



# Aprendizaje No Supervisado: Reglas de Asociación

---

CURSO GRUPO BANCOLOMBIA

Universidad Nacional de Colombia



# Reglas de Asociación

---

- Las Reglas de Asociación son una implicación de la forma  $X \rightarrow Y$ , donde  $X$  es el antecedente y  $Y$  es el consecuente de la regla.
- Un ejemplo de Reglas de Asociación es en el análisis del carrito o canasta de compras. Donde se busca encontrar la dependencia entre dos artículos  $X$  y  $Y$ .
- Una aplicación típica es la venta minorista (retail) donde  $X$  y  $Y$  son artículos vendidos.



# Análisis del carrito de compras

El objetivo es encontrar asociaciones entre productos comprados por el usuario.

- Las personas que compran X normalmente compran Y.
- Si hay un cliente que compra X pero no compra Y, entonces es un potencial cliente de Y.
- Una vez encontrados los clientes se realiza una venta cruzada.
- Se espera encontrar la probabilidad condicional de la forma  $P(Y|X)$ , donde Y es el producto que queremos condicionar con X.
- X es el producto o el grupo de productos que sabemos que el cliente ya compró.

Ejemplo:

Si se encuentra que  $P(\text{Papas fritas} | \text{Cerveza}) = 0,7$ , esto quiere decir que el 70% de los clientes que compraron cerveza también compraron papas fritas.



# Análisis del carrito de compras

- Es posible que se quiera distinguir el tipo de cliente,  $P(Y|X, D)$ , donde  $D$  es un conjunto de atributos del cliente, por ejemplo, género, edad, estado civil.

Otros ejemplos:

- Biblioteca, los productos podrían ser los libros o los autores.
- Página web, los productos podrían ser link a otras páginas y nos podríamos estimar los links que un usuario le gustaría visitar. Usaríamos esa información para descargar dichas páginas anticipadamente para que el usuario acceda a ellas más rápido.



# Reglas de Asociación

Hay tres medidas que normalmente se calculan para la regla de asociación  $X \rightarrow Y$ :

- Soporte

$$\text{Support}(Y|X) \equiv P(X|Y) = \frac{\#(\text{personas que compraron } X \text{ y } Y)}{\#(\text{personas})}$$

- Confianza

$$\text{Confidence}(X \rightarrow Y) \equiv P(Y|X) = \frac{P(X|Y)}{P(X)} = \frac{\#(\text{personas que compraron } X \text{ y } Y)}{\#(\text{personas que compraron } X)}$$

- Interés

$$\text{Lift}(X \rightarrow Y) \equiv \frac{P(X|Y)}{P(X)P(Y)} = \frac{P(Y|X)}{P(Y)}$$

Existen otras medidas, pero estas tres son las más usadas. La confianza es la probabilidad condicional de los productos. Para poder decir que la regla se mantiene con confianza, el valor debe ser cercano a 1 y significativamente mayor que  $P(Y)$  que es la probabilidad que las personas compren  $Y$ .



# Reglas de Asociación

- También estamos interesados en maximizar el soporte de la regla, porque incluso si existe una dependencia con un fuerte valor de confianza, si el número de tales clientes es pequeño, la regla no tiene valor.
- El soporte muestra la importancia estadística de la regla, mientras que la confianza muestra la fuerza de la regla.
- El usuario establece los valores mínimos de soporte y confianza.
- Si  $X$  y  $Y$  son independientes, entonces el interés será cercano a 1, es decir  $P(Y|X) \approx P(Y)$ .
- Si  $P(Y|X) \neq P(Y)$  entonces hay una dependencia entre los elementos.
- Si  $Lift(X \rightarrow Y)$  es mayor que 1, entonces  $X$  hace que  $Y$  sea más probable.
- Si  $Lift(X \rightarrow Y)$  es menor que 1, entonces  $X$  hace que  $Y$  sea menos probable.



# Algoritmo *Apriori*

Este es un algoritmo de reglas de asociación muy eficiente que tiene 2 pasos:

1. **Encontrar conjuntos de elementos frecuentes, es decir, aquellos que tienen suficiente soporte (*Support*).**

Primero se encuentran los conjuntos frecuentes de un elemento y en cada paso, se generan conjuntos candidatos del elemento y luego se verifica si tienen suficiente soporte (*Support*). El algoritmo almacena los conjuntos de elementos frecuentes en una tabla para facilitar el acceso.

2. **Convertirlos en reglas con suficiente confianza (*Confidence*), dividiendo los elementos en dos, antecedente y consecuente.**

Se empieza colocando un solo consecuente y los elementos en el antecedente. Después, para todos los posibles consecuentes se verifica si la regla tiene suficiente confianza (*Confidence*).





# ¡A codificar!

---

