.571

Ganancia (Urea en Sangre) =
0.971 -0.4=0.571

Atributo Gota



Si

$p1 = 0$, $n1 = 3$, $I(p1,n1) = 0$

$$I(p,n) = -(0/3)*\log_2(0/3) - (3/3)*\log_2(3/3) = 0$$

No

$p1 = 2$, $n1 = 0$, $I(p1,n1) = 0$

$I(p,n) =$

$-(2/2)*\log_2(2/2) - (0/2)*\log_2(0/2) = 0$

E (Gota) =
(3*I(p1,n1) + 2*I(p2,n2))/5

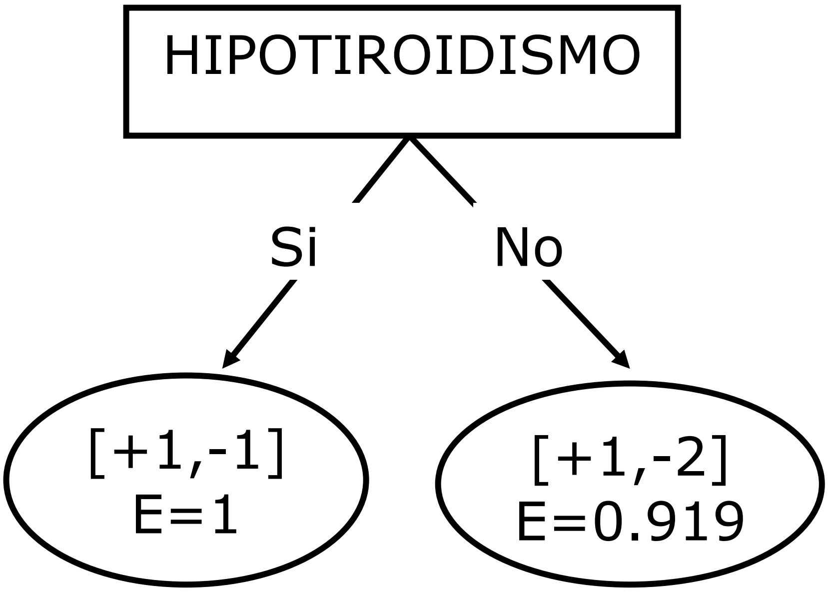
E (Gota) =
(3*(0) + 2*(0))/5

E (Gota) = 0

Ganancia (Gota) =
0.971 - E(Gota) = 0.971

Ganancia (Gota) =
0.971 -0=0.971

Atributo Hipotiroidismo



Si

p1 = 1 , n1 =1 , I(p1,n1) = 1

$$\mathbf{I(p,n) = -(1/2)*\log_2(1/2) - (1/2)*\log_2(1/2) = 1}$$

No

p1 = 1 , n1 = 2 , I(p1,n1) = 0.919

$$I(p,n) = -(1/3)*\log_2(1/3) - (2/3)*\log_2(2/3) = 0.919$$

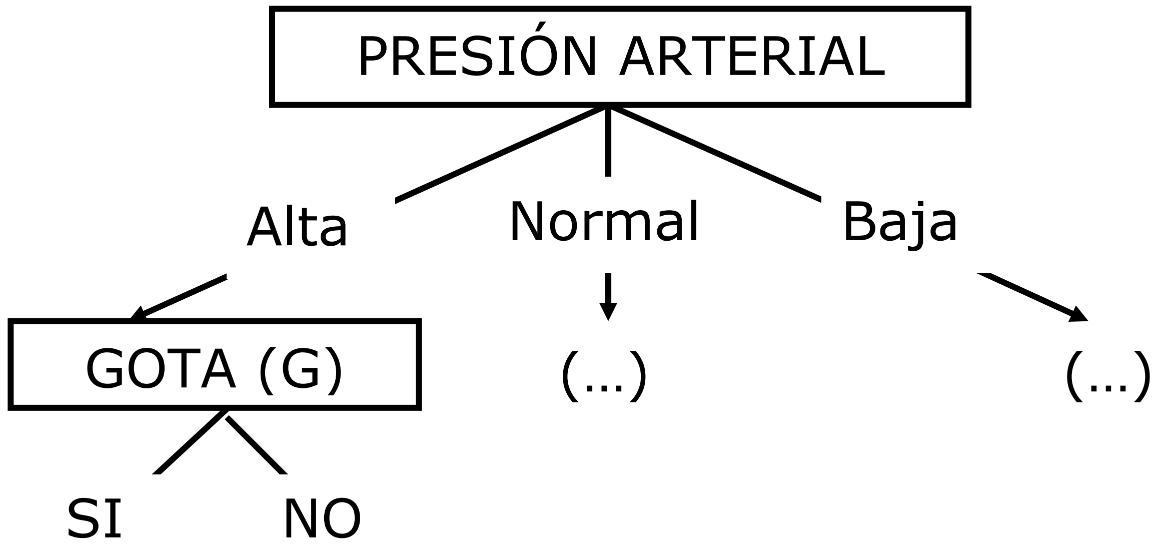
$$E\ (Hipotiroidismo) = (2*I(p1,n1) + 3*I(p2,n2))/5$$

$$E\ (Hipotiroidismo) = (2*(1) + 3*(0.919))/5$$

$$E\ (Hipotiroidismo) = 0.9514$$

$$Ganancia\ (Hipotiroidismo) = 0.971 - E(Hipotiroidismo) = 0.020$$

$$Ganancia\ (Hipotiroidismo) = 0.971 - 0.951 = 0.020$$



Pcte.	***Urea***	***Hipot.***	***"TO"***
1	Alta	No	No
2	Alta	Sí	No
8	Normal	No	No

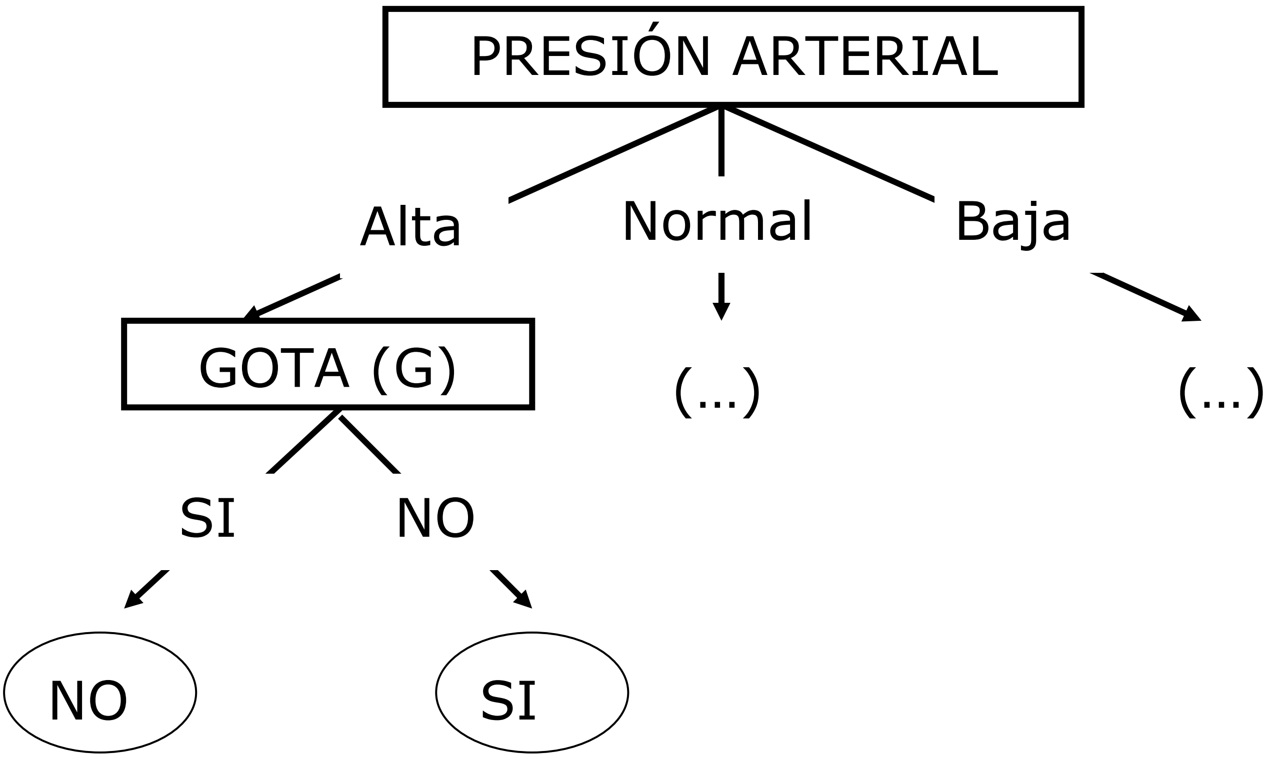
Pcte.	***Urea***	***Hipot.***	***"TO"***
9	Baja	No	Sí
11	Normal	Sí	Sí

Gota - Si

$$I(p,n) = -(0/3)*\log_2(0/3) - (3/3)*\log_2(3/3) = 0 \text{ bits}$$

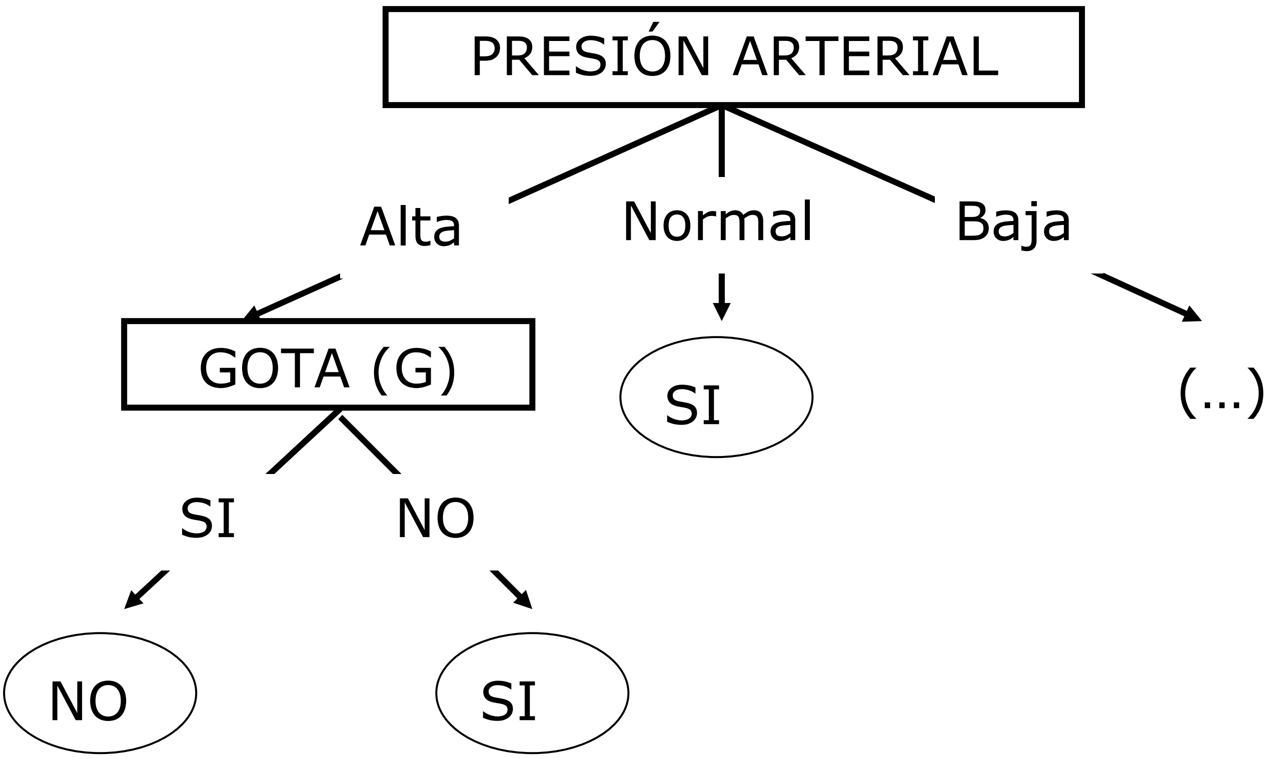
Gota - No

$$I(p,n) = -(2/2)*\log_2(2/2) - (0/2)*\log_2(0/2) = 0 \text{ bits}$$



Presión Arterial Normal

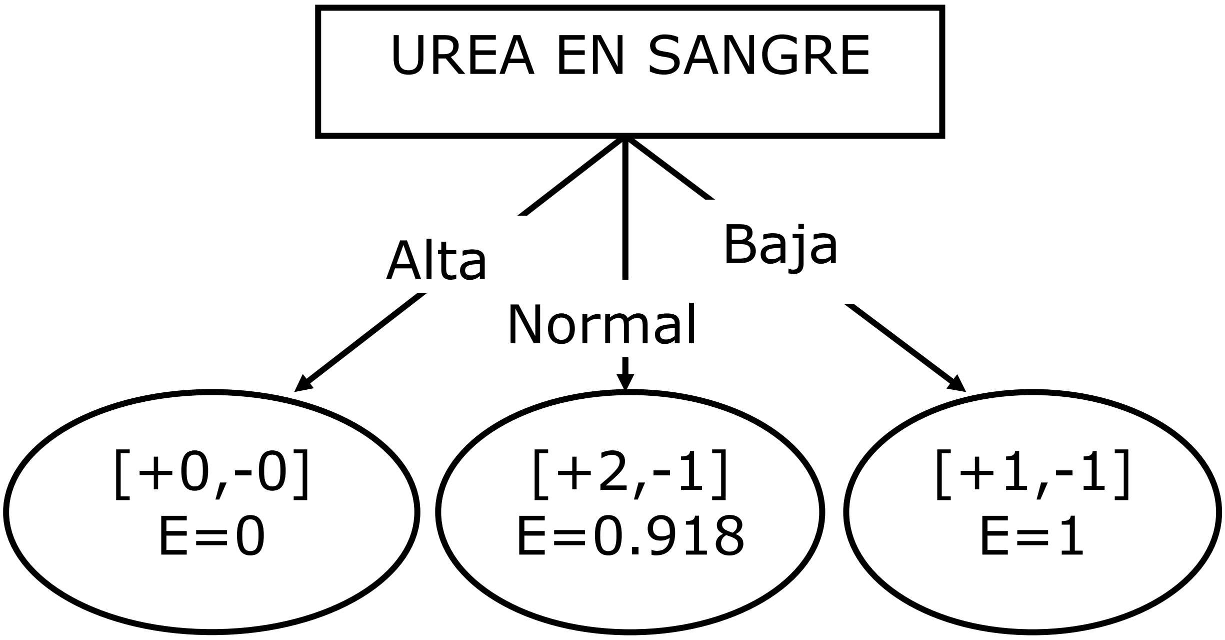
$$I(p,n) = -(4/4)*\log_2(4/4) - (0/4)*\log_2(0/4) = 0 \text{ bits}$$



Presión Arterial Baja

$$I(p,n) = -(3/5)*\log_2(3/5) - (2/5)*\log_2(2/5) = 0.971 \text{ bits}$$

Atributo Urea en Sangre



Alta

$p1 = 0$, $n1 = 0$, $I(p1,n1) = 0$

$$\mathbf{I(p,n) = -(0/0)*\log_2(0/0) - (0/0)*\log_2(0/0) = 0}$$

Normal

p1 = 2 , n1 = 1 , I(p1,n1) = 0.918

I(p,n) =
- (2/3)* $\log_2$(2/3) - (1/3)*$\log_2$ (1/3)= 0.918

Baja

p1 = 1, n1 = 1 , I(p1,n1) = 1

I(p,n) =
- (1/2)*$\log_2$ (1/2) - (1/2)* $\log_2$(1/2)= 1

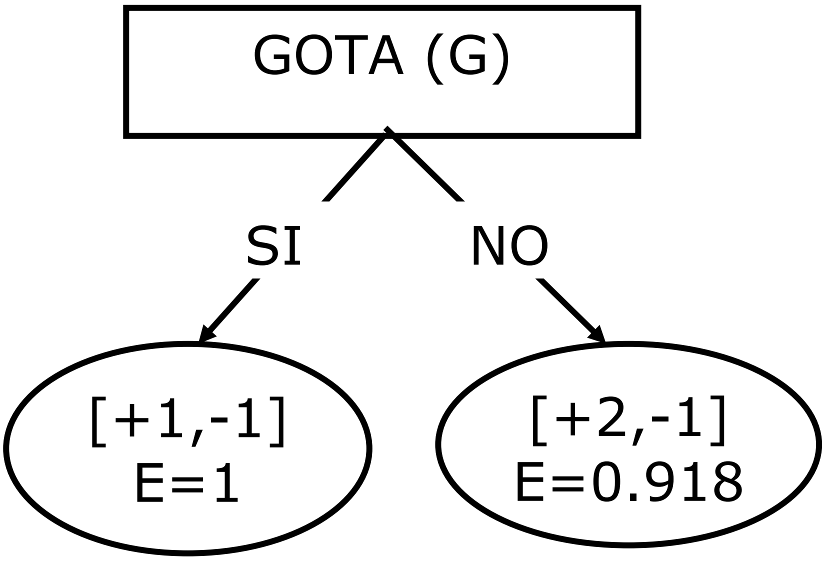
E (Urea en sangre) =
(0*I(p1,n1) + 3*I(p2,n2) + 2*I(p3,n3))/5

E (Urea en sangre) =
(0*(0) + 3*(0.918) + 2*(1))/5

E (Urea en sangre) = 0.951
Ganancia (Urea en sangre) = 0.971 - E(Urea en sangre) = 0.020

Ganancia (Urea en Sangre) = 0.971 -0.951=0.020

Atributo Gota



Si

$p1 = 1$, $n1 = 1$, $I(p1,n1) = 0$

$$I(p,n) = -(1/2)*\log_2(1/2) - (1/2)*\log_2(1/2) = 1$$

No

$p1 = 2$, $n1 = 1$, $I(p1,n1) = 0.918$

$$I(p,n) = -(2/3)*\log_2(2/3) - (1/3)*\log_2(1/3) = 0.918$$

E (Gota) =
(2*I(p1,n1) + 3*I(p2,n2))/5

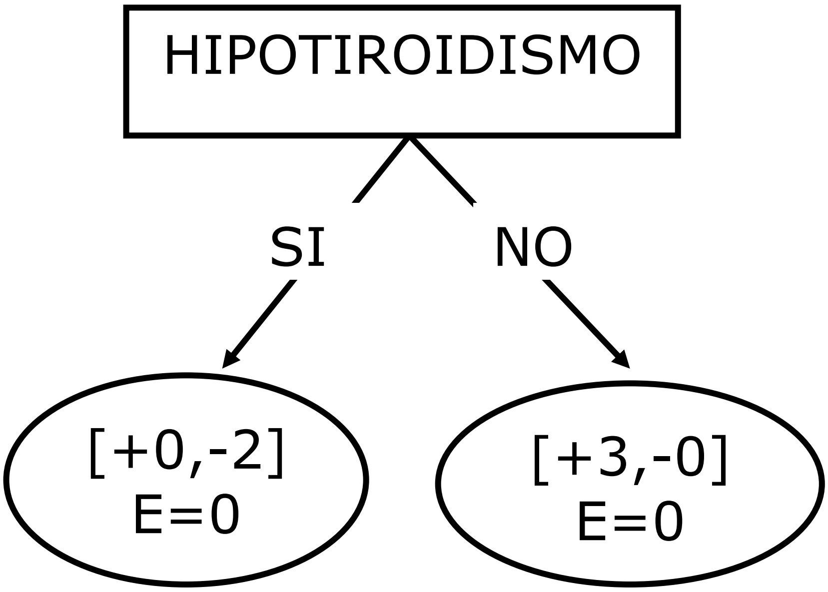
E (Gota) =
(2*(1) + 3*(0.918))/5

E (Gota) = 0.020

Ganancia (Gota) =
0.971 - E(Gota) = 0.020

Ganancia (Gota) =
0.971 -0.951=0.020

Atributo Hipotiroidismo



Si

$p1 = 0 \text{ , } n1 = 2 \text{ , } I(p1,n1) = 0$

$$\mathbf{I(p,n) =}$$
$$\mathbf{-(0/2)*\log_2(0/2) - (2/2)*\log_2 (2/2) = 0}$$

No

$p1 = 3 \text{ , } n1 = 0 \text{ , } I(p1,n1) = 0$

$$\mathbf{I(p,n) =}$$
$$\mathbf{-(3/3)*\log_2(3/3) - (0/3)*\log_2 (0/3) = 0}$$

E (Hipotiroidismo) =

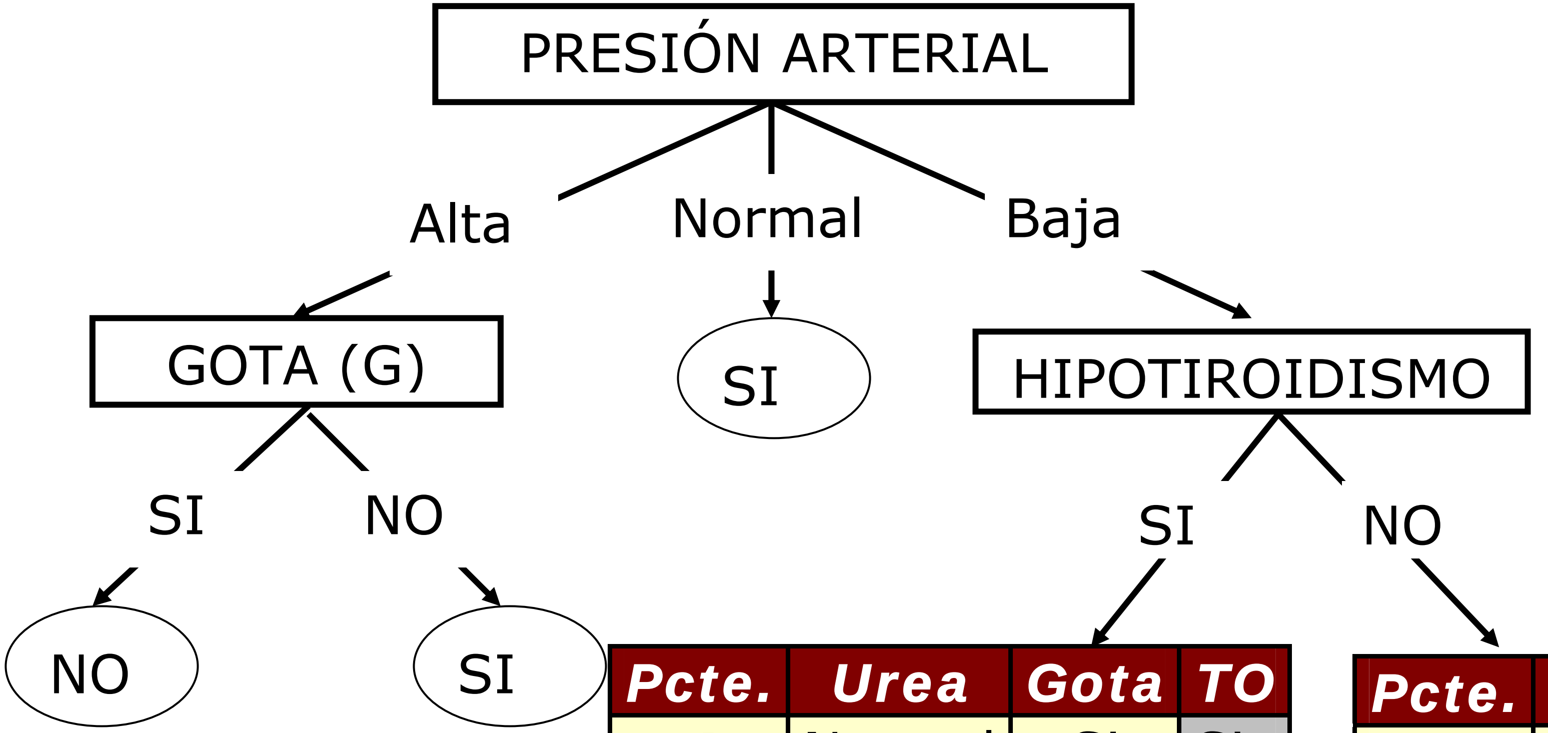
(3*I(p1,n1) + 2*I(p2,n2))/5

E (Hipotiroidismo) =
(3*(0) + 2*(0))/5

E (Hipotiroidismo) = 0

Ganancia (Hipotiroidismo) =
0.971 - E(Hipotiroidismo) = 0.971

Ganancia (Hipotiroidismo) =
0.971 -0=0.971



Pcte.	Urea	Gota	TO
4	Normal	Si	Si
5	Baja	No	Si
10	Normal	no	Si

Pcte.	Urea	Gota	TO
6	Baja	No	No
14	Normal	Si	No

Hipotiroidismo - SI

$$I(p,n) = -(3/3)*\log_2(3/3) - (0/3)*\log_2(0/3) = 0 \text{ bits}$$

Hipotiroidismo - NO

$$I(p,n) = -(0/2)*\log_2(0/2) - (2/2)*\log_2(2/2) = 0 \text{ bits}$$