

# Regras de Associação

# Descoberta de Regras de Associação

- Definição original: tipo especial de dados, chamado “basket data” (dados de cesta)[Agrawal et al 96]
- Cada registro corresponde a uma transação de um cliente, com itens assumindo valores binários (sim/não), indicando se o cliente comprou ou não o respectivo item.

# Exemplo: [Freitas & Lavington 98]

	leite	café	cerveja	pão	manteiga	arroz	feijão
1	não	sim	não	sim	sim	não	não
2	sim	não	sim	sim	sim	não	não
3	não	sim	não	sim	sim	não	não
4	sim	sim	não	sim	sim	não	não
5	não	não	sim	não	não	não	não
6	não	não	não	não	sim	não	não
7	não	não	não	sim	não	não	não
8	não	não	não	não	não	não	sim
9	não	não	não	não	não	sim	sim
10	não	não	não	não	não	sim	não

# Descoberta de Regras de Associação

- Uma regra de associação é um relacionamento
  - SE (X) ENTÃO (Y), onde X e Y são conjuntos de itens, com interseção vazia.
- A cada regra são atribuídos 2 fatores:
  - Suporte (Sup.) =  $\frac{\text{No. de registros com X e Y}}{\text{No. Total de registros}}$
  - Confiança (Conf.) =  $\frac{\text{No. de registros com X e Y}}{\text{No. de registros com X}}$
- Tarefa: descobrir todas as regras de associação com um mínimo Sup e um mínimo Conf.

Sup. = No. de registros com X e Y /No. Total de registros,  
Conf = No. de registros com X e Y/ No. de registros com X

- Conjunto de Items Frequente: café, pão. Sup. = 0,3
- Regra: SE (café) ENTÃO (pão). Conf. = 1
- Conjunto de Items Frequente: café, manteiga. Sup. = 0,3
- Regra: SE (café) ENTÃO (manteiga). Conf. = 1
- Conjunto de Items Frequente: pão, manteiga. Sup = 0,4
- Regra: SE (pão) ENTÃO (manteiga). Conf. = 0,8

Sup. = No. de registros com X e Y / No. Total de registros,  
Conf = No. de registros com X e Y / No. de registros com X

- Regra: SE (manteiga) ENTÃO (pão). Conf. = 0,8
- Conjunto de Items Frequente: café,pão,manteiga  
Sup.=0,3
- Regra: SE (café E pão) ENTÃO (manteiga). Conf.= 1
- Regra: SE (café E manteiga) ENTÃO (pão). Conf.= 1
- Regra: SE (café) ENTÃO (manteiga E pão). Conf.= 1

# Descobrendo regras de associação

- Algoritmo tem 2 fases.
- Fase I: Descobrir conjuntos de itens frequentes. Descobrir todos os conjuntos de itens com suporte maior ou igual ao mínimo suporte especificado pelo usuário.
- Fase II: Descobrir regras com alto fator de confiança. A partir dos conjuntos de itens frequentes, descobrir regras de associação com fator de confiança maior ou igual ao especificado pelo usuário.

# Calculando o suporte de conjuntos de itens

- Suporte = No. de transações contendo o conjunto de itens, dividido pelo No. total de transações.
- Fase I: Passo 1: Calcular suporte de conjuntos com 1 item.
  - leite:  $\text{Sup} = 0,2$ ; café:  $\text{Sup} = 0,3$ ; cerveja:  $\text{Sup} = 0,2$ ; pão:  $\text{Sup} = 0,5$ ; manteiga:  $\text{Sup} = 0,5$ ; arroz:  $\text{Sup} = 0,2$ ; feijão:  $\text{Sup} = 0,2$ ;
  - Itens frequentes ( $\text{Sup} \geq 0,3$ ): café, pão, manteiga

# Calcular suporte de conjuntos com 2 itens

- Passo 2: Calcular suporte de conjuntos com 2 itens
- Otimização: Se um item  $I$  não é frequente, um conjunto com 2 itens, um dos quais é o item  $I$ , não pode ser frequente. Logo, conjuntos contendo item  $I$  podem ser ignorados.
  - Conjunto de itens: café, pão.  $\text{Sup} = 0,3$ .
  - Conjunto de itens: café, manteiga.  $\text{Sup} = 0,3$ .
  - Conjunto de itens: manteiga, pão.  $\text{Sup} = 0,4$ .
  - Conjuntos de itens frequentes ( $\text{Sup} \geq 0,3$ ):
- {café, pão}, {café, manteiga}, {manteiga, pão}

# Calcular suporte de conjuntos com 3 itens.

- Passo 3: Calcular suporte de conjuntos com 3 itens.
  - Otimização: Se o conjunto de itens  $\{I, J\}$  não é frequente, um conjunto com 3 itens incluindo os itens  $\{I, J\}$  não pode ser frequente. Logo, conjuntos contendo itens  $\{I, J\}$  podem ser ignorados.
  - Conjunto de itens: café, pão, manteiga.  $\text{Sup} = 0,3$ .
  - Conjuntos de itens frequentes ( $\text{Sup} \geq 0,3$ ): {café, pão, manteiga}.

# Fator de confiança de regras

- Calculando fator de confiança de regras candidatas, geradas a partir de conjuntos de itens frequentes.
  - Conf. da regra "SE X ENTÃO Y" é No. de transações contendo X e Y dividido pelo No. de transações com X.
- Conjunto de itens: {café, pão}.
  - SE café ENTÃO pão. Conf = 1,0.
  - SE pão ENTÃO café. Conf = 0,6.

- Conjunto de itens: {café, manteiga}.
  - SE café ENTÃO manteiga. Conf = 1,0.
  - SE manteiga ENTÃO café. Conf = 0,6.
- Conjunto de itens: {manteiga, pão}.
  - SE manteiga ENTÃO pão. Conf = 0,8.
  - SE pão ENTÃO manteiga. Conf = 0,8.

# Confiança de regras

- Conjunto de itens: {café, manteiga, pão}.
  - SE café, pão ENTÃO manteiga. Conf = 1,0.
  - SE café, manteiga ENTÃO pão. Conf = 1,0.
  - SE manteiga, pão ENTÃO café. Conf = 0,75.
  - SE café ENTÃO pão, manteiga. Conf = 1,0.
  - SE pão ENTÃO café, manteiga. Conf = 0,6.
  - SE manteiga ENTÃO café, pão. Conf = 0,6.

# Confiança de regras

- Finalmente, selecionam-se regras com Conf. maior ou igual ao valor mínimo especificado pelo usuário (ex. 0,8).