



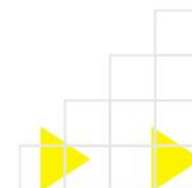
Curso 2 – CD, AM e DM

Mineração de Dados

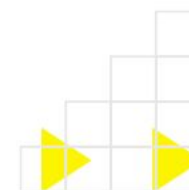
Parte 6

Extração de Padrões
Regras de Associação

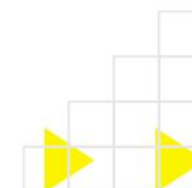
Prof. Ricardo M. Marcacini
ricardo.marcacini@icmc.usp.br



Etapas do Processo de Mineração de Dados



Etapas do Processo de Mineração de Dados

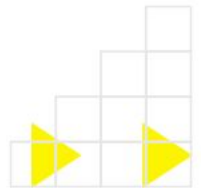


Extração de Padrões



Tarefas Descritivas

Regras de Associação



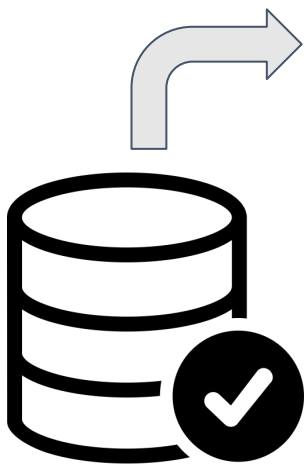
Contextualização e Motivação

Mining association rules between sets of items in large databases.

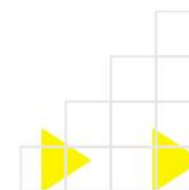
Agrawal, R.; Imieliński, T.; Swami, A.

Proceedings of the ACM SIGMOD International Conference on Management of Data. 1993. pp. 207-216.

- Base de dados com transações de consumidores



Transação	Itens da Transação
#T ₁	I ₁ , I ₂ , I ₄ , I ₇
#T ₂	I ₁ , I ₃ , I ₄
#T ₃	I ₄ , I ₅
...	...
#T _n	I ₁ , I ₂ , I ₇



Contextualização e Motivação

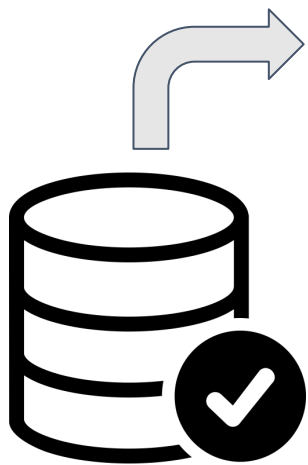


Mining association rules between sets of items in large databases.

Agrawal, R.; Imieliński, T.; Swami, A.

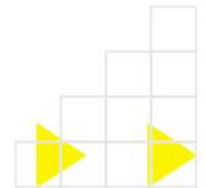
Proceedings of the ACM SIGMOD International Conference on Management of Data. 1993. pp. 207-216.

- Base de dados com transações de consumidores



Transação	Itens da Transação
#T ₁	I ₁ , I ₂ , I ₄ , I ₇
#T ₂	I ₁ , I ₃ , I ₄
#T ₃	I ₄ , I ₅
...	...
#T _n	I ₁ , I ₂ , I ₇

Cada transação consiste de um conjunto de itens



Contextualização e Motivação

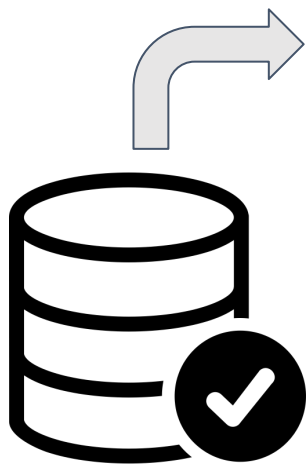


Mining association rules between sets of items in large databases.

Agrawal, R.; Imieliński, T.; Swami, A.

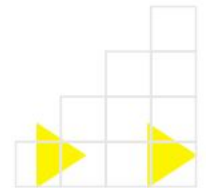
Proceedings of the ACM SIGMOD International Conference on Management of Data. 1993. pp. 207-216.

- Base de dados com transações de consumidores



Transação	Itens da Transação
#T ₁	café, pão, leite, sal
#T ₂	café, açúcar, leite
#T ₃	leite, chocolate
...	...
#T _n	café, pão, sal

Cada transação consiste de um conjunto de itens



Contextualização e Motivação

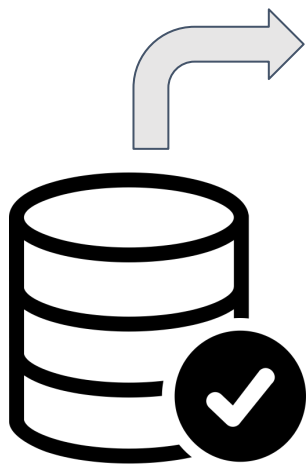


Mining association rules between sets of items in large databases.

Agrawal, R.; Imieliński, T.; Swami, A.

Proceedings of the ACM SIGMOD International Conference on Management of Data. 1993. pp. 207-216.

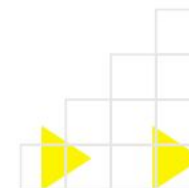
- Base de dados com transações de consumidores



Transação	Itens da Transação
#T ₁	I ₁ , I ₂ , I ₄ , I ₇
#T ₂	I ₁ , I ₃ , I ₄
#T ₃	I ₄ , I ₅
...	...
#T _n	I ₁ , I ₂ , I ₇

Queremos encontrar regras de associação entre os itens

$$A \Rightarrow B$$



Contextualização e Motivação

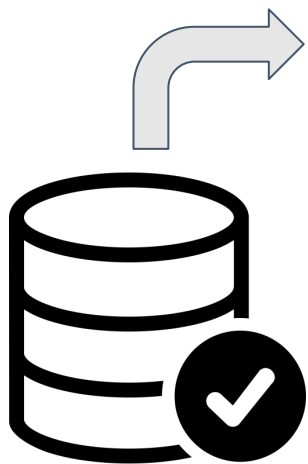


Mining association rules between sets of items in large databases.

Agrawal, R.; Imieliński, T.; Swami, A.

Proceedings of the ACM SIGMOD International Conference on Management of Data. 1993. pp. 207-216.

- Base de dados com transações de consumidores



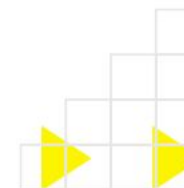
Transação	Itens da Transação
#T ₁	I ₁ , I ₂ , I ₄ , I ₇
#T ₂	I ₁ , I ₃ , I ₄
#T ₃	I ₄ , I ₅
...	...
#T _n	I ₁ , I ₂ , I ₇

Queremos encontrar regras de associação entre os itens

Antecedente

Consequente

$A \Rightarrow B$



Contextualização e Motivação

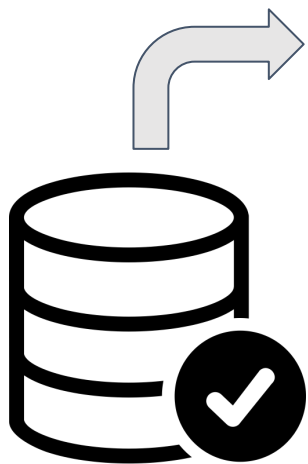


Mining association rules between sets of items in large databases.

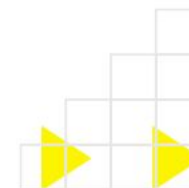
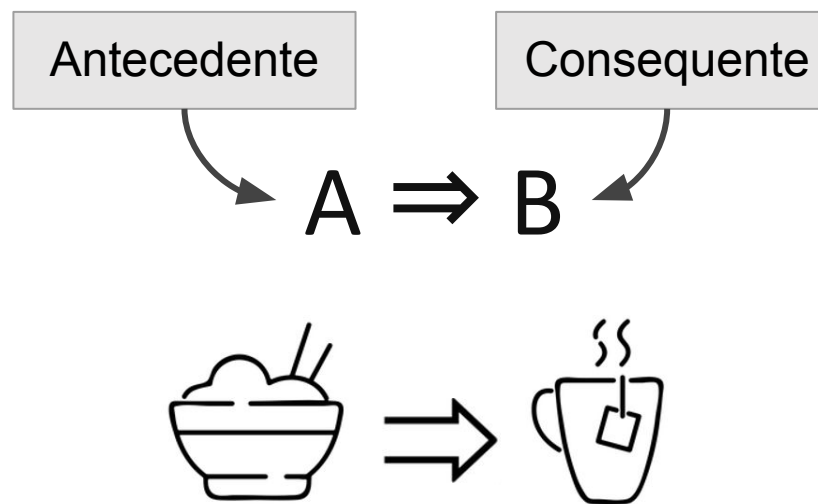
Agrawal, R.; Imieliński, T.; Swami, A.

Proceedings of the ACM SIGMOD International Conference on Management of Data. 1993. pp. 207-216.

- Base de dados com transações de consumidores



Transação	Itens da Transação
#T ₁	I ₁ , I ₂ , I ₄ , I ₇
#T ₂	I ₁ , I ₃ , I ₄
#T ₃	I ₄ , I ₅
...	...
#T _n	I ₁ , I ₂ , I ₇



Contextualização e Motivação

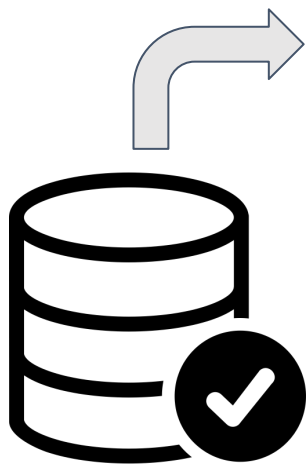


Mining association rules between sets of items in large databases.

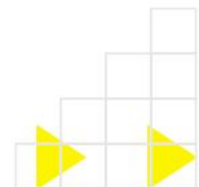
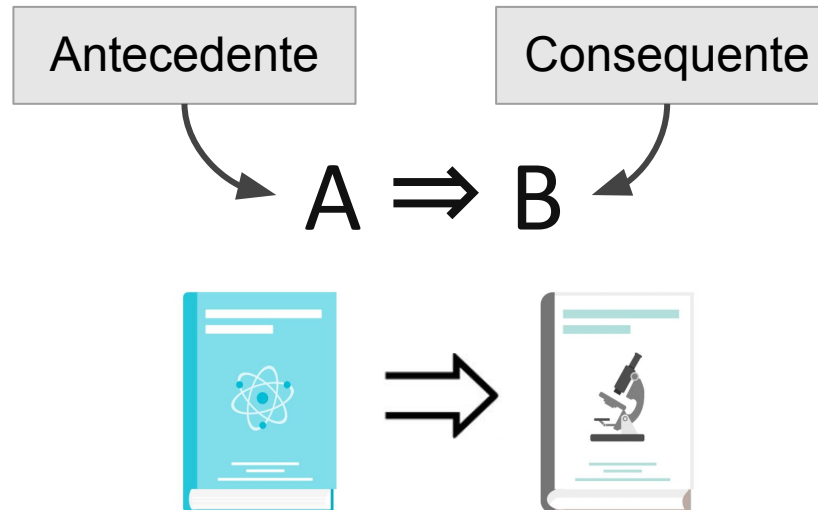
Agrawal, R.; Imieliński, T.; Swami, A.

Proceedings of the ACM SIGMOD International Conference on Management of Data. 1993. pp. 207-216.

- Base de dados com transações de consumidores



Transação	Itens da Transação
#T ₁	I ₁ , I ₂ , I ₄ , I ₇
#T ₂	I ₁ , I ₃ , I ₄
#T ₃	I ₄ , I ₅
...	...
#T _n	I ₁ , I ₂ , I ₇



Contextualização e Motivação

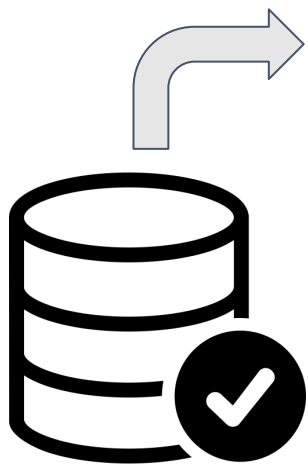


Mining association rules between sets of items in large databases.

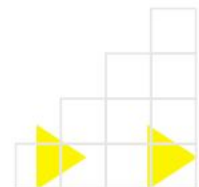
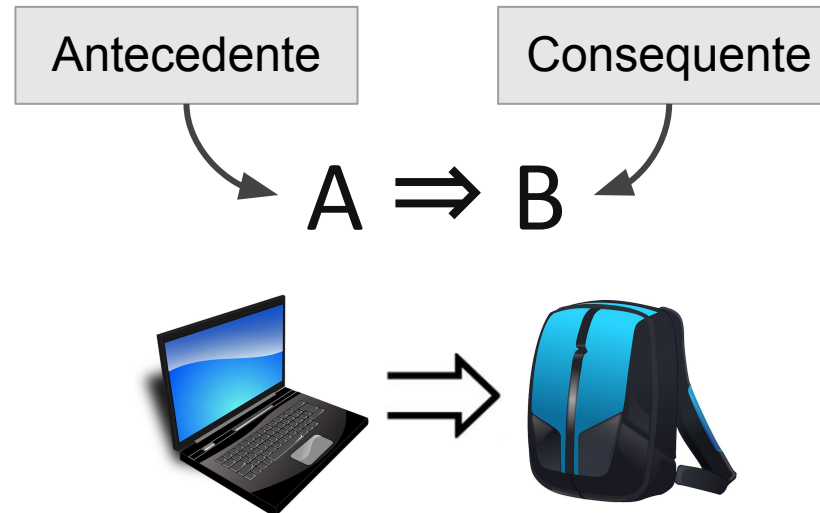
Agrawal, R.; Imieliński, T.; Swami, A.

Proceedings of the ACM SIGMOD International Conference on Management of Data. 1993. pp. 207-216.

- Base de dados com transações de consumidores

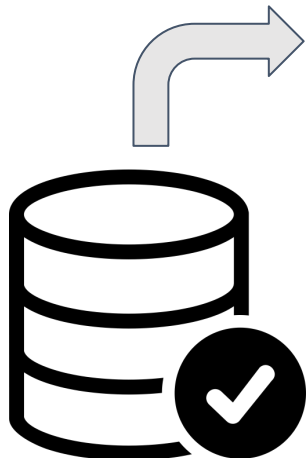


Transação	Itens da Transação
#T ₁	I ₁ , I ₂ , I ₄ , I ₇
#T ₂	I ₁ , I ₃ , I ₄
#T ₃	I ₄ , I ₅
...	...
#T _n	I ₁ , I ₂ , I ₇

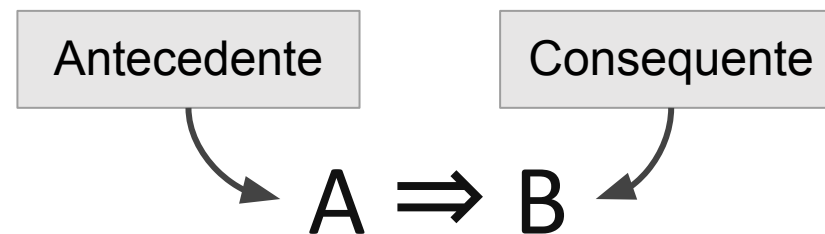


Contextualização e Motivação

- Base de dados com transações de consumidores
- Regras de associação entre os itens (simbólico)
- Técnica não supervisionada

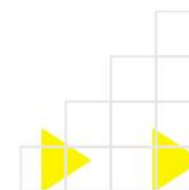


Transação	Itens da Transação
#T ₁	I ₁ , I ₂ , I ₄ , I ₇
#T ₂	I ₁ , I ₃ , I ₄
#T ₃	I ₄ , I ₅
...	...
#T _n	I ₁ , I ₂ , I ₇



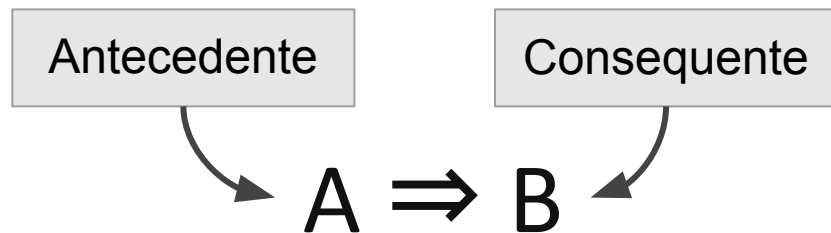
Exemplos de aplicações:

- *Market Basket Analysis*
- *Web Usage Mining*
- *Bioinformática*
- *Medicina*

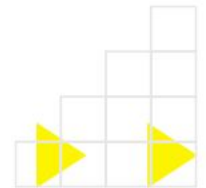


Contextualização e Motivação

- Base de dados com transações de consumidores
- Regras de associação entre os itens
- Técnica não supervisionada

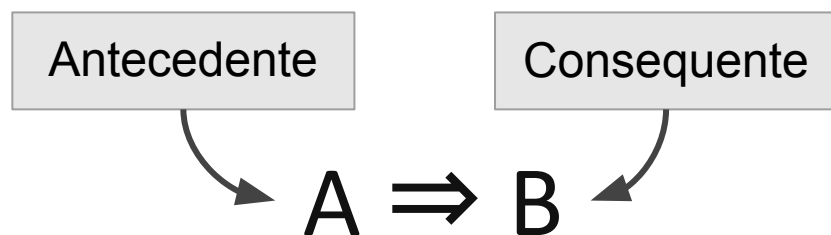


Como extrair regras interessantes
(não triviais) das transações?



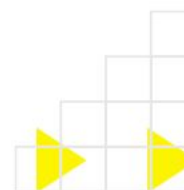
Contextualização e Motivação

- Base de dados com transações de consumidores
- Regras de associação entre os itens
- Técnica não supervisionada



Como extrair regras interessantes
(não triviais) das transações?

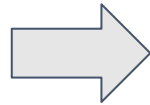
Medidas objetivas para analisar a significância,
correlação e força das regras de associação!



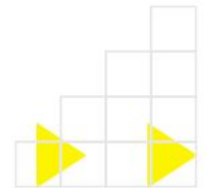
Regras de Associação

- Matriz atributo-valor
- Representação usando atributos binários

T	Itens
#T ₁	I ₁ , I ₂
#T ₂	I ₃
#T ₃	I ₄ , I ₅
#T ₄	I ₁ , I ₂ , I ₃
#T ₅	I ₂



T	I ₁	I ₂	I ₃	I ₄	I ₅
#T ₁	1	1	0	0	0
#T ₂	0	0	1	0	0
#T ₃	0	0	0	1	1
#T ₄	1	1	1	0	0
#T ₅	0	1	0	0	0



Regras de Associação

- Notações

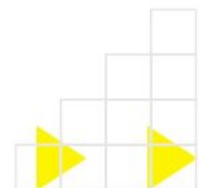
- *Itemset* (conjunto de itens)

$$\mathcal{I} = \{I_1, I_2, \dots, I_m\}$$

- Transação

$$T \subseteq \mathcal{I}$$

T	I_1	I_2	I_3	I_4	I_5
#T ₁	1	1	0	0	0
#T ₂	0	0	1	0	0
#T ₃	0	0	0	1	1
#T ₄	1	1	1	0	0
#T ₅	0	1	0	0	0



Regras de Associação

- Notações

- *Itemset* (conjunto de itens)

$$\mathcal{I} = \{I_1, I_2, \dots, I_m\}$$

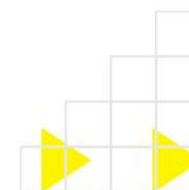
- Transação

$$T \subseteq \mathcal{I}$$

- Regra de Associação: $A \Rightarrow B$

- $A \subset \mathcal{I}, B \subset \mathcal{I}$
- $A \neq \emptyset, B \neq \emptyset$
- $A \cap B = \emptyset$

T	I_1	I_2	I_3	I_4	I_5
#T ₁	1	1	0	0	0
#T ₂	0	0	1	0	0
#T ₃	0	0	0	1	1
#T ₄	1	1	1	0	0
#T ₅	0	1	0	0	0



Regras de Associação

- Notações

- *Itemset* (conjunto de itens)

$$\mathcal{I} = \{I_1, I_2, \dots, I_m\}$$

- Transação

$$T \subseteq \mathcal{I}$$

- Regra de Associação: $A \Rightarrow B$

- $A \subset \mathcal{I}, B \subset \mathcal{I}$
- $A \neq \emptyset, B \neq \emptyset$
- $A \cap B = \emptyset$

T	leite	pão	café	sal	açúcar
#T ₁	1	1	0	0	0
#T ₂	0	0	1	0	0
#T ₃	0	0	0	1	1
#T ₄	1	1	1	0	0
#T ₅	0	1	0	0	0

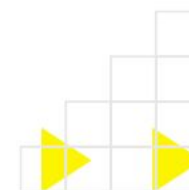
A = {pão, café}

B = {leite}

Regra de Associação:

$A \Rightarrow B$

{pão, café} \Rightarrow {leite}

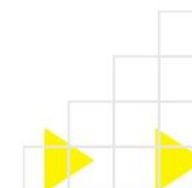


Regras de Associação

- Suporte (relativo)
 - Indica a frequência (relativa) de *itemsets* na base de transações

$$\text{sup}(A \Rightarrow B) = P(A \cup B) = \frac{\text{cont}(A \& B)}{|T|}$$

T	leite	pão	café	sal	açúcar
#T ₁	1	1	0	0	0
#T ₂	0	0	1	0	0
#T ₃	0	0	0	1	1
#T ₄	1	1	1	0	0
#T ₅	0	1	0	0	0



Regras de Associação

- Suporte (relativo)
 - Indica a frequência (relativa) de *itemsets* na base de transações

$$\text{sup}(A \Rightarrow B) = P(A \cup B) = \frac{\text{cont}(A \& B)}{|T|}$$

A = {pão, café}

B = {leite}

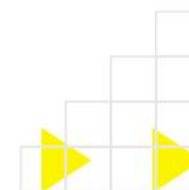
Regra de Associação:

A \Rightarrow B

{pão, café} \Rightarrow {leite}

T	leite	pão	café	sal	açúcar
#T ₁	1	1	0	0	0
#T ₂	0	0	1	0	0
#T ₃	0	0	0	1	1
#T ₄	1	1	1	0	0
#T ₅	0	1	0	0	0

Suporte de {pão, café} \Rightarrow {leite}
é de $\frac{1}{5} = 0.2$ (20%)



Regras de Associação

- Confiança

- Indica quantas vezes itens em B aparecem em transações que contém A

$$conf(A \Rightarrow B) = \frac{sup(A \Rightarrow B)}{sup(A)}$$

A = {pão, café}

B = {leite}

Regra de Associação:

$A \Rightarrow B$

$\{\text{pão, café}\} \Rightarrow \{\text{leite}\}$

T	leite	pão	café	sal	açúcar
#T ₁	1	1	0	0	0
#T ₂	0	0	1	0	0
#T ₃	0	0	0	1	1
#T ₄	1	1	1	0	0
#T ₅	0	1	0	0	0

Suporte de $\{\text{pão, café}\} \Rightarrow \{\text{leite}\} = 0.2$

Suporte de $\{\text{pão, café}\} = 0.2$

Confiança de $\{\text{pão, café}\} \Rightarrow \{\text{leite}\} = 1$



Regras de Associação

- Como minerar regras de associação?

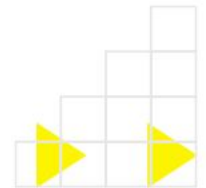
Etapa 1

Encontrar *itemsets* frequentes de acordo com um **suporte mínimo**



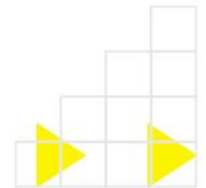
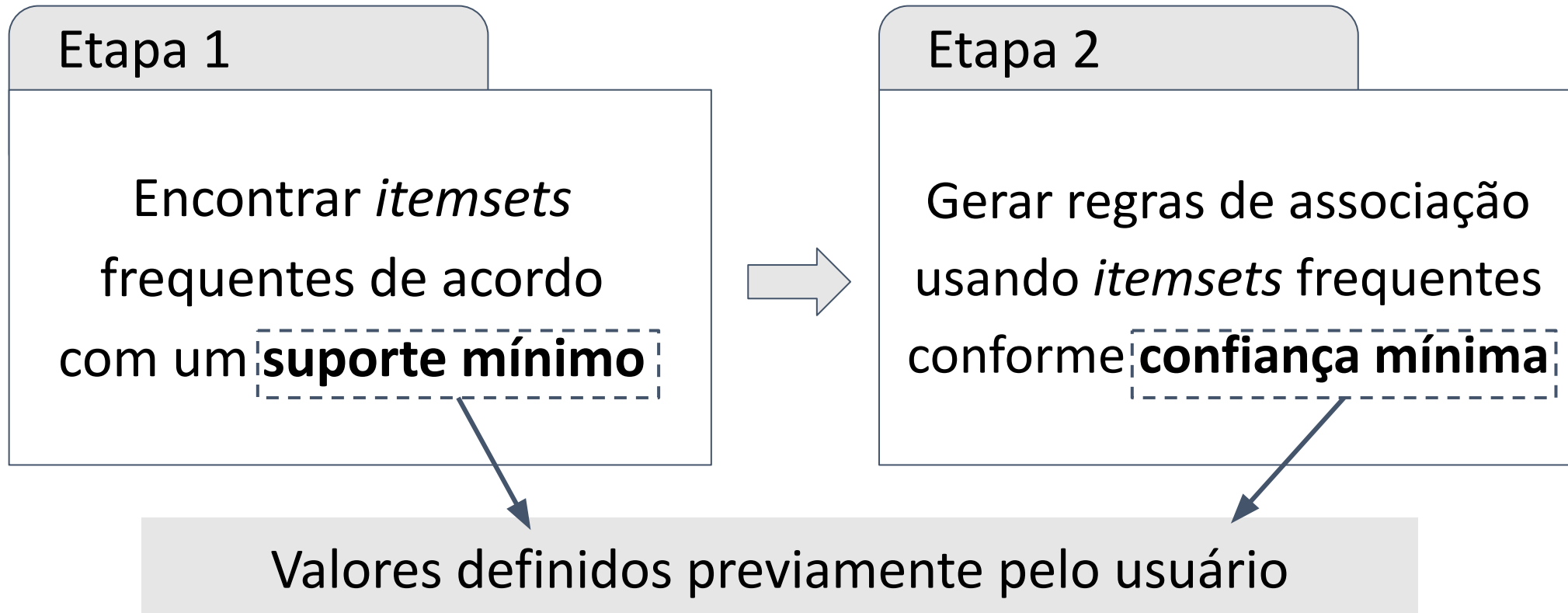
Etapa 2

Gerar regras de associação usando *itemsets* frequentes conforme **confiança mínima**



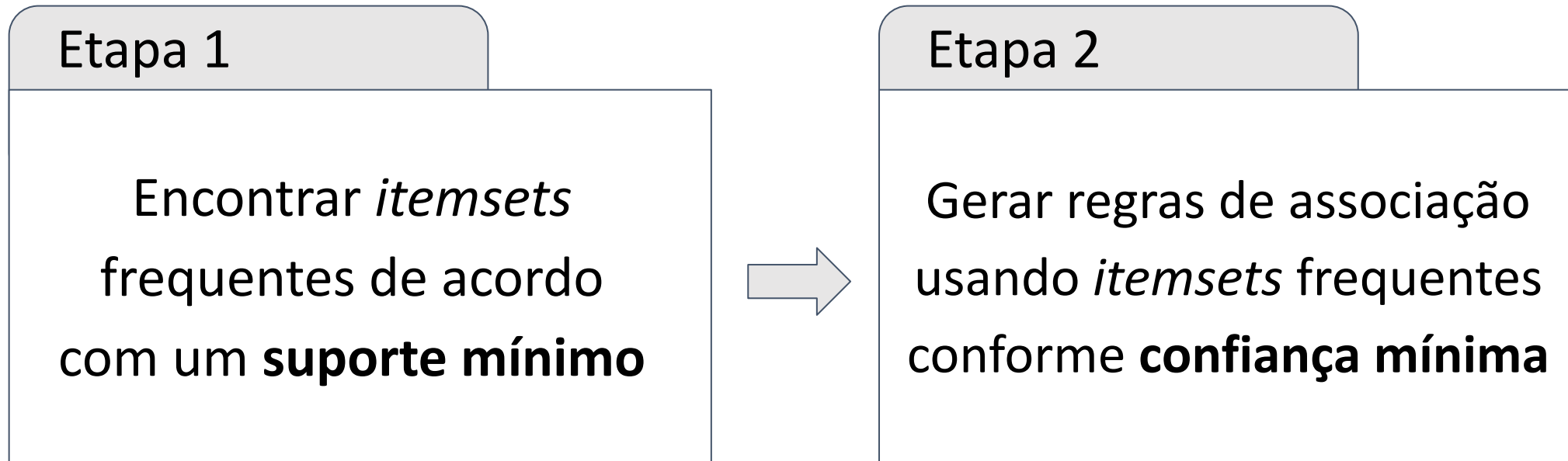
Regras de Associação

- Como minerar regras de associação?

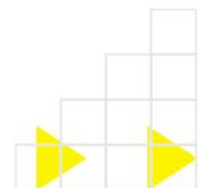


Regras de Associação

- Como minerar regras de associação?

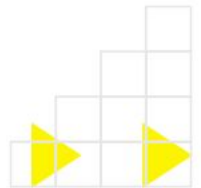


Etapa 1 apresenta maior custo computacional



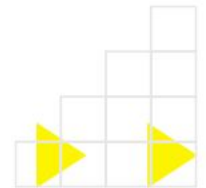
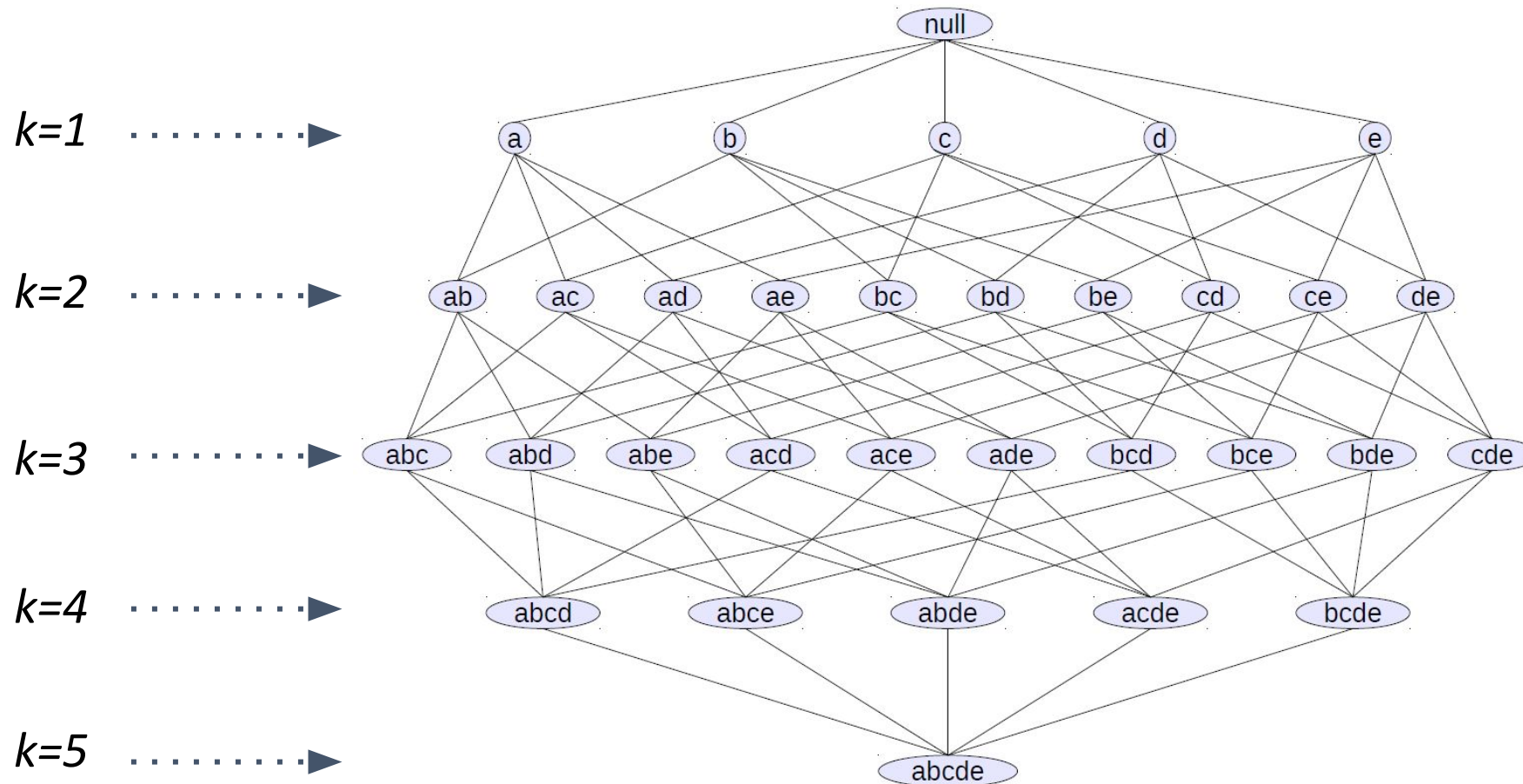
Regras de Associação

- Algoritmo APRIORI para extração de itemsets frequentes
 - Processo iterativo
 - Os *k-itemsets* de uma iteração são utilizados para gerar os próximos *(k+1)-itemsets*



Regras de Associação

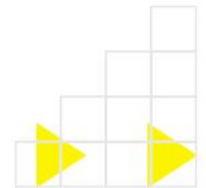
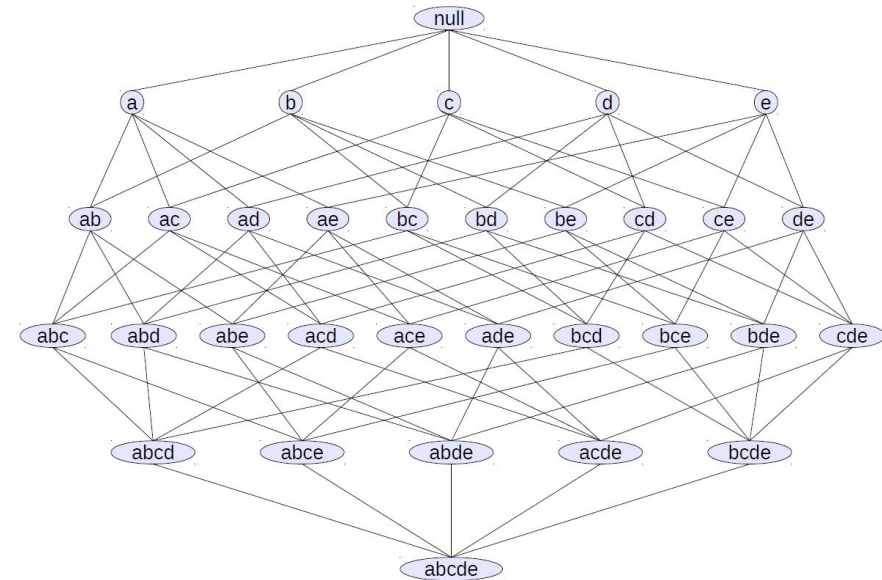
- Algoritmo APRIORI para extração de itemsets frequentes



Regras de Associação

- Algoritmo APRIORI para extração de itemsets frequentes
 - Processo iterativo
 - Os k -itemsets de uma iteração são utilizados para gerar os próximos $(k+1)$ -itemsets

Encontrar todos os itemsets frequentes (força bruta) gera um total de $2^m - 1$ combinações, em que m é a quantidade de itens.

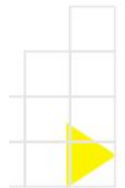
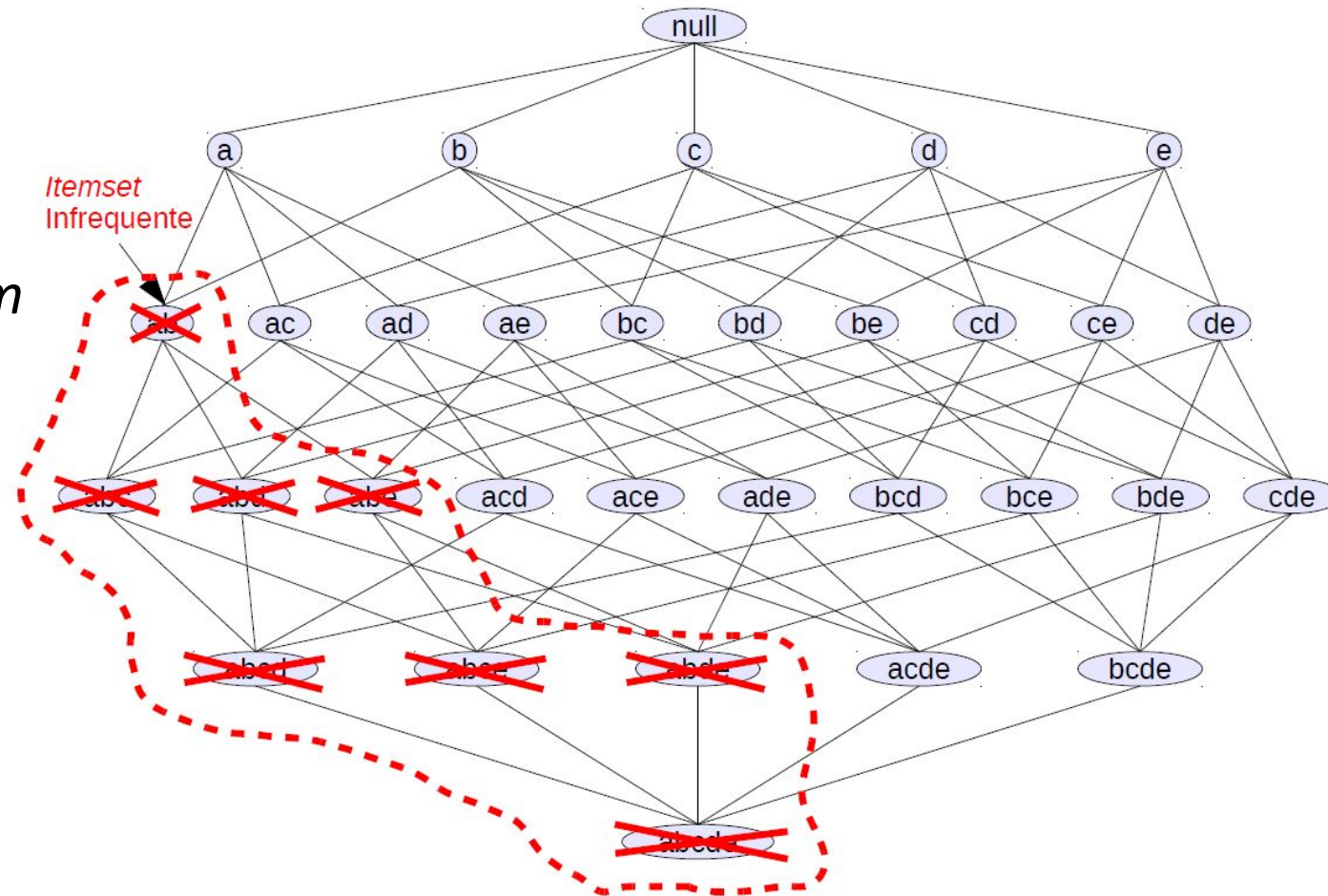


Regras de Associação

- Algoritmo APRIORI para extração de itemsets frequentes

Propriedade

Todo subconjunto de um itemset frequente também é frequente



Regras de Associação



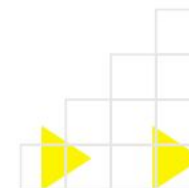
- APRIORI: Exemplo de geração de *itemsets* frequentes

T	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
#T ₁	0	1	0	1	1	0	0
#T ₂	1	0	1	1	1	0	0
#T ₃	0	1	0	1	1	0	0
#T ₄	1	1	0	1	1	0	0
#T ₅	0	0	1	0	0	0	0
#T ₆	0	0	0	0	1	0	0
#T ₇	0	0	0	1	0	0	0
#T ₈	0	0	0	0	0	0	1
#T ₉	0	0	0	0	0	1	1
#T ₁₀	0	0	0	0	0	1	0

Vamos escolher um
Suporte Mínimo = 0.3

Suporte com $k=1$ (1-itemsets frequentes)

- $\text{sup}(\{\text{Leite}\}) = 0.2$
- $\text{sup}(\{\text{Café}\}) = 0.3$
- $\text{sup}(\{\text{Cerveja}\}) = 0.2$
- $\text{sup}(\{\text{Pão}\}) = 0.5$
- $\text{sup}(\{\text{Manteiga}\}) = 0.5$
- $\text{sup}(\{\text{Arroz}\}) = 0.2$
- $\text{sup}(\{\text{Feijão}\}) = 0.2$



Regras de Associação



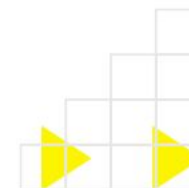
- APRIORI: Exemplo de geração de *itemsets* frequentes

T	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
#T ₁	0	1	0	1	1	0	0
#T ₂	1	0	1	1	1	0	0
#T ₃	0	1	0	1	1	0	0
#T ₄	1	1	0	1	1	0	0
#T ₅	0	0	1	0	0	0	0
#T ₆	0	0	0	0	1	0	0
#T ₇	0	0	0	1	0	0	0
#T ₈	0	0	0	0	0	0	1
#T ₉	0	0	0	0	0	1	1
#T ₁₀	0	0	0	0	0	1	0

Vamos escolher um
Suporte Mínimo = 0.3

Suporte com $k=1$ (1-itemsets frequentes)

- ~~$\text{sup}(\{\text{Leite}\}) = 0.2$~~
- $\text{sup}(\{\text{Café}\}) = 0.3$
- ~~$\text{sup}(\{\text{Cerveja}\}) = 0.2$~~
- $\text{sup}(\{\text{Pão}\}) = 0.5$
- $\text{sup}(\{\text{Manteiga}\}) = 0.5$
- ~~$\text{sup}(\{\text{Arroz}\}) = 0.2$~~
- ~~$\text{sup}(\{\text{Feijão}\}) = 0.2$~~



Regras de Associação

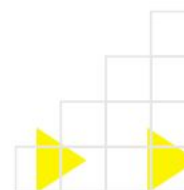
- APRIORI: Exemplo de geração de *itemsets* frequentes

T	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
#T ₁	0	1	0	1	1	0	0
#T ₂	1	0	1	1	1	0	0
#T ₃	0	1	0	1	1	0	0
#T ₄	1	1	0	1	1	0	0
#T ₅	0	0	1	0	0	0	0
#T ₆	0	0	0	0	1	0	0
#T ₇	0	0	0	1	0	0	0
#T ₈	0	0	0	0	0	0	1
#T ₉	0	0	0	0	0	1	1
#T ₁₀	0	0	0	0	0	1	0

Vamos escolher um
Suporte Mínimo = 0.3

Suporte com $k=2$ (2-itemsets frequentes)

- $\text{sup}(\{\text{Café}, \text{Manteiga}\}) = 0.3$
- $\text{sup}(\{\text{Café}, \text{Pão}\}) = 0.3$
- $\text{sup}(\{\text{Manteiga}, \text{Pão}\}) = 0.4$



Regras de Associação

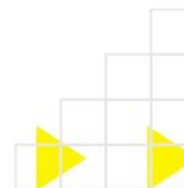
- APRIORI: Exemplo de geração de *itemsets* frequentes

T	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
#T ₁	0	1	0	1	1	0	0
#T ₂	1	0	1	1	1	0	0
#T ₃	0	1	0	1	1	0	0
#T ₄	1	1	0	1	1	0	0
#T ₅	0	0	1	0	0	0	0
#T ₆	0	0	0	0	1	0	0
#T ₇	0	0	0	1	0	0	0
#T ₈	0	0	0	0	0	0	1
#T ₉	0	0	0	0	0	1	1
#T ₁₀	0	0	0	0	0	1	0

Vamos escolher um
Suporte Mínimo = 0.3

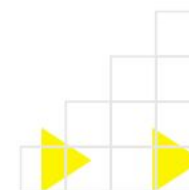
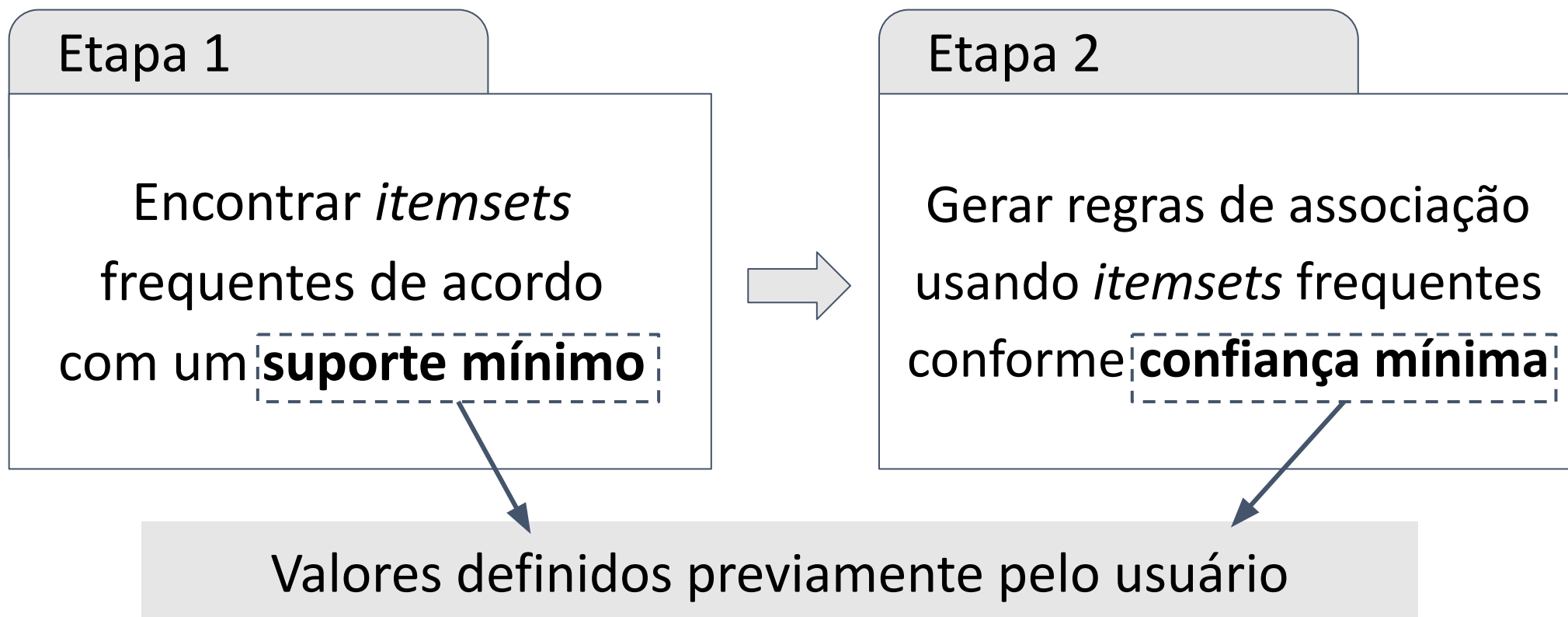
Suporte com $k=3$ (3-itemsets frequentes)

- $\text{sup}(\{\text{Café}, \text{Manteiga}, \text{Pão}\}) = 0.3$



Regras de Associação

- Como minerar regras de associação?



Regras de Associação

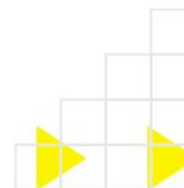
- Exemplo: Calculando confiança de regras de associação

T	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
#T ₁	0	1	0	1	1	0	0
#T ₂	1	0	1	1	1	0	0
#T ₃	0	1	0	1	1	0	0
#T ₄	1	1	0	1	1	0	0
#T ₅	0	0	1	0	0	0	0
#T ₆	0	0	0	0	1	0	0
#T ₇	0	0	0	1	0	0	0
#T ₈	0	0	0	0	0	0	1
#T ₉	0	0	0	0	0	1	1
#T ₁₀	0	0	0	0	0	1	0

Vamos escolher uma
Confiança Mínima = 0.8

Itemset frequente {Café,Manteiga} pode gerar duas regras de associação

- $\{\text{Café}\} \Rightarrow \{\text{Manteiga}\}$
- $\{\text{Manteiga}\} \Rightarrow \{\text{Café}\}$



Regras de Associação

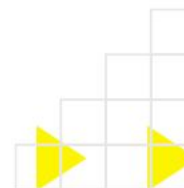
- Exemplo: Calculando confiança de regras de associação

T	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
#T ₁	0	1	0	1	1	0	0
#T ₂	1	0	1	1	1	0	0
#T ₃	0	1	0	1	1	0	0
#T ₄	1	1	0	1	1	0	0
#T ₅	0	0	1	0	0	0	0
#T ₆	0	0	0	0	1	0	0
#T ₇	0	0	0	1	0	0	0
#T ₈	0	0	0	0	0	0	1
#T ₉	0	0	0	0	0	1	1
#T ₁₀	0	0	0	0	0	1	0

Vamos escolher uma
Confiança Mínima = 0.8

Vamos calcular a confiança:

- $\text{conf}(\{\text{Café}\} \Rightarrow \{\text{Manteiga}\}) = 1.0$
- $\text{conf}(\{\text{Manteiga}\} \Rightarrow \{\text{Café}\}) = 0.6$



Regras de Associação

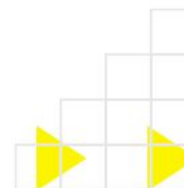
- Exemplo: Calculando confiança de regras de associação

T	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
#T ₁	0	1	0	1	1	0	0
#T ₂	1	0	1	1	1	0	0
#T ₃	0	1	0	1	1	0	0
#T ₄	1	1	0	1	1	0	0
#T ₅	0	0	1	0	0	0	0
#T ₆	0	0	0	0	1	0	0
#T ₇	0	0	0	1	0	0	0
#T ₈	0	0	0	0	0	0	1
#T ₉	0	0	0	0	0	1	1
#T ₁₀	0	0	0	0	0	1	0

Vamos escolher uma
Confiança Mínima = 0.8

Vamos calcular a confiança:

- $\text{conf}(\{\text{Café}\} \Rightarrow \{\text{Manteiga}\}) = 1.0$
- ~~$\text{conf}(\{\text{Manteiga}\} \Rightarrow \{\text{Café}\}) = 0.6$~~



Regras de Associação

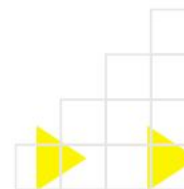
- Exemplo: Calculando confiança de regras de associação

T	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
#T ₁	0	1	0	1	1	0	0
#T ₂	1	0	1	1	1	0	0
#T ₃	0	1	0	1	1	0	0
#T ₄	1	1	0	1	1	0	0
#T ₅	0	0	1	0	0	0	0
#T ₆	0	0	0	0	1	0	0
#T ₇	0	0	0	1	0	0	0
#T ₈	0	0	0	0	0	0	1
#T ₉	0	0	0	0	0	1	1
#T ₁₀	0	0	0	0	0	1	0

Vamos escolher uma
Confiança Mínima = 0.8

Vamos calcular a confiança de regras geradas por {Café,Pão,Manteiga}:

- $\text{conf}(\{\text{Café}, \text{Pão}\} \Rightarrow \{\text{Manteiga}\}) = 1.0$
- $\text{conf}(\{\text{Manteiga}, \text{Pão}\} \Rightarrow \{\text{Café}\}) = 0.75$
- $\text{conf}(\{\text{Café}\} \Rightarrow \{\text{Pão}, \text{Manteiga}\}) = 1.0$
- ...



Regras de Associação



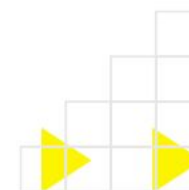
- Exemplo: Calculando confiança de regras de associação

T	Leite	Café	Cerveja	Pão	Manteiga	Arroz	Feijão
#T ₁	0	1	0	1	1	0	0
#T ₂	1	0	1	1	1	0	0
#T ₃	0	1	0	1	1	0	0
#T ₄	1	1	0	1	1	0	0
#T ₅	0	0	1	0	0	0	0
#T ₆	0	0	0	0	1	0	0
#T ₇	0	0	0	1	0	0	0
#T ₈	0	0	0	0	0	0	1
#T ₉	0	0	0	0	0	1	1
#T ₁₀	0	0	0	0	0	1	0

Vamos escolher uma
Confiança Mínima = 0.8

Vamos calcular a confiança de regras geradas por {Café,Pão,Manteiga}:

- $\text{conf}(\{\text{Café}, \text{Pão}\} \Rightarrow \{\text{Manteiga}\}) = 1.0$
- ~~$\text{conf}(\{\text{Manteiga}, \text{Pão}\} \Rightarrow \{\text{Café}\}) = 0.75$~~
- $\text{conf}(\{\text{Café}\} \Rightarrow \{\text{Pão}, \text{Manteiga}\}) = 1.0$
- ...

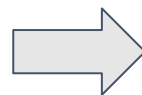


Regras de Associação

- Minerar regras de associação

Etapa 1

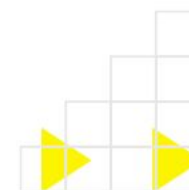
Encontrar *itemsets* frequentes de acordo com um **suporte mínimo**



Etapa 2

Gerar regras de associação usando *itemsets* frequentes conforme **confiança mínima**

Valor alto para suporte mínimo reduz a quantidade *itemsets* frequentes e regras de associação, MAS pode gerar regras de associação muito óbvias e não interessantes



Regras de Associação

- Minerar regras de associação

Etapa 1

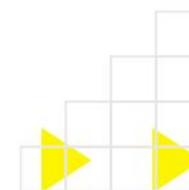
Encontrar *itemsets* frequentes de acordo com um **suporte mínimo**



Etapa 2

Gerar regras de associação usando *itemsets* frequentes conforme **confiança mínima**

Valores baixos para suporte e confiança mínima geram muitas regras de associação, mas quais são potencialmente mais interessantes?



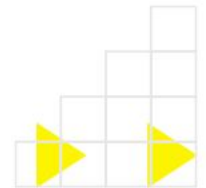
Regras de Associação

- Limitações da medida de confiança
 - Não mede dependência entre *itemsets*
- Outras medidas objetivas:
 - Lift [1]
 - Novidade (*leverage ou PS*) [2]
 - Convicção (*conviction*) [3]

[1] S. Brin, R. Motwani, J. D. Ullman e S. Tsur, Dynamic Itemset Counting and Implication Rules for Market Basket Data. In Proc. of the ACM SIGMOD Intl. Conf. on Management of Data, Arizona, Estados Unidos, 1997, 255–264.

[2] G. Piatetsky-Shapiro, Discovery, Analysis and Presentation of Strong Rules. Knowledge Discovery in Databases, AAAI/MIT Press, 1991, 229– 248

[3] Sergey Brin, Rajeev Motwani, Jeffrey D. Ullman, and Shalom Turk. Dynamic itemset counting and implication rules for market basket data. In SIGMOD 1997, Proceedings ACM SIGMOD International Conference on Management of Data, 255-264



Regras de Associação



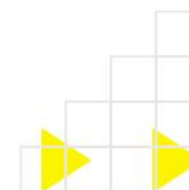
- *Lift*

- Avalia a força da dependência entre *itemsets* da regra
- Medida simétrica

$$lift(A \Rightarrow B) = \frac{conf(A \Rightarrow B)}{sup(B)}$$

- Interpretação da medida:

- $lift(A \Rightarrow B) = 1$: A e B são independentes
- $lift(A \Rightarrow B) < 1$: A e B são negativamente dependentes
- $lift(A \Rightarrow B) > 1$: A e B são positivamente dependentes
- Intervalo: $[0, \infty[$
- Quanto maior o *lift*, mais interessante é a regra de associação



Regras de Associação



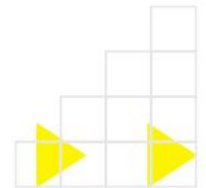
- *Novidade (leverage ou PS)*

- Avalia a força da dependência entre *itemsets* da regra
- Medida simétrica

$$\text{leverage}(A \Rightarrow B) = \text{sup}(A \cup B) - \text{sup}(A) \times \text{sup}(B)$$

- Interpretação da medida:

- $\text{leverage}(A \Rightarrow B) = 0$: A e B são independentes
- $\text{leverage}(A \Rightarrow B) < 0$: A e B são negativamente dependentes
- $\text{leverage}(A \Rightarrow B) > 0$: A e B são positivamente dependentes



Regras de Associação



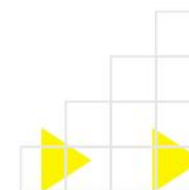
- *Convicção (conviction)*

- Avalia a força da dependência entre *itemsets* da regra
- Medida assimétrica

$$conviction(A \Rightarrow B) = \frac{1 - sup(B)}{1 - conf(A \Rightarrow B)}$$

- Interpretação da medida:

- Valores próximos de 1 indicam independência entre A e B
- Regras com alta confiança (próximos de 1) possuem valores de *conviction* tendendo ao infinito
- Intervalo: $[0, \infty[$



Regras de Associação

- Dezenas de medidas objetivas para regras de associação
- Algoritmos alternativos ao APRIORI

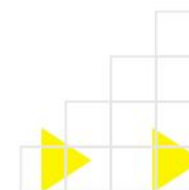
- *FP-growth*

Utiliza uma estrutura de árvore e técnica de divisão e conquista para reduzir custo computacional

- Análise de Regras de Associação

- Capítulo 5 do livro:

Tan, P.N.; Steinbach, M.; Karpatne, A.; Kumar, V. (2016).
Introduction to Data Mining (2nd Edition). Pearson.



Bibliografia

Rezende, S. O. (2003). *Sistemas inteligentes: fundamentos e aplicações*. Editora Manole Ltda.

Tan, P.N.; Steinbach, M.; Karpatne, A.; Kumar, V. (2016). *Introduction to Data Mining (2nd Edition)*. Pearson.

