

Regras

# Regras Rudimentarias

- 1R: aprende uma regra por atributo
  - atribuí a classe mais freqüente
  - taxa de erro: proporção de instâncias que não pertence a classe majoritária
  - escolhe o atributo com menor taxa de erro

# Pseudo-código para 1R

- Para cada atributo
  - para cada valor do atributo, faça uma regra como:
    - conte a frequência de cada classe
    - encontre a classe mais freqüente
    - atribua a classe mais freqüente a esta regra
    - calcule a taxa de erro da regra
  - escolha a regra com taxa de erro menor

# Avaliando os atributos da base Weather

Outlook	Temp	Humidity	Windy	Play	Attribute	Rules	Errors	Total errors
Sunny	Hot	High	False	No	Outlook	Sunny → No	2/5	4/14
Sunny	Hot	High	True	No		Overcast → Yes	0/4	
Overcast	Hot	High	False	Yes	Temp	Rainy → Yes	2/5	5/14
Rainy	Mild	High	False	Yes		Hot → No*	2/4	
Rainy	Cool	Normal	False	Yes		Mild → Yes	2/6	
Rainy	Cool	Normal	True	No		Cool → Yes	1/4	
Overcast	Cool	Normal	True	Yes	Humidity	High → No	3/7	4/14
Sunny	Mild	High	False	No		Normal → Yes	1/7	
Sunny	Cool	Normal	False	Yes	Windy	False → Yes	2/8	5/14
Rainy	Mild	Normal	False	Yes		True → No*	3/6	
Sunny	Mild	Normal	True	Yes				
Overcast	Mild	High	True	Yes				
Overcast	Hot	Normal	False	Yes				
Rainy	Mild	High	True	No				

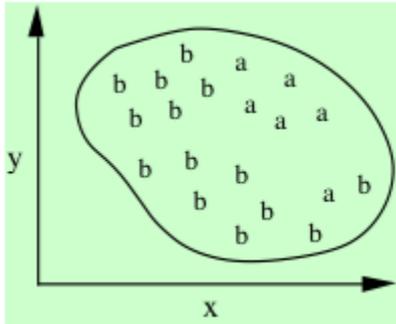
# Discussão

Apesar de simples apresenta resultados similares a arvores de decisão!!!!

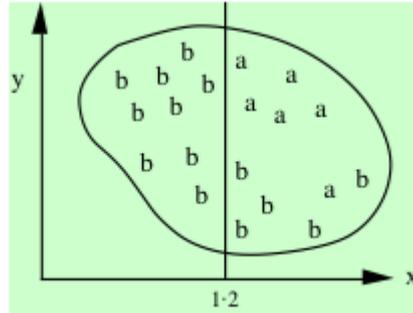
# Algoritmos 'Covering'

- Para cada classe são geradas um conjunto de regras.
- Abordagem 'Covering'
  - a cada passo são identificadas regras que cobrem algumas instâncias

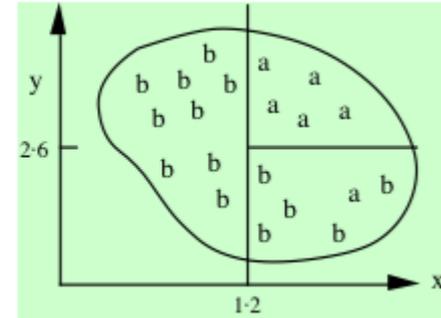
# Exemplo



Se True então a



Se  $x > 1.2$  então a



Se  $x > 1.2$  e  $y > 2.6$   
então a

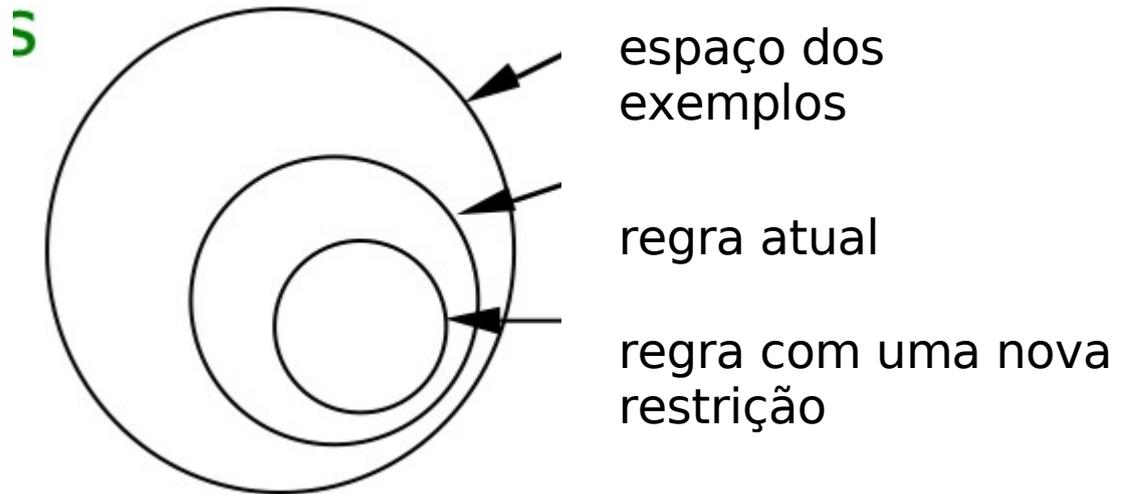
Possível conjunto de regras para a classe b

Se  $x < 1.2$  então classe b

se  $x > 1.2$  and  $y < 2.6$  então classe b

# Algorithm 'Covering'

- Gere uma regra adicionando restrições que maximizem a precisão da regra.
- Similar a situação das árvores



# Selecionando a Restrição

- Objetivo: Melhorar a precisão
  - t numero total de instâncias cobertas pela regra
  - p exemplos positivos da classe cobertos pela regra
  - t-p numero de erros cometidos
  - Maximizar  $p/t$  ??

# Exemplo: Receita de óculos ???

Age	Spectacle prescription	Astigmatism	Tear production rate	Recommended lenses
Young	Myope	No	Reduced	None
Young	Myope	No	Normal	Soft
Young	Myope	Yes	Reduced	None
Young	Myope	Yes	Normal	Hard
Young	Hypermetrope	No	Reduced	None
Young	Hypermetrope	No	Normal	Soft
Young	Hypermetrope	Yes	Reduced	None
Young	Hypermetrope	Yes	Normal	hard
Pre-presbyopic	Myope	No	Reduced	None
Pre-presbyopic	Myope	No	Normal	Soft
Pre-presbyopic	Myope	Yes	Reduced	None
Pre-presbyopic	Myope	Yes	Normal	Hard
Pre-presbyopic	Hypermetrope	No	Reduced	None
Pre-presbyopic	Hypermetrope	No	Normal	Soft
Pre-presbyopic	Hypermetrope	Yes	Reduced	None
Pre-presbyopic	Hypermetrope	Yes	Normal	None
Presbyopic	Myope	No	Reduced	None
Presbyopic	Myope	No	Normal	None
Presbyopic	Myope	Yes	Reduced	None
Presbyopic	Myope	Yes	Normal	Hard
Presbyopic	Hypermetrope	No	Reduced	None
Presbyopic	Hypermetrope	No	Normal	Soft
Presbyopic	Hypermetrope	Yes	Reduced	None
Presbyopic	Hypermetrope	Yes	Normal	None

# Primeira Restrição

```
If ?  
    then recommendation = hard
```

Age = Young	2/8
Age = Pre-presbyopic	1/8
Age = Presbyopic	1/8
Spectacle prescription = Myope	3/12
Spectacle prescription = Hypermetrope	1/12
Astigmatism = no	0/12
Astigmatism = yes	4/12
Tear production rate = Reduced	0/12
Tear production rate = Normal	4/12

## Regra

```
If astigmatism = yes  
then recommendation = hard
```

## Instâncias cobertas pela regra

Age	Spectacle prescription	Astigmatism	Tear production rate	Recommended lenses
Young	Myope	Yes	Reduced	None
Young	Myope	Yes	Normal	Hard
Young	Hypermetrope	Yes	Reduced	None
Young	Hypermetrope	Yes	Normal	hard
Pre-presbyopic	Myope	Yes	Reduced	None
Pre-presbyopic	Myope	Yes	Normal	Hard
Pre-presbyopic	Hypermetrope	Yes	Reduced	None
Pre-presbyopic	Hypermetrope	Yes	Normal	None
Presbyopic	Myope	Yes	Reduced	None
Presbyopic	Myope	Yes	Normal	Hard
Presbyopic	Hypermetrope	Yes	Reduced	None
Presbyopic	Hypermetrope	Yes	Normal	None

# Refinando

estado Atual

```
If astigmatism = yes  
and ?  
then recommendation = hard
```

possíveis atributos

Age = Young	2/4
Age = Pre-presbyopic	1/4
Age = Presbyopic	1/4
Spectacle prescription = Myope	3/6
Spectacle prescription = Hypermetrope	1/6
Tear production rate = Reduced	0/6
Tear production rate = Normal	4/6

# Regra Modificada e dados

regra com o melhor atributo adicionado

```
If astigmatism = yes  
    and tear production rate = normal  
then recommendation = hard
```

Instâncias cobertas pela regra modificada

Age	Spectacle prescription	Astigmatism	Tear production rate	Recommended lenses
Young	Myope	Yes	Normal	Hard
Young	Hypermetrope	Yes	Normal	hard
Pre-presbyopic	Myope	Yes	Normal	Hard
Presbyopic	Hypermetrope	Yes	Normal	None
Presbyopic	Myope	Yes	Normal	Hard
Presbyopic	Hypermetrope	Yes	Normal	None

# Continuando

Estado atual

```
If astigmatism = yes
    and tear production rate = normal
    and ?
then recommendation = hard
```

Possíveis testes

Age = Young	2/2
Age = Pre-presbyopic	1/2
Age = Presbyopic	1/2
Spectacle prescription = Myope	3/3
Spectacle prescription = Hypermetrope	1/3

# Resultado

- Regra final

```
If astigmatism = yes  
and tear production rate = normal  
and spectacle prescription = myope  
then recommendation = hard
```

- Segunda regra (construída a partir dos exemplos que não são cobertos pela regra anterior)

```
If age = young and astigmatism = yes  
and tear production rate = normal  
then recommendation = hard
```

# 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.) =  $\frac{\text{No. de registros com X e Y}}{\text{Total de registros}}$
  - 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: Sup = 0,2; café: Sup = 0,3; cerveja: Sup = 0,2; pão: Sup = 0,5; manteiga: Sup = 0,5; arroz: Sup = 0,2; feijão: Sup = 0,2;
  - Itens frequentes (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 ( $\text{Sup} \geq 0,3$ ):  $\{\text{café, pão, manteiga}\}$ .

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