



UNIVERSIDAD
DE SANTIAGO
DE CHILE



Ayudantía

Análisis de datos – PEP 2

Ayudante Gustavo Hurtado

Contenido incluido en la PEP 2

- Capítulo V – Reglas de Asociación.
- Capítulo VI – Clasificación Bayesiana.
- Capítulo VII – Árboles de Decisión.

Objetivos

- Comprender el entorno de problemas y la génesis de las reglas de asociación.
- Conocer las definiciones formales de las reglas de asociación.
- Comprender y manipular los parámetros necesarios para evaluar en forma cualitativa una regla.
- Comprender los problemas combinatorios subyacentes a la búsqueda de relaciones frecuentes.
- Evaluar las diferentes medidas de calidad de una regla de asociación.
- Dominar los algoritmos para la búsqueda de reglas frecuentes.

Problema:

Mejorar la calidad de decisiones comerciales a través del análisis de transacciones de compras que se han realizado en el pasado.

Definiciones básicas de Probabilidades :

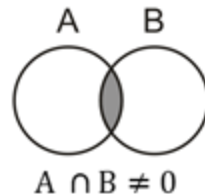
Sucesos mutuamente excluyentes, no pueden ocurrir al mismo tiempo son disjuntos, la ocurrencia de uno impide la del otro.



$$P(A \cup B) = P(A) + P(B)$$

Reglas de Adición

Sucesos no excluyentes, pueden ocurrir al mismo tiempo pero no dependen uno del otro para que se realice su ocurrencia.



$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Definiciones básicas de **Probabilidades** :

Sucesos independientes, son sucesos cuya ocurrencia no tiene efecto sobre la probabilidad de ocurrencia de otro.

Sucesos dependientes, son sucesos cuya ocurrencia tiene efecto directo sobre la probabilidad de ocurrencia de otro.

$$P(A \cap B) = P(A)P(B)$$

Reglas de Multiplicación

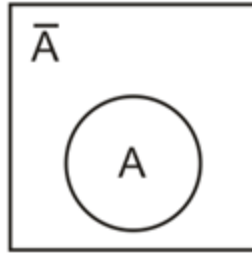
$$P(A \cap B) = P(A)P(B|A)$$

- Dónde $P(B|A)$ probabilidad condicional de B dado A , y al despejar la probabilidad condicional obtenemos:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Definiciones básicas de Probabilidades:

- **Sucesos complementarios**, la no aparición de ellos obliga la ocurrencia del otro.



A = Conjunto de A

\bar{A} = Complemento de A

Se deduce que $P(A) + P(\bar{A}) = 1$

Definición: Sistema de Información Operacional

Considere una base de datos o *Sistema de Información Operacional* compuesto por la 4-tupla:

$$SI = \langle U, Q, V, f \rangle$$

Dónde:

U es un universo cerrado (conjunto finito) no vacío de n objetos $[x_1, x_2, \dots, x_n]$.

Q es un conjunto finito, no vacío, de p atributos $[q_1, q_2, \dots, q_p]$.

V representa el dominio o valores que pueden adquirir los atributos q .

$f: U \times Q \rightarrow V$ es una función de decisión o *función de información* (correspondencia funcional) $f(x, q) \in V_q$ para cualquier $q \in Q$ y $x \in U$; y un par “atributo, valor” o (q, v) para todo $q \in Q$, $v \in V_q$ es llamado **descriptor**.

Definición: **Transacción**

- Establezcamos a S como un conjunto de transacciones del sistema de información SI , y en donde cada transacción x_i agrupa un conjunto de atributos (o productos), tal que $x_i \subseteq V$.

1 ['q8', 'q1', 'q6', 'q7', 'q5']

2 ['q14', 'q10', 'q12', 'q13', 'q5']

3 ['q3', 'q12', 'q14', 'q7', 'q11', 'q5', 'q8', 'q4']

4 ['q5', 'q14', 'q2', 'q3', 'q11', 'q4', 'q7', 'q8']

5 ['q15', 'q14', 'q7', 'q13', 'q6']

6 ['q7', 'q3', 'q12', 'q4', 'q5', 'q13']

7 ['q12', 'q1', 'q11', 'q4', 'q3']

8 ['q1', 'q3', 'q15', 'q6', 'q7', 'q13', 'q8', 'q5', 'q11']

...

x_i ['q9', 'q13', 'q8', 'q5', 'q10', 'q11', 'q4', 'q12']

Definición: Transacción

- Por consiguiente se dice que cada transacción s_i contiene un conjunto A de productos q en V , si $A \subseteq SI$.

U	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	q11	q12	q13	q14	q15
1	V	F	F	F	V	V	V	V	F	F	F	F	F	F	F
2	F	F	F	F	V	F	F	F	F	V	F	V	V	V	F
3	F	F	V	V	V	F	V	V	F	F	V	V	F	V	F
4	F	V	V	V	V	F	V	V	F	F	V	F	F	V	F
5	F	F	F	F	F	V	V	F	F	F	F	F	V	V	V
...
996	F	F	F	V	V	F	F	F	F	V	F	F	F	F	F
997	V	V	F	F	F	V	F	V	F	F	V	F	F	F	F
998	F	F	F	F	V	F	V	F	F	V	V	F	F	F	F
999	F	V	V	F	F	V	F	F	V	F	F	V	F	F	F
1000	F	F	F	F	F	F	V	F	V	V	F	V	F	F	F

Definición: **Regla de Asociación**

- Una *Regla de Asociación* es una implicación de la forma:

$$A \Rightarrow B$$

- En donde A y B representan dos conjuntos de atributos que cumplen con las condiciones $A \subset V, B \subset V$ y $(A \cap B) \cap V = \emptyset$. A es conjunto de productos de la **condición de la regla** denominado *Antecedente* y B es el conjunto de productos de la **conclusión de la regla** denominado *Consecuente*.

Definición: Soporte

- El **soporte de un conjunto** A de transacciones $Sop(A)$, se define como el número de transacciones de los atributos de A que toman el valor verdadero.
- El **soporte de una regla** $Sop(A, B)$ es el número de transacciones en el conjunto S en donde $A=B=Verdadero$.

$$Sop(A, B) = A \cap B$$

Ejemplo:

$$A = q1, B = q6$$

$$Sop(q1, q6) = 176$$



Definición: **Soporte normalizado**

- De igual forma para el cálculo de los distintos indicadores, se utiliza la proporción de las transacciones conjuntas entre A y B divididas por el número total de transacciones n . Así el **soporte normalizado** corresponde a la estimación de la probabilidad de la intersección (o juntura) entre A y B se define como:

$$Sopn(A, B) = \frac{Sop(A, B)}{n} = P(A \cap B)$$

Definición: **Soporte**

- Esta una medida es **anti monótona**, al especializar una regla se mantiene o disminuye el soporte.

$$Sop(A, B) = A \cap B$$

Ejemplo:

$$A = q1; B = q6$$

$$Sop(q1, q6) = 176$$

$$A = [q1, q3]; B = q6$$

$$Sop([q1, q3], q6) = 79$$

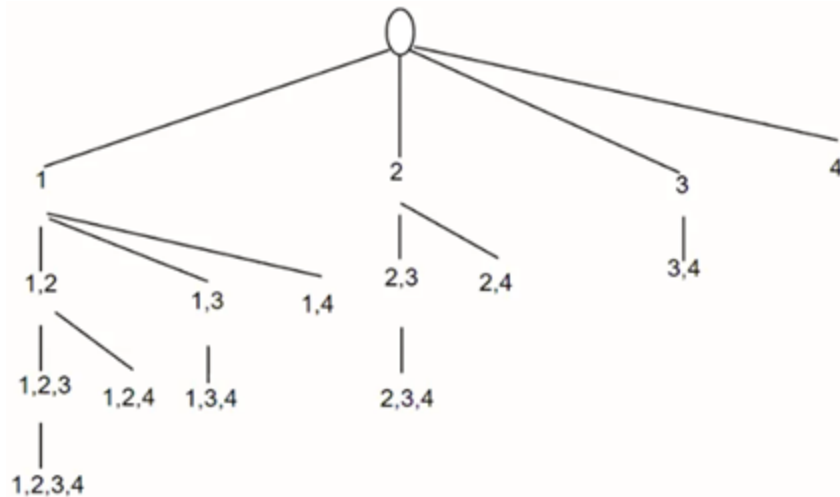
$$A = [q1, q3, q7]; B = q6$$

$$Sop([q1, q3, q7], q6) = 23$$

Capítulo V – Reglas de Asociación

Definición: Soporte mínimo (*minsop*)

- El algoritmo inicial de Agrawal utiliza esta combinatoria para generar el árbol paso a paso y va realizando una pre-poda por soporte.
- Cuando encuentra un nodo cuyo soporte es inferior al ***minsop*** entonces no genera las combinaciones.



Definición: **Confianza**

- Dada una regla de asociación $A \Rightarrow B$, la confianza de dicha regla en un conjunto S se define como la proporción de transacciones conjuntas entre A y B divididas por el número de casos de A .

$$Conf(A \Rightarrow B) = \frac{Sop(A, B)}{Sop(A)}$$

Definición: **Confianza normalizada**

- De igual forma que en el soporte, podemos normalizar el indicador al dividir por el número total de transacciones n . Consideremos entonces la estimación de la probabilidad de A corresponde a dividir al soporte de A por n , así por definición de probabilidad condicional obtenemos:

$$Conf_n(A \Rightarrow B) = \frac{\frac{Sop(A,B)}{n}}{\frac{Sop(A)}{n}} = \frac{P(A \cap B)}{P(A)} = P(B/A)$$

Medidas de Calidad: **Lift**

- Representa una medida de independencia dependencia entre A y B.
- Es una medida de eficiencia a través de correlación entre variables binarias.
- Es **monótona en confianza** al especializar una regla, el *Lift* disminuye proporcional a A , puesto que $P(B)$ se mantiene constante para el proceso de especialización.

$$Lift(A, B) = \frac{Conf(A \Rightarrow B)}{Sopn(B)}$$

$$Lift(A, B) = \frac{P(B|A)}{P(B)}$$

$$Lift(A, B) = \frac{P(A \cap B)}{P(A)P(B)}$$

Medidas de Calidad: Lift

- $Lift = 1$, el conjunto aparece una cantidad de veces acorde a lo esperado bajo condiciones de independencia.
- $Lift > 1$, el conjunto aparece una cantidad de veces superior a lo esperado bajo condiciones de independencia. La ocurrencia de los items de “B” influye en la probabilidad de la ocurrencia de los items de “A”.
- $Lift < 1$, el conjunto aparece una cantidad de veces inferior a lo esperado bajo condiciones de independencia. La ocurrencia de los items de “B” influye en la probabilidad de NO ocurrencia de los items de “A”.



Medidas de Calidad: Convicción

- Es una medida que evalúa el grado en que el antecedente influye en la ocurrencia del consecuente, representa la independencia de A y B, y es monótona en confianza. A diferencia del *Lift* es unidireccional.

$$\text{Conviccion}(A \Rightarrow B) = \frac{n - \text{Sop}(B)}{n[1 - \text{Conf}(A \Rightarrow B)]}$$

$$\text{Conviccion}(A \Rightarrow B) = \frac{P(A) \cdot P(\bar{B})}{P(A \cap \bar{B})}$$

- El valor tiene un rango entre $[1, +\infty]$.
- Cuanto más alto, mayor probabilidad de ocurrencia del consecuente cuando el antecedente ocurra. El valor 1 indica la independencia de los términos de la regla.



Medidas de Calidad: Laplace

- Es una medida monótona en confianza, para un soporte constante, y es monótona en soporte, para un el confianza constante.

$$Laplace(A \Rightarrow B) = \frac{Sop(A,B)+1}{Sop(A)+k}$$

$$Laplace(A \Rightarrow B) = \frac{Sop(A,B)+1}{\frac{Sop(A, B)}{c+k}}$$

Dónde: k es una la constante entera mayor que 1, (en clasificadores basados en reglas es el número de clases) y c es igual a la confianza.

Medidas de Calidad: **Ganancia**

$$Ganancia = Sop(A, B) - \theta Sop(A)$$

con $\theta < q < 1$

- Si la confianza se mantiene constante y $c > \theta$ se observa claramente que la ganancia disminuirá al disminuir el soporte de la regla, puesto que son proporcionales.

Pregunta 3

Una parte de la medida *Gini* para evaluar reglas de asociación está definida como:

$$Gini(A \rightarrow B) = P(A)[1 - P(B/A)]^2 + [1 - P(B/A)]^2$$

- Determine su monotonidad (**monótona o anti-monótona**) para el caso del soporte (con confianza constante) y para el caso de confianza (con soporte constante).



Pregunta 3 - Desarrollo

$$Gini(A \rightarrow B) = \overbrace{P(A)}^{Sopn(A)} \overbrace{[1 - P(B/A)]^2}^{Conf n(A \Rightarrow B)} + \overbrace{[1 - P(B/A)]^2}^{Conf n(A \Rightarrow B)}$$

$$Gini(A \rightarrow B) = P(A)[1 - 2P(B/A) + (B/A)^2] + [1 - 2P(B/A) + (B/A)^2]$$

Pregunta 3 - Desarrollo

$$Gini(A \rightarrow B) = \overbrace{P(A)}^{Sopn(A)} \overbrace{[1 - P(B/A)]^2}^{Conf n(A \Rightarrow B)} + \overbrace{[1 - P(B/A)]^2}^{Conf n(A \Rightarrow B)}$$

$$Gini(A \rightarrow B) = P(A)[1 - 2P(B/A) + (B/A)^2] + [1 - 2P(B/A) + (B/A)^2]$$

Evaluar: Caso del soporte (con confianza constante).

$$Sopn(A)=0,2$$

$$Conf n(A \Rightarrow B)=0,5$$

$$Gini(A \rightarrow B) = (0,2)[1 - 2(0,5) + (0,5)^2] + [1 - 2(0,5) + (0,5)^2] = 0,3$$

$$Sopn(A)=0,1$$

$$Conf n(A \Rightarrow B)=0,5$$

$$Gini(A \rightarrow B) = (0,1)[1 - 2(0,5) + (0,5)^2] + [1 - 2(0,5) + (0,5)^2] = 0,275$$

Pregunta 3 - Desarrollo

$$\begin{aligned}
 & \text{Sopn}(A) \quad \text{Conf}(A \Rightarrow B) \quad \text{Conf}(A \Rightarrow B) \\
 & \text{Gini}(A \rightarrow B) = P(A)[1 - P(B/A)]^2 + [1 - P(B/A)]^2 \\
 & \text{Gini}(A \rightarrow B) = P(A)[1 - 2P(B/A) + (B/A)^2] + [1 - 2P(B/A) + (B/A)^2]
 \end{aligned}$$

Evaluar: Caso del soporte (con confianza constante).

$$\text{Sopn}(A)=0,2$$

$$\text{Conf}(A \Rightarrow B)=0,5$$

$$\text{Gini}(A \rightarrow B) = (0,2)[1 - 2(0,5) + (0,5)^2] + [1 - 2(0,5) + (0,5)^2] = 0,3$$

$$\text{Sopn}(A)=0,1$$

$$\text{Conf}(A \Rightarrow B)=0,5$$

$$\text{Gini}(A \rightarrow B) = (0,1)[1 - 2(0,5) + (0,5)^2] + [1 - 2(0,5) + (0,5)^2] = 0,275$$

Al especializar el soporte es Anti-monótona

Pregunta 3 - Desarrollo

$$\begin{aligned}
 & \text{Sopn}(A) \quad \text{Confn}(A \Rightarrow B) \quad \text{Confn}(A \Rightarrow B) \\
 & \text{Gini}(A \rightarrow B) = \overbrace{P(A)}^{\text{Sopn}(A)} \left[\overbrace{1 - P(B/A)}^{\text{Confn}(A \Rightarrow B)} \right]^2 + \left[\overbrace{1 - P(B/A)}^{\text{Confn}(A \Rightarrow B)} \right]^2 \\
 & \text{Gini}(A \rightarrow B) = P(A)[1 - 2P(B/A) + (B/A)^2] + [1 - 2P(B/A) + (B/A)^2]
 \end{aligned}$$

Evaluar: Caso de confianza (con soporte del *Antecedente* constante).

Pregunta 3 - Desarrollo

$$\begin{aligned}
 & \text{Sopn}(A) \quad \text{Conf}(A \Rightarrow B) \quad \text{Conf}(A \Rightarrow B) \\
 & \text{Gini}(A \rightarrow B) = P(A)[1 - P(B/A)]^2 + [1 - P(B/A)]^2 \\
 & \text{Gini}(A \rightarrow B) = P(A)[1 - 2P(B/A) + (B/A)^2] + [1 - 2P(B/A) + (B/A)^2]
 \end{aligned}$$

Evaluar: Caso de confianza (con soporte del *Antecedente* constante).

$$\text{Sopn}(A)=0,1$$

$$\text{Conf}(A \Rightarrow B)=0,5$$

$$\text{Gini}(A \rightarrow B) = (0,1)[1 - 2(0,5) + (0,5)^2] + [1 - 2(0,5) + (0,5)^2] = 0,275$$

$$\text{Sopn}(A)=0,1$$

$$\text{Conf}(A \Rightarrow B)=0,6$$

$$\text{Gini}(A \rightarrow B) = (0,1)[1 - 2(0,6) + (0,6)^2] + [1 - 2(0,6) + (0,6)^2] = 0,176$$

Pregunta 3 - Desarrollo

$$\begin{aligned}
 & \text{Sopn}(A) \quad \text{Confn}(A \Rightarrow B) \quad \text{Confn}(A \Rightarrow B) \\
 & \text{Gini}(A \rightarrow B) = P(A)[1 - P(B/A)]^2 + [1 - P(B/A)]^2 \\
 & \text{Gini}(A \rightarrow B) = P(A)[1 - 2P(B/A) + (B/A)^2] + [1 - 2P(B/A) + (B/A)^2]
 \end{aligned}$$

Evaluar: Caso de confianza (con soporte del *Antecedente* constante).

$$\text{Sopn}(A)=0,1$$

$$\text{Confn}(A \Rightarrow B)=0,5$$

$$\text{Gini}(A \rightarrow B) = (0,1)[1 - 2(0,5) + (0,5)^2] + [1 - 2(0,5) + (0,5)^2] = 0,275$$

$$\text{Sopn}(A)=0,1$$

$$\text{Confn}(A \Rightarrow B)=0,6$$

$$\text{Gini}(A \rightarrow B) = (0,1)[1 - 2(0,6) + (0,6)^2] + [1 - 2(0,6) + (0,6)^2] = 0,176$$

Al especializar la confianza es

Anti-monótona

Pregunta 5

- Un analista de marketing tiene la hipótesis que existe una fuerte relación (más del 50%) entre personas que están suscritas al diario *El Mercurio* y *Estrategia*, con las personas que toman sus vacaciones en el extranjero y en el sur del país. Para comprobar esta hipótesis posee las bases de datos de las suscripciones a los periódicos y de la empresa *Latam*.
- Construya una base de datos de no más de 10 ejemplos que compruebe la hipótesis. Considerando un soporte de más del 10%, cuando el universo está constituido por las personas comunes a las tres bases de datos.

Pregunta 5 - Desarrollo

	Rut Cliente	El Mercurio (EM)	Estrategia (ES)	Latam (LT)
1	11200200-0	V	V	V
2	12300300-1	V	F	V
3	13400400-2	V	V	F
4	14500500-3	V	V	V
5	15600600-4	F	V	V
6	16700700-5	V	V	V
7	11800800-6	V	F	F
8	12900900-7	V	V	V
9	13010010-8	V	V	F
10	14020020-9	V	F	V

$$H_0: \text{Confn}(A, B) \leq 0,5$$

$$H_A: \text{Confn}(A, B) > 0,5$$

Pregunta 5 - Desarrollo

		<i>A</i>		<i>B</i>
	Rut Cliente	El Mercurio (EM)	Estrategia (ES)	Latam (LT)
1	11200200-0	V	V	V
2	12300300-1	V	F	V
3	13400400-2	V	V	F
4	14500500-3	V	V	V
5	15600600-4	F	V	V
6	16700700-5	V	V	V
7	11800800-6	V	F	F
8	12900900-7	V	V	V
9	13010010-8	V	V	F
10	14020020-9	V	F	V

$$Sop(A, B) = A \cap B$$

$$Sopn(A, B) = \frac{Sop(A, B)}{n} = P(A \cap B)$$

$$A = [EM, ES]; B = LT$$

$$n = 10$$

$$Sop([EM, ES], LT) = 4$$

$$Sopn([EM, ES], LT) = 4/10 = 0,4$$

Pregunta 5 - Desarrollo

		<i>A</i>		<i>B</i>
	Rut Cliente	El Mercurio (EM)	Estrategia (ES)	Latam (LT)
1	11200200-0	V	V	V
2	12300300-1	V	F	V
3	13400400-2	V	V	F
4	14500500-3	V	V	V
5	15600600-4	F	V	V
6	16700700-5	V	V	V
7	11800800-6	V	F	F
8	12900900-7	V	V	V
9	13010010-8	V	V	F
10	14020020-9	V	F	V

$$Conf(A \Rightarrow B) = \frac{Sop(A, B)}{Sop(A)}$$

$$Conf_n(A \Rightarrow B) = \frac{\frac{Sop(A, B)}{n}}{\frac{Sop(A)}{n}} = \frac{P(A \cap B)}{P(A)} = P(B/A)$$

$$Conf_n(A \Rightarrow B) = \frac{0,4}{0,6} = 0,66$$

Pregunta 5 - Desarrollo

		<i>A</i>		<i>B</i>
	Rut Cliente	El Mercurio (EM)	Estrategia (ES)	Latam (LT)
1	11200200-0	V	V	V
2	12300300-1	V	F	V
3	13400400-2	V	V	F
4	14500500-3	V	V	V
5	15600600-4	F	V	V
6	16700700-5	V	V	V
7	11800800-6	V	F	F
8	12900900-7	V	V	V
9	13010010-8	V	V	F
10	14020020-9	V	F	V

$$Conf(A \Rightarrow B) = \frac{Sop(A, B)}{Sop(A)}$$

$$Conf_n(A \Rightarrow B) = \frac{\frac{Sop(A, B)}{n}}{\frac{Sop(A)}{n}} = \frac{P(A \cap B)}{P(A)} = P(B/A)$$

$$Conf_n(A \Rightarrow B) = \frac{0,4}{0,6} = 0,66$$

$$H_0: Conf_n(A, B) \leq 0,5$$

$$H_A: Conf_n(A, B) > 0,5$$

Objetivos

- Cuantificar **probabilidad a priori**.
- Comprender el **costo del error** de clasificar basado en **probabilidad a priori**.
- Cuantificar el **riesgo condicional**.
- Usar los conceptos anteriores para obtener un método de clasificación a mínimo riesgo condicional para un **problema multivariado**.
- Analizar clasificador **Bayesiano simple**.
- Comprender los fundamentos de un clasificador **Bayesiano ingenuo**.

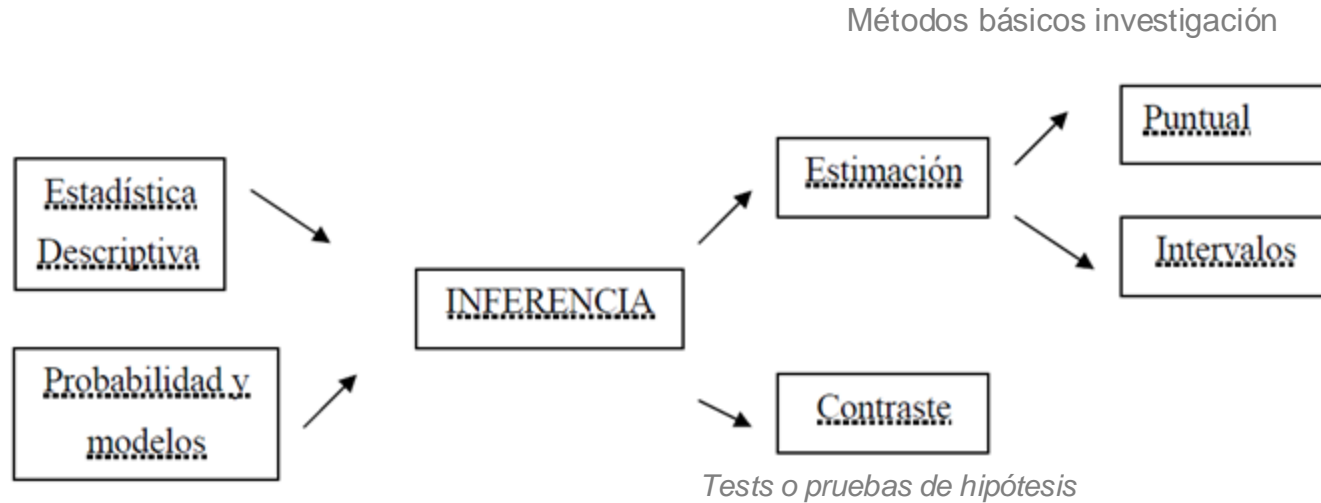
Antecedentes teóricos

- En la cotidianidad muchas situaciones poseen resultados diferentes aunque las condiciones iniciales en las que se producen son las mismas, (manifestación física de situaciones que **envuelven incertidumbre**). Estos fenómenos son denominados **fenómenos aleatorios**.
- El objetivo del **Cálculo de Probabilidades** es el estudio de **métodos de análisis del comportamiento** de **fenómenos aleatorios**.
- **Cálculo de Probabilidades** y la **Estadística** son disciplinas íntimamente relacionadas, **ambas se refieren al estudio de un mismo tipo de situaciones**.
- Mientras que **Cálculo de probabilidades desarrolla los modelos teóricos** para tratar fenómenos aleatorios, la **Estadística ajusta dichos modelos** a situaciones concretas.



Capítulo VI – Clasificación Bayesiana

Antecedentes teóricos



Antecedentes teóricos

Paralelismo entre la **Estadística** y el **Cálculo de Probabilidades**:

ESTADÍSTICA	PROBABILIDAD
f_i, F_i	Probabilidad
Variable Unidimensional	Variable aleatoria
Variable Bidimensional	Vectores aleatorios
Distribución de frecuencias	Distribución de Probabilidad (Función de distribución)
Medias, Momentos	Esperanza, Momentos
Independencia Estadística	Independencia Estocástica
Series Temporales	Procesos Estocásticos



Variable aleatoria v.a.

- Este concepto surge de la necesidad de **calcular probabilidades de conjuntos de interés definidos** en términos de **dicha variable**.
- Es un valor numérico que corresponde al resultado de un **experimento aleatorio** (las v.a. pueden ser *discretas o continuas*).
- Permiten definir la probabilidad como una **función numérica de variable real** (en lugar de como una función de un conjunto dado).
- Dado un espacio de probabilidad (Ω, \mathcal{A}, P) , una variable aleatoria es cualquier función X :

$$X : \Omega \longrightarrow \mathbb{R}. \quad \mathcal{A} = \mathcal{P}(\Omega); \quad A_i \in \mathcal{A}$$

- Cada valor de X se corresponde con el subconjunto de puntos de Ω que se aplica en dicho valor, esto es $\{\omega \in \Omega \mid X(\omega) = x\}$.

Distribución de probabilidad de una variable aleatoria

- Cuando se considera una v.a. X sobre un espacio de probabilidad $(\Omega, \mathcal{A}, \Pr)$, los únicos sucesos de interés son los que se expresan en términos de esta variable, bajo la forma:

$$\{\omega \in \Omega / X(\omega) \in B\} = X^{-1}(B) = \{X \in B\} \in \mathcal{A}$$

- Las probabilidades de estos sucesos describen completamente el comportamiento de la variable X y dan lugar a lo que se denomina la **distribución de probabilidad** de X .

Distribución de probabilidad de una variable aleatoria

La función de distribución de una variable aleatoria X satisface:

- Es monótona (no decreciente).
- Es continua a la derecha.
- $\lim_{x \rightarrow +\infty} F(x) = 1$ y $\lim_{x \rightarrow -\infty} F(x) = 0$

Teorema de Bayes

Planteada por el filósofo inglés *Thomas Bayes* (1702-1761), expresa la **probabilidad condicional de un evento aleatorio A dado B** en términos de la **distribución de probabilidad condicional del evento B dado A** y la **distribución de probabilidad marginal** de sólo A .



Teorema de Bayes

- Sea $\{A_1, A_2, \dots, A_n\}$ un conjunto de **sucesos mutuamente excluyentes y exhaustivos**, tal que la probabilidad de cada uno de ellos es distinta de cero (0).
- Sea B un suceso cualquiera del que se **conocen las probabilidades condicionales** $P(B|A_i)$.
Entonces, la probabilidad $P(A_i|B)$ viene dada por la expresión:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)}$$

Dónde:

$P(A_i)$ son las **probabilidades a priori** de la hipótesis A_i .

$P(B)$ es la **probabilidad** de observar el **conjunto de entrenamiento** B .

$P(B | A_i)$ es la **probabilidad** de observar el **conjunto de entrenamiento** B en un universo donde se verifica la **hipótesis** A_i .

$P(A_i|B)$ es la **probabilidad a posteriori** de A_i , cuando se ha observado el **conjunto de entrenamiento** B .

Teorema de Bayes

- Es válido en **todas las aplicaciones** de la teoría de la probabilidad.
- La **Estadística Bayesiana** está demostrando su utilidad en ciertas estimaciones basadas en el conocimiento subjetivo a priori y el hecho de permitir revisar esas estimaciones en función de la evidencia empírica es lo que está abriendo nuevas formas de hacer conocimiento.
- Se utiliza también en herramientas de Inteligencia de Negocios (BI).
- Como observación, se tiene:

$$\sum_{i=1}^n P(A_i|B) = 1$$

Clasificación a priori

La **probabilidad a priori** $p(c_i)$ será la probabilidad de que el próximo sujeto se clasifique en la clase c_i con $i=1, 2$.

$$p(c_i) = \lim_{n \rightarrow \infty} \frac{n_i}{n}$$

sí $p(c_1) > p(c_2)$ entonces $\text{Clase}(x) = c_1$

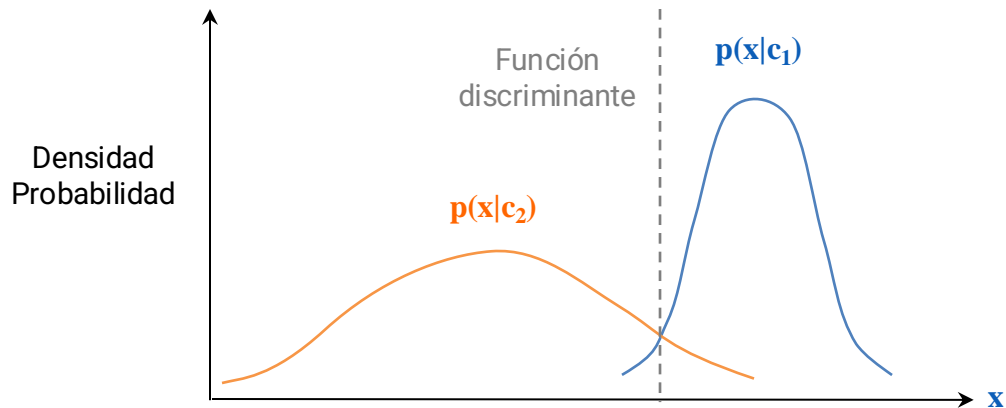
sí $p(c_2) > p(c_1)$ entonces $\text{Clase}(x) = c_2$

Para un sujeto en particular la probabilidad de **error de la clasificación** será:

$$p(\text{error de clasificación}) = \begin{cases} p(c_2) & \text{si se decide } C = c_1 \\ p(c_1) & \text{si se decide } C = c_2 \end{cases}$$

Clasificación condicionada

- El interés será contar con **funciones de densidad de probabilidad condicionales** $p(x|c_i)$, con $i=1,2$.



- $p(x|c_i)$ es la función de densidad de probabilidades condicionales para un valor de x dada la clase.

Clasificación condicionada

- Usando las definiciones anteriores se tiene el *Teorema de Bayes*;

$$p(c_i|x) = \frac{p(x|c_i)p(c_i)}{p(x)}$$

Clasificación condicionada

- Usando las definiciones anteriores se tiene el *Teorema de Bayes*;

$$p(c_i|x) = \frac{p(x|c_i)p(c_i)}{p(x)} \quad p(x) = \sum_{i=1}^2 p(x|c_i)p(c_i)$$

Clasificación condicionada

- Usando las definiciones anteriores se tiene el *Teorema de Bayes*;

$$p(c_i|x) = \frac{p(x|c_i)p(c_i)}{p(x)} \quad p(x) = \sum_{i=1}^2 p(x|c_i)p(c_i) \quad p(c_i|x) = \frac{p(x|c_i)p(c_i)}{\sum_{i=1}^2 p(x|c_i)p(c_i)}$$

- Para obtener el **conocimiento a posteriori** solo se requiere el **conocimiento a priori** y el **conocimiento de la verosimilitud** de las clases c_i respecto a x .

Minimización del Riesgo Condicional

- **Maximización de la Probabilidad a Posteriori (MAP):** Al clasificar una v.a. x , se asigna a la clase c_i que tenga el mayor valor de $p(c_i|x)$.

$$Clase_{MAP} = \arg \max_{c_i} \{p(c_i / x)\} \quad \text{ó} \quad Clase_{MAP} = \arg \max_{c_i} \{p(x / c_i)p(c_i)\}$$

- La mejor clasificación será aquella que **minimice la probabilidad de error** (*regla de clasificación estadística*).

Clasificación multivariada

- Dado un sujeto con un vector de características \vec{x} determinar la clase a la cual pertenece dicho sujeto $p(c_i/\vec{x})$

$$Clase_{MAP} = \arg \max_{c_i} \{p(\vec{x}/c_i)p(c_i)\}$$

- Definiendo la probabilidad como la **función discriminante** $d_i(\vec{x}) = p(\vec{x}/c_i)P(c_i)$ y su logaritmo $L_i(\vec{x}) = \ln d_i(\vec{x})$

$$L_i(\vec{x}) = \ln \{p(\vec{x}/c_i)P(c_i)\}$$

Clasificador Bayesiano Ingenuo

- El clasificador óptimo obtenido por MAP supone los atributos reales y dependientes entre sí.
- Al generalizar, se pueden considerar diferentes tipos de atributos $\{a_1, a_2, \dots, a_p\}$ que pueden ser *reales*, *nominales* o *binarios* y el clasificador óptimo será determinado por:

$$Clase_{MAP} = \arg \max \{p(c_i / a_1, a_2, \dots, a_p)\}$$

- o al usar la verosimilitud y la probabilidad a priori:

$$Clase_{MAP} = \arg \max_{c_i} \{p(a_1, a_2, \dots, a_p / c_i) p(c_i)\}$$

Clasificador Bayesiano Ingenuo

- Dada una clase, la probabilidad de observar la conjunción de los $\{a_1, a_2, \dots, a_p\}$ corresponde al producto de las probabilidades individuales de los atributos:

$$p(a_1, a_2, \dots, a_p / c_i) = \prod_{j=1}^p P(a_j / c_i)$$

- Sustituyendo esto en el clasificador Bayesiano óptimo obtenido por MAP, se tiene el **clasificador Bayesiano ingenuo**:

$$Clase_{NBC} = \arg \max_{c_i} \left\{ p(c_i) \prod_{j=1}^p P(a_j / c_i) \right\}$$

- Así, es muy fácil estimar los $p(a_j/c_i)$ del conjunto de datos.

Clasificador Bayesiano Ingenuo

En caso de **atributos nominales**:

Ej: Suponiendo que a_j puede tomar 4 valores $k \in \{0,1,2,3\}$ entonces:

$$\hat{p}(a_j = k / c_i) = \frac{\text{Nº casos en que } a_j = k \text{ en la clase } c_i}{\text{Total de casos de la clase } i}$$

En **atributos reales** se puede asumir una *distribución normal*.

Se estima la media \bar{a}_j y la varianza $\hat{\sigma}_{a_j}^2$ del atributo y se obtiene la estimación de la probabilidad como:

$$\hat{p}(a_j / c_i) = \frac{1}{\hat{\sigma}_{a_j} \sqrt{2\pi}} e^{-\frac{(a_j - \bar{a}_j)^2}{2\hat{\sigma}_{a_j}^2}}$$

Clasificador Bayesiano Ingenuo

En caso de **atributos nominales**:

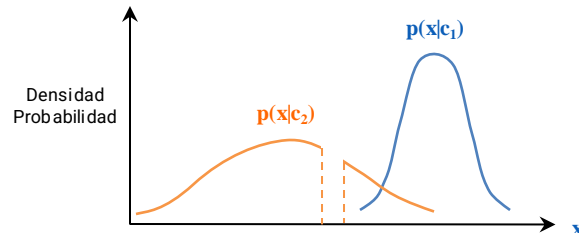
Ej: Suponiendo que a_j puede tomar 4 valores $k \in \{0,1,2,3\}$ entonces:

$$\hat{p}(a_j = k / c_i) = \frac{N^{\circ} \text{ casos en que } a_j = k \text{ en la clase } c_i}{\text{Total de casos de la clase } i}$$

En **atributos reales** se puede asumir una *distribución normal*.

Se estima la media \bar{a}_j y la varianza $\hat{\sigma}_{a_j}^2$ del atributo y se obtiene la estimación de la probabilidad como:

$$\hat{p}(a_j / c_i) = \frac{1}{\hat{\sigma}_{a_j} \sqrt{2\pi}} e^{-\frac{(a_j - \bar{a}_j)^2}{2\hat{\sigma}_{a_j}^2}}$$



Clasificador Bayesiano Ingenuo

En caso de **atributos nominales**:

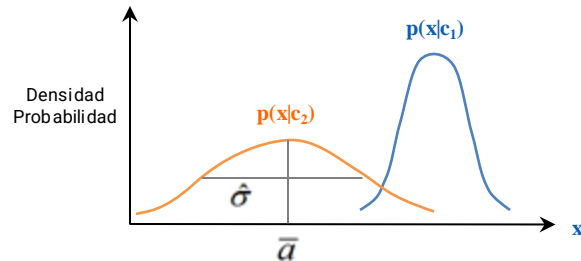
Ej: Suponiendo que a_j puede tomar 4 valores $k \in \{0,1,2,3\}$ entonces:

$$\hat{p}(a_j = k / c_i) = \frac{N^{\circ} \text{ casos en que } a_j = k \text{ en la clase } c_i}{\text{Total de casos de la clase } i}$$

En **atributos reales** se puede asumir una *distribución normal*.

Se estima la media \bar{a}_j y la varianza $\hat{\sigma}_{a_j}^2$ del atributo y se obtiene la estimación de la probabilidad como:

$$\hat{p}(a_j / c_i) = \frac{1}{\hat{\sigma}_{a_j} \sqrt{2\pi}} e^{-\frac{(a_j - \bar{a}_j)^2}{2\hat{\sigma}_{a_j}^2}}$$



Pregunta 1. Accidentabilidad conductores

- Diseñe un **clasificador Bayesiano ingenuo** para el siguiente problema. Una compañía de seguros especializada en pólizas de automóviles, requiere tener un modelo que determine si el postulante a un seguro **tendrá o no accidente**, basado en las variables **sexo** (M=masculino; F=femenino) y el **tipo de auto** que conduce (P=particular; T=trabajo).

Pregunta 1. Accidentabilidad conductores - Desarrollo

Para diseñar el clasificador se cuenta con la siguiente BD, donde se indica si el cliente ha tenido accidentes previamente:

#	Sexo	Tipo	Accidente
1	M	P	Sí
2	F	P	No
3	M	P	Sí
4	F	T	No
5	M	T	No
6	F	P	No
7	M	T	Sí
8	F	P	Sí

Pregunta 1. Accidentabilidad conductores - Desarrollo

$$Clase_{NBC} = \arg \max_{c_i} \left\{ p(c_i) \prod_{j=1}^p P(a_j / c_i) \right\}$$

Pregunta 1. Accidentabilidad conductores - Desarrollo

$$Clase_{NBC} = \arg \max_{c_i} \left\{ p(c_i) \prod_{j=1}^p P(a_j / c_i) \right\}$$

Probabilidad a priori $P(C_i)$

$$P(C_1) = P(\text{Accidente=Sí}) = 4/8 = 0,5 = 50\%$$

$$P(C_2) = P(\text{Accidente=No}) = 4/8 = 0,5 = 50\%$$

Pregunta 1. Accidentabilidad conductores - Desarrollo

$$Clase_{NBC} = \arg \max_{c_i} \left\{ p(c_i) \prod_{j=1}^p P(a_j / c_i) \right\}$$

Probabilidad a priori $P(C_i)$

$$P(C_1) = P(\text{Accidente=Sí}) = 4/8 = 0,5 = 50\%$$

$$P(C_2) = P(\text{Accidente=No}) = 4/8 = 0,5 = 50\%$$

Verosimilitudes $P(x/C_i)$

$$P(\text{Sexo=M} \mid \text{Accidente=Sí}) = 3/4 = 0,75 = 75\%$$

$$P(\text{Sexo=M} \mid \text{Accidente=No}) = 1/4 = 0,25 = 25\%$$

$$P(\text{Sexo=F} \mid \text{Accidente=Sí}) = 1/4 = 0,25 = 25\%$$

$$P(\text{Sexo=F} \mid \text{Accidente=No}) = 3/4 = 0,75 = 75\%$$

$$P(\text{Tipo=P} \mid \text{Accidente=Sí}) = 3/4 = 0,75 = 75\%$$

$$P(\text{Tipo=P} \mid \text{Accidente=No}) = 2/4 = 0,50 = 50\%$$

$$P(\text{Tipo=T} \mid \text{Accidente=Sí}) = 1/4 = 0,25 = 25\%$$

$$P(\text{Tipo=T} \mid \text{Accidente=No}) = 2/4 = 0,50 = 50\%$$

Pregunta 1. Accidentabilidad conductores - Desarrollo

$$Clase_{NBC} = \arg \max_{c_i} \left\{ p(c_i) \prod_{j=1}^p P(a_j / c_i) \right\}$$

Ejemplo Clasificar: <Sexo=F, Tipo=T>

$P(\text{Accidente}=\text{Sí}) P(\text{Sexo}=\text{F}|\text{Accidente}=\text{Sí}) P(\text{Tipo}=\text{T}|\text{Accidente}=\text{Sí}) = (0,5) (0,25) (0,25) = 0,03125$

$P(\text{Accidente}=\text{No}) P(\text{Sexo}=\text{F}|\text{Accidente}=\text{No}) P(\text{Tipo}=\text{T}|\text{Accidente}=\text{No}) = (0,5) (0,75) (0,50) = 0,1875$

$c_1 \quad c_2$

$Clase_{NB} = \arg \max \{ 0,03125; 0,1875 \}$

Solución: Accidente=No

Pregunta 3. Evaluación de riesgo en pacientes

- Estime las probabilidades para diseñar un **clasificador Bayesiano Ingenuo** que clasifique el **riesgo de los pacientes** (riesgoso / no riesgoso), utilizando los datos de sus **edades** y **tipos de cirugías** (codificados en 1, 2 o 3), incluidos en la siguiente tabla.
- Aplicando el clasificador determine si los pacientes de 25 años con cirugía tipo 2 son riesgosos o no.

Capítulo VI – Clasificación Bayesiana

Pregunta 3. Evaluación de riesgo en pacientes

Base de datos (20 observaciones).

Edad paciente	Tipo de cirugía	Riesgo
15	3	Riesgoso
55	1	No Riesgoso
36	1	No Riesgoso
20	2	Riesgoso
40	1	No Riesgoso
17	3	Riesgoso
33	1	No Riesgoso
18	1	No Riesgoso
62	3	Riesgoso
26	2	No Riesgoso
38	2	No Riesgoso
57	3	Riesgoso
22	3	No Riesgoso
70	1	Riesgoso
16	1	No Riesgoso
72	1	No Riesgoso
24	1	No Riesgoso
45	2	No Riesgoso
15	2	No Riesgoso
26	2	No Riesgoso

Pregunta 3. Evaluación de riesgo en pacientes – Desarrollo

$$Clase_{NBC} = \arg \max_{c_i} \left\{ p(c_i) \prod_{j=1}^p P(a_j / c_i) \right\}$$

Probabilidad a priori $P(C_i)$

$$P(C_1) = P(\text{Riesgo}=\text{No Riesgoso}) = 14/20 = 0,7 = 70\%$$

$$P(C_2) = P(\text{Riesgo}=\text{Riesgoso}) = 6/20 = 0,3 = 30\%$$

Pregunta 3. Evaluación de riesgo en pacientes – Desarrollo

$$Clase_{NBC} = \arg \max_{c_i} \left\{ p(c_i) \prod_{j=1}^p P(a_j / c_i) \right\}$$

Probabilidad a priori $P(C_i)$

$$P(C_1) = P(\text{Riesgo}=\text{No Riesgoso}) = 14/20 = 0,7 = 70\%$$

$$P(C_2) = P(\text{Riesgo}=\text{Riesgoso}) = 6/20 = 0,3 = 30\%$$

Verosimilitudes $P(x|C_i)$

$$P(\text{Tipo de cirugía}=1|\text{No Riesgoso}) = 8/14 = 0,571 = 57,1\%$$

$$P(\text{Tipo de cirugía}=1|\text{Riesgoso}) = 1/6 = 0,167 = 16,7\%$$

$$P(\text{Tipo de cirugía}=2|\text{No Riesgoso}) = 5/14 = 0,357 = 35,7\%$$

$$P(\text{Tipo de cirugía}=2|\text{Riesgoso}) = 1/6 = 0,167 = 16,7\%$$

$$P(\text{Tipo de cirugía}=3|\text{No Riesgoso}) = 1/14 = 0,071 = 7,1\%$$

$$P(\text{Tipo de cirugía}=3|\text{Riesgoso}) = 4/6 = 0,667 = 66,7\%$$

Pregunta 3. Evaluación de riesgo en pacientes – Desarrollo

$$Clase_{NBC} = \arg \max_{c_i} \left\{ p(c_i) \prod_{j=1}^p P(a_j / c_i) \right\}$$

Probabilidad a priori $P(C_i)$

$$P(C_1) = P(\text{Riesgo}=\text{No Riesgoso}) = 14/20 = 0,7 = 70\%$$

$$P(C_2) = P(\text{Riesgo}=\text{Riesgoso}) = 6/20 = 0,3 = 30\%$$

Verosimilitudes $P(x|C_i)$

Edad Paciente|No Riesgoso =

{ 55, 36, 40, 33, 18, 26, 38, 22, 16, 72, 24, 45, 15, 26 }

$$\mu(\text{Edad Paciente}|\text{Riesgo}=\text{No Riesgoso}) = 33,3$$

$$\sigma(\text{Edad Paciente}|\text{Riesgo}=\text{No Riesgoso}) = 16,1$$

$$\text{DN Edad Paciente}|\text{No Riesgoso} = N(\bar{\mu}=33.3, \sigma=16.1)$$

Pregunta 3. Evaluación de riesgo en pacientes – Desarrollo

$$Clase_{NBC} = \arg \max_{c_i} \left\{ p(c_i) \prod_{j=1}^p P(a_j / c_i) \right\}$$

Probabilidad a priori $P(C_i)$

$$P(C_1) = P(\text{Riesgo}=\text{No Riesgoso}) = 14/20 = 0,7 = 70\%$$

$$P(C_2) = P(\text{Riesgo}=\text{Riesgoso}) = 6/20 = 0,3 = 30\%$$

Verosimilitudes $P(x|C_i)$

$$\text{Edad Paciente}|\text{Riesgoso} = \{ 15, 20, 17, 62, 57, 70 \}$$

$$\mu(\text{Edad Paciente}|\text{Riesgo}=\text{Riesgoso}) = 40,2$$

$$\sigma(\text{Edad Paciente}|\text{Riesgo}=\text{Riesgoso}) = 25,4$$

$$\text{DN Edad Paciente}|\text{Riesgoso} = N(\mu=33.3, \sigma=16.1)$$

P3. Evaluación de riesgo en pacientes

Función de Densidad de
Probabilidad (Normal) Gaussiana

$$Clase_{NBC} = \arg \max_{c_i} \left\{ p(c_i) \prod_{j=1}^p P(a_j / c_i) \right\}$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{(2\sigma^2)}}$$

Clasificar: <Edad Paciente=25, Cirugía=2>

$P(C_1) = P(\text{Riesgo}=\text{No Riesgoso}) P(\text{Edad Paciente}=25 \mid \text{No Riesgoso}) P(\text{Cirugia}=2 \mid \text{No Riesgoso}) =$
 $(0,7) (0,021) (0,357) = 0,005$

$P(C_2) = P(\text{Riesgo}=\text{Riesgoso}) P(\text{Edad Paciente}=25 \mid \text{Riesgoso}) P(\text{Cirugia}=2 \mid \text{Riesgoso}) =$
 $(0,3) (0,013) (0,167) = 0,0006$

Solución: ClaseNB = argmax { 0,005; 0,0006 } => Riesgo=No Riesgoso

Aprendizaje Bayesiano

El aprendizaje se puede ver como el proceso de encontrar la hipótesis más probable, dado un conjunto de ejemplos de entrenamiento B y un *conocimiento a priori* sobre la probabilidad de cada hipótesis.

Importancia

- Los algoritmos de aprendizaje bayesiano pueden calcular probabilidades explícitas para cada hipótesis.
- También nos proporcionan un marco para estudiar otros algoritmos de aprendizaje.

Razonamiento Bayesiano

- Provee un **enfoque probabilístico** de la inferencia.
- Está basado en asumir que las **incógnitas de interés** siguen distribuciones probabilísticas.
- Es posible obtener una **solución óptima** por medio de estas distribuciones y datos observados.
- Nos da la posibilidad de realizar una **ponderación de la posibilidad** de ocurrencia de una **hipótesis de manera cuantitativa**.

Capítulo VII – Árboles de Decisión

- Comprender la generación de un árbol de decisión.
- Cuantificar la ganancia de información para un atributo en un conjunto de datos.
- Comprender los algoritmos de generación de los árboles de decisión.
- Establecer los mecanismos de poda de los árboles de decisión.
- Comprender los mecanismos de equivalencia de reglas y la generalización de reglas simples.

Definición: **Sistema de Información Operacional**

Considere una base de datos o *Sistema de Información Operacional* compuesto por la 4-tupla:

$$SI = \langle U, Q, V, f \rangle$$

Dónde:

$S \subseteq U$ es un universo cerrado (conjunto finito) no vacío de n objetos $[x_1, x_2, \dots, x_n]$.

Q es un conjunto finito, no vacío, de p atributos $[q_1, q_2, \dots, q_p]$.

$V = \bigcup_{q \in Q} V_i^q$, donde V_i^q representa el dominio o valores que pueden adquirir los atributos q .

$f: S \times Q \rightarrow V$ es una función de decisión llamada *función de información*, tal que $f(x, q) \in V^q$ para cualquier $q \in Q, x \in S$.



Capítulo VII – Árboles de Decisión

- El SI puede ser representado por una tabla finita de datos, donde las columnas están indicadas por los atributos y las filas por los objetos.

Objeto	Atributos						
S	q_1	q_2	...	q_j	...	q_{p-1}	q_p
x_1	V_1^1	V_2^2	V_3^j			V_1^{p-1}	V_1^p
x_2	V_3^1	V_1^2	V_2^j			V_2^{p-1}	V_2^p
· ·	V_4^1	V_4^2	V_4^j			V_2^{p-1}	V_k^p

- Se denominan atributos estudiantes a los atributos comprendidos entre q_1 a q_{p-1} .
- Se denomina atributo experto o de características al atributo q_p que separa los n objetos en k clases $\{V_1^p, V_2^p, \dots, V_k^p\}$.

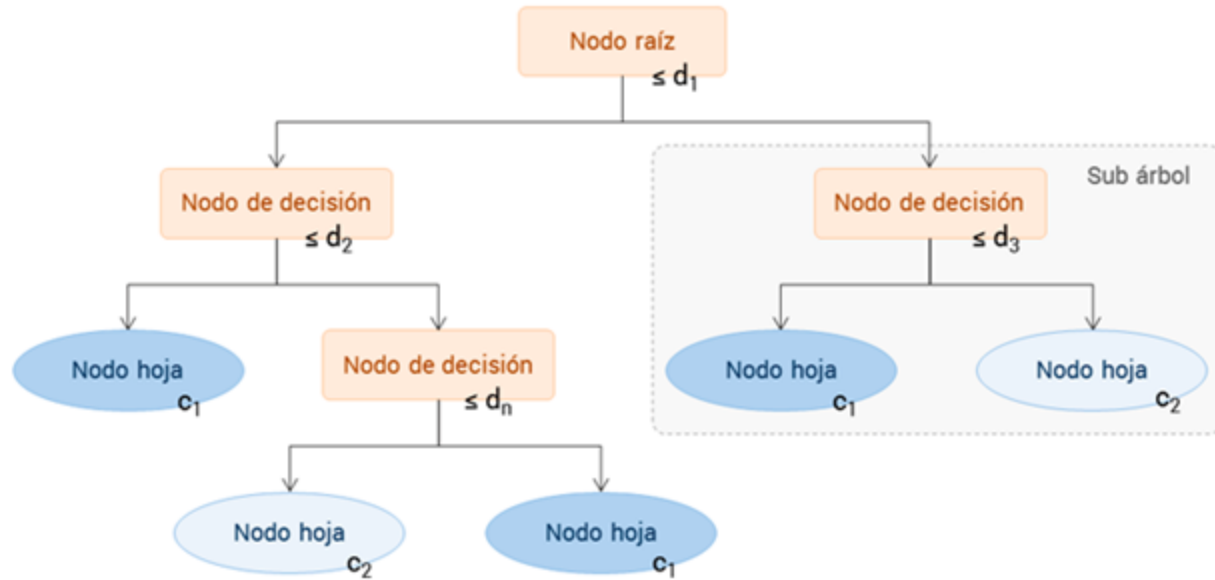


Capítulo VII – Árboles de Decisión

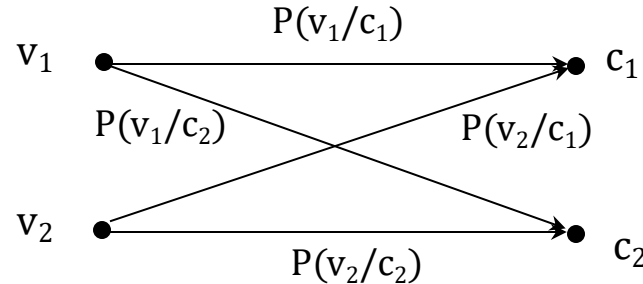
- ■ La idea inicial del método de Hunt es construir un árbol de decisión desde un conjunto de casos de entrenamiento S que consiste de n ejemplos, pertenecientes a k diferentes clases $\{C_1, C_2, \dots, C_k\}$ indicadas por el atributo experto q_p .
- La tarea es dividir el conjunto de entrenamiento S en conjuntos disjuntos T_1, T_2, \dots, T_n ; creando una partición, basada en una característica simple.



Árbol de decisión



Cálculo de Entropía



Se mide la independencia de V_i y C_k mediante probabilidad conjunta $p(v_i \cap c_k)$ o de juntura $p(v_i; c_k)$.

$$\frac{p(v_i; c_k)}{P(v_i)P(c_k)}$$

- Si son independientes la relación será 1.
- Si son completamente dependientes $p(v_i; c_k) = p(v_i) = p(c_k)$ con lo cual la relación será o $1/p(v_i)$ o $1/p(c_k)$.

Cálculo de Entropía de la información

Para que esta medida sea cero en el caso de independencia, se toma el logaritmo en base 2 de la relación, resultando una medida de información, en bit.

$$\text{ld}\left(\frac{p(v_i; c_k)}{P(v_i)P(c_k)}\right) [bit]$$

- Esta medida es cero en el caso de ser independientes V y C.
- En el caso de ser completamente dependientes la información es: $-\text{ld}(p(v_i))$ o $-\text{ld}(p(c_k))$.

Cálculo de la Ganancia de información

Para cuantificar la relación de dependencia entre cualquiera de los atributos estudiantes V_j y el atributo experto C , se toma la **Ganancia de información** (promedio de la información) entre los atributos.

$$Ganancia(v^j, c) = \sum_i \sum_k p(v_i; c_k) \log \left(\frac{p(v_i; c_k)}{P(v_i)P(c_k)} \right)$$

Usando la definición de probabilidad condicional $p(v_i; c_k) = p(c_k / v_i) p(v_i)$ se puede separar en:

$$Ganancia(V, C) = \sum_i \sum_k p(v_i; c_k) \log(p(c_k / v_i)) - \sum_i \sum_k p(v_i; c_k) \log(p(c_k))$$

- La **información de la clase C condicionada** (particionada) **por la instancia v_i** del atributo estudiante V_j .

$$\text{Inf}(C/v_i) = - \sum_k p(c_k / v_i) \log(p(c_k / v_i))$$

El **promedio de la información** de C condicionada por V_j será la ponderación de la información particionada:

$$\text{Inf}(C/V) = \sum_i P(v_i) \text{inf}(C / v_i)$$

$$\text{Inf}(C) = \sum_k P(c_k) \log P(c_k)$$

La **ganancia (de información)** será $\text{Ganancia}(V) = \text{Inf}(C) - \text{Inf}(C/V)$

Capítulo VII – Árboles de Decisión

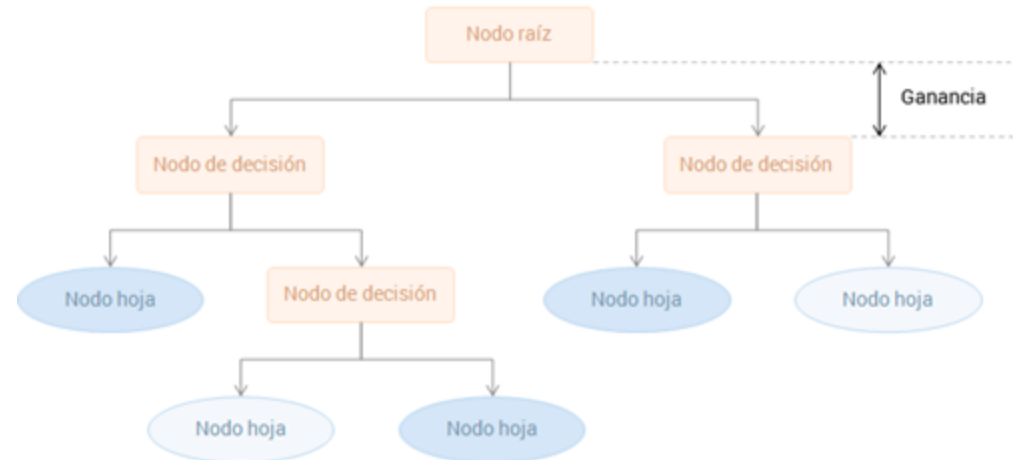
Con esta ganancia es posible determinar cual de los atributos estudiantes V^j ($j=1..p-1$) **separa o caracteriza de una forma más adecuada las clases** c_k .

Para realizar esto se calcula:

$$\text{Max (Ganancia}(V^j))$$

j

El atributo j será la raíz del árbol.



Principales pasos del modelo Árboles de Decisión

- Cálculo de Ganancia o Razón de Ganancia (variables mixtas).

$$\text{Razón Ganancia}(V) = \text{Ganancia}(V) / \text{Inf}(V)$$

- Pre poda o poda en Árboles de Decisión (probabilidad de error a través de la binomial).
- Transformación de Árboles de Decisión en reglas.
 - Reducción de reglas redundantes.

Pregunta 1. Razón de ganancia

Para la siguiente base de datos muestre los cálculos de la razón de ganancia propuestos por Quinlan, para **seleccionar el atributo que presenta mejor discriminación de la clase**.

A	B	Clase
0	0	Sí
0	0	No
0	1	No
0	1	Sí
1	0	Sí
1	1	Sí
1	0	Sí
1	1	Sí
2	0	No
2	0	No
2	1	Sí
2	1	Sí

Capítulo VII – Árboles de Decisión

Pregunta 1. Razón de ganancia – Desarrollo

Atributos estudiantes

- $A = \{ 0, 1, 2 \}$
- $B = \{ 0, 1 \}$

Atributo experto

- Clase = { Sí, No }

#	A	B	Clase
1	0	0	Sí
2	0	0	No
3	0	1	No
4	0	1	Sí
5	1	0	Sí
6	1	1	Sí
7	1	0	Sí
8	1	1	Sí
9	2	0	No
10	2	0	No
11	2	1	Sí
12	2	1	Sí

Capítulo VII – Árboles de Decisión

Pregunta 1. Razón de ganancia – Desarrollo

Probabilidades atributos estudiantes

- $P(A=0) = 4/12 = 1/3 = 0,33$
- $P(A=1) = 4/12 = 1/3 = 0,33$
- $P(A=2) = 4/12 = 1/3 = 0,33$
- $P(B=0) = 6/12 = 1/2 = 0,50$
- $P(B=1) = 6/12 = 1/2 = 0,50$

#	A	B	Clase
1	0	0	Sí
2	0	0	No
3	0	1	No
4	0	1	Sí
5	1	0	Sí
6	1	1	Sí
7	1	0	Sí
8	1	1	Sí
9	2	0	No
10	2	0	No
11	2	1	Sí
12	2	1	Sí

Capítulo VII – Árboles de Decisión

Pregunta 1. Razón de ganancia – Desarrollo

Probabilidades atributos estudiantes

- $P(A=0) = 4/12 = 1/3 = 0,33$
- $P(A=1) = 4/12 = 1/3 = 0,33$
- $P(A=2) = 4/12 = 1/3 = 0,33$
- $P(B=0) = 6/12 = 1/2 = 0,50$
- $P(B=1) = 6/12 = 1/2 = 0,50$

Probabilidades Atributo experto

- $P(\text{Clase}=Sí) = 8/12 = 2/3 = 0,67$
- $P(\text{Clase}=No) = 4/12 = 1/3 = 0,33$

#	A	B	Clase
1	0	0	Sí
2	0	0	No
3	0	1	No
4	0	1	Sí
5	1	0	Sí
6	1	1	Sí
7	1	0	Sí
8	1	1	Sí
9	2	0	No
10	2	0	No
11	2	1	Sí
12	2	1	Sí

Capítulo VII – Árboles de Decisión

Pregunta 1. Razón de ganancia – Desarrollo

Partición 1

	Sí	No	Entropía	Ganancia = $\text{Inf}(C) - \text{Inf}(C/V)$
	8	4		
A=0	2	2		Ganancia(C/A)
A=1	2	2		
A=2	2	2		
B=0	3	3		Ganancia(C/B)
B=1	5	1		

Capítulo VII – Árboles de Decisión

Pregunta 1. Razón de ganancia – Desarrollo

Partición 1

$$Inf(C) = -\sum_k P(c_k) \log P(c_k)$$


	Sí	No	Entropía	Ganancia = $Inf(C) - Inf(C/V)$
	8	4		
A=0	2	2		Ganancia(C/A)
A=1	2	2		
A=2	2	2		
B=0	3	3		Ganancia(C/B)
B=1	5	1		

Capítulo VII – Árboles de Decisión

Pregunta 1. Razón de ganancia – Desarrollo

Partición 1

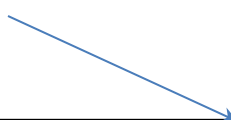
$$Inf(C) = -\sum_k P(c_k) \log P(c_k)$$


	Sí	No	Entropía	Ganancia = $Inf(C) - Inf(C/V)$
	8	4	$Inf(C) = -(P(Clase=Sí) \log(P(Clase=Sí)) + P(Clase=No) \log(P(Clase=No)))$	
A=0	2	2		Ganancia(C/A)
A=1	2	2		
A=2	2	2		
B=0	3	3		Ganancia(C/B)
B=1	5	1		

Capítulo VII – Árboles de Decisión

Pregunta 1. Razón de ganancia – Desarrollo

Partición 1

$$Inf(C/V) = \sum_i P(v_i) inf(C / v_i)$$


	Sí	No	Entropía	Ganancia = Inf(C)-Inf(C/V)
	8	4	$Inf(C) = -(P(Clase=Sí) \log(P(Clase=Sí)) + P(Clase=No) \log(P(Clase=No)))$	
A=0	2	2		Ganancia(C/A)
A=1	2	2		
A=2	2	2		
B=0	3	3		Ganancia(C/B)
B=1	5	1		

Capítulo VII – Árboles de Decisión

Pregunta 1. Razón de ganancia – Desarrollo

Partición 1

$$Inf(C/V) = - \sum_k p(c_k / v_i) \log(p(c_k / v_i))$$

$$Inf(C/V) = \sum_i P(v_i) inf(C / v_i)$$

Diagram showing the relationship between the two equations: An arrow points from the term $inf(C / v_i)$ in the second equation to the term $Inf(C/V)$ in the first equation. Another arrow points from the term $inf(C / v_i)$ in the second equation to the term $inf(C/V)$ in the first equation.

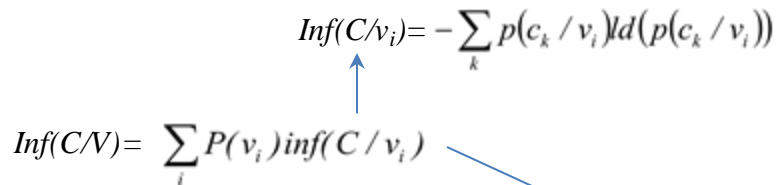
	Sí	No	Entropía	Ganancia = $Inf(C) - Inf(C/V)$
	8	4	$Inf(C) = -(P(Clase=Sí) \log(P(Clase=Sí)) + P(Clase=No) \log(P(Clase=No)))$	
A=0	2	2		
A=1	2	2		
A=2	2	2		
B=0	3	3		
B=1	5	1		

Capítulo VII – Árboles de Decisión

Pregunta 1. Razón de ganancia – Desarrollo

Partición 1

$$Inf(C/V) = - \sum_k p(c_k / v_i) \log(p(c_k / v_i))$$

$$Inf(C/V) = \sum_i P(v_i) inf(C / v_i)$$


	Sí	No	Entropía	Ganancia = Inf(C)-Inf(C/V)
	8	4	$Inf(C) = -(P(Clase=Sí) \log(P(Clase=Sí)) + P(Clase=No) \log(P(Clase=No)))$	
A=0	2	2	$Inf(C A=0) = -(P(Clase=Sí A=0) \log(P(Clase=Sí A=0)) + P(Clase=No A=0) \log(P(Clase=No A=0)))$	Ganancia(C/A)
A=1	2	2	$Inf(C A=1) = -(P(Clase=Sí A=1) \log(P(Clase=Sí A=1)) + P(Clase=No A=1) \log(P(Clase=No A=1)))$	
A=2	2	2	$Inf(C A=2) = -(P(Clase=Sí A=2) \log(P(Clase=Sí A=2)) + P(Clase=No A=2) \log(P(Clase=No A=2)))$	
B=0	3	3		
B=1	5	1		

Capítulo VII – Árboles de Decisión

Pregunta 1. Razón de ganancia – Desarrollo

Partición 1

$$Inf(C/V) = - \sum_k p(c_k / v_i) \log(p(c_k / v_i))$$

$$Inf(C/V) = \sum_i P(v_i) inf(C / v_i)$$

Diagram showing the relationship between the two equations: An arrow points from the term $inf(C / v_i)$ in the second equation to the term $Inf(C/v_i)$ in the first equation. Another arrow points from the term $inf(C / v_i)$ in the second equation to the term $Ganancia(C/V)$ in the table below.

	Sí	No	Entropía	Ganancia = $Inf(C) - Inf(C/V)$
	8	4	$Inf(C) = -(P(Clase=Sí) \log(P(Clase=Sí)) + P(Clase=No) \log(P(Clase=No)))$	
A=0	2	2	$Inf(C A=0) = -(P(Clase=Sí A=0) \log(P(Clase=Sí A=0)) + P(Clase=No A=0) \log(P(Clase=No A=0)))$	Ganancia(C/A)
A=1	2	2	$Inf(C A=1) = -(P(Clase=Sí A=1) \log(P(Clase=Sí A=1)) + P(Clase=No A=1) \log(P(Clase=No A=1)))$	
A=2	2	2	$Inf(C A=2) = -(P(Clase=Sí A=2) \log(P(Clase=Sí A=2)) + P(Clase=No A=2) \log(P(Clase=No A=2)))$	
B=0	3	3	$Inf(C B=0) = -(P(Clase=Sí B=0) \log(P(Clase=Sí B=0)) + P(Clase=No B=0) \log(P(Clase=No B=0)))$	Ganancia(C/B)
B=1	5	1	$Inf(C B=1) = -(P(Clase=Sí B=1) \log(P(Clase=Sí B=1)) + P(Clase=No B=1) \log(P(Clase=No B=1)))$	

Capítulo VII – Árboles de Decisión

Pregunta 1. Razón de ganancia – Desarrollo

Partición 1

$$Inf(C/V) = \sum_i P(v_i) \inf(C / v_i)$$

	Sí	No	Entropía	Ganancia = $Inf(C) - Inf(C/V)$
	8	4	$Inf(C) = -(P(Clase=Sí) \log(P(Clase=Sí)) + P(Clase=No) \log(P(Clase=No)))$	
A=0	2	2	$Inf(C A=0) = -(P(Clase=Sí A=0) \log(P(Clase=Sí A=0)) + P(Clase=No A=0) \log(P(Clase=No A=0)))$	$Ganancia(A) = Inf(C) - (P(A=0) * Inf(C A=0) + P(A=1) * Inf(C A=1) + P(A=2) * Inf(C A=2))$
A=1	2	2	$Inf(C A=1) = -(P(Clase=Sí A=1) \log(P(Clase=Sí A=1)) + P(Clase=No A=1) \log(P(Clase=No A=1)))$	
A=2	2	2	$Inf(C A=2) = -(P(Clase=Sí A=2) \log(P(Clase=Sí A=2)) + P(Clase=No A=2) \log(P(Clase=No A=2)))$	
B=0	3	3	$Inf(C B=0) = -(P(Clase=Sí B=0) \log(P(Clase=Sí B=0)) + P(Clase=No B=0) \log(P(Clase=No B=0)))$	$Ganancia(B) = Inf(C) - (P(B=0) * Inf(C B=0) + P(B=1) * Inf(C B=1))$
B=1	5	1	$Inf(C B=1) = -(P(Clase=Sí B=1) \log(P(Clase=Sí B=1)) + P(Clase=No B=1) \log(P(Clase=No B=1)))$	

Pregunta 1. Razón de ganancia – Desarrollo

Partición 1

	Sí	No	Entropía	Ganancia = $\text{Inf}(C) - \text{Inf}(C/V)$
	8	4	$\text{Inf}(C) = 0,918$	
A=0	2	2	$\text{Inf}(C A=0) = 0,862$	1,206
A=1	2	2	$\text{Inf}(C A=1) = 0,862$	
A=2	2	2	$\text{Inf}(C A=2) = 0,862$	
B=0	3	3	$\text{Inf}(C B=0) = 1,000$	0,831
B=1	5	1	$\text{Inf}(C B=1) = 0,825$	

$$J = \text{Max} (\text{Ganancia}(\{\text{Ganancia}(A), \text{Ganancia}(B)\})) = \text{Ganancia}(A)$$

Capítulo VII – Árboles de Decisión

Pregunta 1. Razón de ganancia – Desarrollo

Partición 1

```

|--- A <= 0.50
|   |--- clase: Sí
|--- A > 0.50
|   |--- B <= 0.50
|       |--- clase: No
|       |--- B > 0.50
|           |--- clase: Sí
    
```

	Sí	No	Entropía	Ganancia = $\text{Inf}(C) - \text{Inf}(C/V)$
	8	4	$\text{Inf}(C) = 0,918$	
A=0	2	2	$\text{Inf}(C A=0) = 0,862$	1,206
A=1	2	2	$\text{Inf}(C A=1) = 0,862$	
A=2	2	2	$\text{Inf}(C A=2) = 0,862$	
B=0	3	3	$\text{Inf}(C B=0) = 1,000$	0,831
B=1	5	1	$\text{Inf}(C B=1) = 0,825$	

$$J = \text{Max} (\text{Ganancia}(\{\text{Ganancia}(A), \text{Ganancia}(B)\})) = \text{Ganancia}(A)$$

Pregunta 3. Posiciones promotor genético

- En genética es muy importante determinar si un gen está expresado en un organismo particular. Una forma de averiguar si el gen está expresado o no, es saber si existe un *promotor* antes del gen. La existencia de un determinado *promotor* se puede medir por la existencia de aminoácidos en el inicio y final del *promotor*. Para el caso de un organismo específico, se sabe que en el inicio del *promotor* existen dos valores posibles: Glicina (G) y Prolina (P); y en el fin del *promotor* existen tres alternativas: Histidina (H), Metionina (M) y Lisina (L).
- Usando el método de máxima entropía propuesto por Quinlan, determine cuál de las dos posiciones (inicio o fin) del *promotor* discrimina mejor su existencia.

Inicio	Fin	Promotor
G	H	Sí
G	H	No
P	H	Sí
G	M	Sí
P	L	Sí
P	M	Sí
P	H	No
G	M	Sí
G	L	No
G	L	No
P	M	Sí
P	L	Sí

Pregunta 3. Posiciones promotor genético

	Sí	No	Entropía	Ganancia = $\text{Inf}(C) - \text{Inf}(C/V)$
	8	4		
Inicio=G	3	3		
Inicio=P	5	1		
Fin=H	2	2		
Fin=L	2	2		
Fin=M	4	0		

$$J = \text{Max} (\text{Ganancia}(\{\text{Ganancia}(\text{Inicio}), \text{Ganancia}(\text{Fin})\}))$$

Pregunta 3. Posiciones promotor genético

$$Inf(C) = -\sum_k P(c_k) \log P(c_k)$$

	Sí	No	Entropía	Ganancia = Inf(C)-Inf(C/V)
	8	4	$Inf(C) = -(P(\text{Promotor}=Sí) \log(P(\text{Promotor}=Sí)) + P(\text{Promotor}=No) \log(P(\text{Promotor}=No)))$	
Inicio=G	3	3		
Inicio=P	5	1		
Fin=H	2	2		
Fin=L	2	2		
Fin=M	4	0		

Capítulo VII – Árboles de Decisión

Pregunta 3. Posiciones promotor genético

$$Inf(C/v_i) = - \sum_k p(c_k / v_i) \log(p(c_k / v_i))$$

$$Inf(C/V) = \sum_i P(v_i) inf(C / v_i)$$

	Sí	No	Entropía	Ganancia = Inf(C)-Inf(C/V)
	8	4	$Inf(C) = -(P(\text{Promotor}=\text{Sí})\log(P(\text{Promotor}=\text{Sí})) + P(\text{Promotor}=\text{No})\log(P(\text{Promotor}=\text{No})))$	
Inicio=G	3	3	$Inf(C Inicio=G) = -(P(\text{Promotor}=\text{Sí} Inicio=G)\log(P(\text{Promotor}=\text{Sí} Inicio=G))) + P(\text{Promotor}=\text{No} Inicio=G)\log(P(\text{Promotor}=\text{Sí} Inicio=G)))$	Ganancia(C/Inicio)
Inicio=P	5	1	$Inf(C Inicio=P) = -(P(\text{Promotor}=\text{Sí} Inicio=P)\log(P(\text{Promotor}=\text{Sí} Inicio=P))) + P(\text{Promotor}=\text{No} Inicio=P)\log(P(\text{Promotor}=\text{Sí} Inicio=P)))$	
Fin=H	2	2	$Inf(C Inicio=H) = -(P(\text{Promotor}=\text{Sí} Inicio=P)\log(P(\text{Promotor}=\text{Sí} Inicio=P))) + P(\text{Promotor}=\text{No} Inicio=P)\log(P(\text{Promotor}=\text{Sí} Inicio=P)))$	
Fin=L	2	2	$Inf(C Fin=L) = -(P(\text{Promotor}=\text{Sí} Fin=L)\log(P(\text{Promotor}=\text{Sí} Fin=L))) + P(\text{Promotor}=\text{No} Fin=L)\log(P(\text{Promotor}=\text{Sí} Fin=L)))$	Ganancia(C/Fin)
Fin=M	4	0	$Inf(C Fin=M) = -(P(\text{Promotor}=\text{Sí} Fin=M)\log(P(\text{Promotor}=\text{Sí} Fin=M))) + P(\text{Promotor}=\text{No} Fin=M)\log(P(\text{Promotor}=\text{Sí} Fin=M)))$	

Pregunta 3. Posiciones promotor genético

$$Inf(C/V) = \sum_i P(v_i) inf(C / v_i)$$

	Sí	No	Entropía	Ganancia = Inf(C)-Inf(C/V)
	8	4	$Inf(C) = -(P(Promotor=Sí)ld(P(Promotor=Sí)) + P(Promotor=No)ld(P(Promotor=No)))$	
Inicio=G	3	3	$Inf(C Inicio=G) = -(P(Promotor=Sí Inicio=G)ld(P(Promotor=Sí Inicio=G)) + P(Promotor=No Inicio=G)ld(P(Promotor=Sí Inicio=G)))$	$Ganancia(Inicio) = Inf(C) - (P(Inicio=G)*Inf(C Inicio=G) + P(Inicio=P)*Inf(C Inicio=P))$
Inicio=P	5	1	$Inf(C Inicio=P) = -(P(Promotor=Sí Inicio=P)ld(P(Promotor=Sí Inicio=P)) + P(Promotor=No Inicio=P)ld(P(Promotor=Sí Inicio=P)))$	
Fin=H	2	2	$Inf(C Fin=H) = -(P(Promotor=Sí Fin=H)ld(P(Promotor=Sí Fin=H)) + P(Promotor=No Fin=H)ld(P(Promotor=Sí Fin=H)))$	$Ganancia(Fin) = Inf(C) - (P(Fin=H)*Inf(C Fin=H) + P(Fin=L)*Inf(C Fin=L) + P(Fin=M)*Inf(C Fin=M))$
Fin=L	2	2	$Inf(C Fin=L) = -(P(Promotor=Sí Fin=L)ld(P(Promotor=Sí Fin=L)) + P(Promotor=No Fin=L)ld(P(Promotor=Sí Fin=L)))$	
Fin=M	4	0	$Inf(C Fin=M) = -(P(Promotor=Sí Fin=M)ld(P(Promotor=Sí Fin=M)) + P(Promotor=No Fin=M)ld(P(Promotor=Sí Fin=M)))$	

Pregunta 3. Posiciones promotor genético

	Sí	No	Entropía	Ganancia = $\text{Inf}(C) - \text{Inf}(C/V)$
	8	4	$\text{Inf}(C) = 0,918$	
Inicio=G	3	3	$\text{Inf}(C \text{Inicio}=G) = 1,000$	Ganancia(Inicio) = 0,831
Inicio=P	5	1	$\text{Inf}(C \text{Inicio}=P) = 0,825$	
Fin=H	2	2	$\text{Inf}(C \text{Fin}=H) = 0,862$	Ganancia(Fin) = 1,094
Fin=L	2	2	$\text{Inf}(C \text{Fin}=L) = 0,862$	
Fin=M	4	0	$\text{Inf}(C \text{Fin}=M) = 0,528$	

Pregunta 3. Posiciones promotor genético

	Sí	No	Entropía	Ganancia = $\text{Inf}(C) - \text{Inf}(C/V)$
	8	4	$\text{Inf}(C) = 0,918$	
Inicio=G	3	3	$\text{Inf}(C \text{Inicio}=G) = 1,000$	Ganancia(Inicio) = 0,831
Inicio=P	5	1	$\text{Inf}(C \text{Inicio}=P) = 0,825$	
Fin=H	2	2	$\text{Inf}(C \text{Fin}=H) = 0,862$	Ganancia(Fin) = 1,094
Fin=L	2	2	$\text{Inf}(C \text{Fin}=L) = 0,862$	
Fin=M	4	0	$\text{Inf}(C \text{Fin}=M) = 0,528$	

$$J = \text{Max} (\text{Ganancia}(\{\text{Ganancia}(\text{Inicio}), \text{Ganancia}(\text{Fin})\})) = \text{Ganancia}(\text{Fin})$$



UNIVERSIDAD
DE SANTIAGO
DE CHILE



Consultas