

# ALGORITMO APRIORI

# EXEMPLO

- 2 passos
- gerar conj. de itens frequentes
  - gerar regras só do conj. de tamanho  $> 1$

de associação  $\{b, c\} \Rightarrow \{I\}$

Regra: nº de Registos  $b, c, I$  / nº total de Registos

se  $\text{conf}_{\min} = 100\%$   
REGRAS DE ASSOCIAÇÃO FORTES

confiança:  $\{A_1, A_4\} \Rightarrow A_3$   
itens preverem o item  $A_3$

com  
 $\frac{\text{nº de Registos } b, c, I}{\text{nº de Registos com } b \text{ e } c}$

$2^n - 1 \Rightarrow$  nº de conjunto de itens

- Conjuntos frequentes de tamanho 1 -  $F_1$ :  $\{1\}, \{2\}, \{3\}, \{5\}$
- Os conjuntos de tamanho 2 são obtidos a partir dos conjuntos frequentes de tamanho 1

como suporte individual  $\leq 50\%$  então podemos descartar o 4, pois qd combinado c/ outros vai ter sempre suporte  $\min \leq 25\%$

Cliente	item 4 é de suporte de 25% max.
100	1 3 (4)
200	2 3 5
300	1 2 3 5
400	2 5

4 linhas  
n=5 itens

Item - Set	Supporte
$\{1, 2\}$	0.25
$\{1, 3\}$	0.5
$\{1, 5\}$	0.25
$\{2, 3\}$	0.5
$\{2, 5\}$	0.75
$\{3, 5\}$	0.5

se o suporte mínimo for 50% então podemos descartar

$F_2$

construção de conj. frequentes  
ex.:  $F_3$

$$\left\{ \begin{array}{l} F_2 \quad \{1,3\}, \{2,3\}, \{2,5\}, \{3,5\} \\ F_1 \quad \{1\}, \{2\}, \{3\}, \{5\} \end{array} \right.$$

$F_3$

Item-Set	Supporte
$\{1, 2, 3\}$	$1/4 = 0.25$
$\{1, 2, 5\}$	$1/4 = 0.25$
$\{1, 3, 5\}$	$1/4 = 0.25$
$\{2, 3, 5\}$	$2/4 = 0.50$

Nº de Registos c/ b, c e I  
Nº Total de Registos

descartar

$$\left\{ \begin{array}{l} F_3 \quad \{2, 3, 5\} \\ F_1 \quad \{1\}, \{2\}, \{3\}, \{5\} \end{array} \right.$$

item-set	supporte
$F_4 \quad \{1, 2, 3, 5\}$	$1/4 = 0.25$

Se os conjuntos forem gerados por enumeração exaustiva, o nº de item-sets no nível  $k$  é dado por:

$$\binom{d}{k} = \frac{d!}{(d-k)! \times k!}$$

$${}^5C_1 + {}^5C_2 + {}^5C_3 + {}^5C_4 = 5 + 10 + 10 + 5 + 1 = 31$$

Com a poda de item-sets baseada no suporte:  $4 + 4 + 1 = 9$

1.  $F_1: \{1\}, \{2\}, \{3\}, \{5\}$
2.  $F_2: \{1,3\}, \{2,3\}, \{2,5\}, \{3,5\}$
3.  $F_3: \{2,3,5\}$

## AVALIAÇÃO DE REGRAS DE ASSOCIAÇÃO

$ABC \rightarrow D$

$$\begin{aligned} \text{confiança} | (ABC \rightarrow D) &= \text{suporte}(ABCD) \\ &= \text{suporte}(ABC) - ABC \text{ combinados} \leq \text{sup}(AB) \leq \text{sup}(A) \end{aligned}$$

$\swarrow$  sempre  
 $\swarrow$  por sua vez sempre

SUPORTE DO ITEM INDIVIDUAL  
SEMPRE MAIOR DO QUE  
QUANDO COMBINADOS  
C/ OUTROS

	caté	$\overline{\text{caté}}$	
chá	15	5	20
$\overline{\text{chá}}$	75	5	80
	90	10	100

chá |  $\text{suporte} = 15/100 = 15\%$

chá |  $\text{confiança} = 15/20 = 75\% \rightarrow$  pode ser enganoso, pois  
 só entao o chá e  
 neste caso o caté aparece +  
 registros