

## ITEMSET

È un insieme di  $n$  oggetti.

## K-ITEMSET

Un insieme di  $k$  oggetti.

## SUPPORT (ASSOLUTO)

Metica che indica il numero di occorrenze di un dato itemset.

## SUPPORT (RELATIVO)

È il rapporto fra il numero di transazioni che contengono l'itemset  $I$  e il totale delle transazioni.

Esempio:

$D = \{ \underline{A}, B, C \}$   
 $\{ A, C \}$   
 $\{ \underline{A}, B, D \}$   
 $\{ \underline{A}, B, E \}$

Il supporto di  $I = \{ A, B \}$  è:

Supporto assoluto:

$$\text{supp}_a(I) = 3$$

Supporto relativo:

$$\text{supp}_r(I) = \frac{3}{4} = 75\%$$

## TRANSAZIONI T

Insieme di itemset per cui vale  $T \subseteq D$

## ASSOCIATION RULES

Una regola di associazione ci dà indicazione sul fatto che dati i sottoinsiemi  $X \subset I$  e  $Y \subset I$ , se la regola dice  $X \Rightarrow Y$  allora è molto probabile che in una transazione con  $X$  abbia anche  $Y$ .

L'importante è che  $X \cup Y \neq \emptyset$

### SUPPORTO DI $X \Rightarrow Y$

Ci dà indicazione sulla percentuale di transazioni che presentano contemporaneamente  $X$  e  $Y$  ( $X \cup Y$ )

$$\text{Supp}(X \Rightarrow Y) = P(X \cup Y)$$

### CONFIDENZA DI $X \Rightarrow Y$

È la probabilità condizionata che dato  $X$  ci sia  $Y$ .

$$\text{Conf}(X \Rightarrow Y) = P(X | Y) = \frac{\text{Supp}(X \Rightarrow Y)}{\text{Supp}(X)}$$

Se la confidenza è 1 allora tutte le transazioni aventi  $X$  hanno anche  $Y$ .

### Esempio:

- D:
1. BIRRA, NOCI, PANNOCCINO
  2. BIRRA, CAFFÈ, PANNOCCINO
  3. BIRRA, PANNOCCINO, UOVA
  4. NOCI, UOVA, LATTE
  5. NOCI, CAFFÈ, PANNOCCINO, UOVA, LATTE

Poniamo  $\text{minSupp} = \text{minConf} = 50\%$

Gli itemset frequenti sono:

$$\{B\}: 3 \mid \{N\}: 3 \mid \{P\}: 4 \mid \{U\}: 3 \mid \{B, P\}: 3$$



Due regole associative possono essere  $B \Rightarrow P$  e  $P \Rightarrow B$

$$\begin{aligned} \text{supp}(B \Rightarrow P) &= \frac{3}{5} = 60\% \\ \text{conf}(B \Rightarrow P) &= \frac{3}{3} = 100\% \end{aligned} \quad \left. \vphantom{\begin{aligned} \text{supp}(B \Rightarrow P) &= \frac{3}{5} = 60\% \\ \text{conf}(B \Rightarrow P) &= \frac{3}{3} = 100\% \end{aligned}} \right\} B \Rightarrow P$$

$$\begin{aligned} \text{supp}(P \Rightarrow B) &= \frac{3}{5} = 60\% \\ \text{conf}(P \Rightarrow B) &= \frac{3}{4} = 75\% \end{aligned} \quad \left. \vphantom{\begin{aligned} \text{supp}(P \Rightarrow B) &= \frac{3}{5} = 60\% \\ \text{conf}(P \Rightarrow B) &= \frac{3}{4} = 75\% \end{aligned}} \right\} P \Rightarrow B$$

## CLOSE ITEMSET

Un itemset  $X$  è definito **close** se esso è:

- 1) Frequente in  $D$ .
- 2) **Non** devono esistere itemset  $Y$ :  $Y \supset X$  con un supporto uguale a  $X$ . (differenza con il max-I)  
Ovvero non deve esistere un itemset che contiene  $X$  e che abbia anche supporto uguale a  $X$ .

## MAX ITEMSET

Un itemset  $X$  è definito **max** se esso è:

- 1) Frequente in  $D$ .
- 2) **Non** devono esistere itemset  $Y$  **frequenti**:  $Y \supset X$ .  
Ovvero non deve esistere un itemset che contiene  $X$  e allo stesso tempo sia frequente.

Esempio : MINSUP = 30%

D: {LATTE, PANNE, UOVA}  
{LATTE, PANNE}  
{PANNE, UOVA}  
{LATTE, UOVA}  
{LATTE, PANNE, ZUCCHERO}  
{PANNE, UOVA, FARINA}  
{LATTE, UOVA, ZUCCHERO}  
{LATTE, ZUCCHERO, CAFFE}

FREQUENT ITEMSET

{LATTE} = 75%.  
{PANNE} = 62,5%.  
{UOVA} = 62,5%.  
{ZUCCHERO} = 37,5%.  
{LATTE, PANNE} = 37,5%.  
{PANNE, UOVA} = 37,5%.  
{LATTE, ZUCCHERO} = 37,5%.  
{LATTE, UOVA} = 37,5%.

CROSS-ITEMSETS :

{LATTE}  
{PANNE}  
{UOVA}  
{ZUCCHERO}  
{LATTE, PANNE}  
{PANNE, UOVA}  
{LATTE, ZUCCHERO}  
{LATTE, UOVA}

MAX-ITEMSET

{LATTE, PANNE}  
{PANNE, UOVA}  
{LATTE, ZUCCHERO}  
{LATTE, UOVA}