Tarefas de Descoberta de Conhecimento em Bases de Dados (DCBD)

Profa. Leticia T. M. Zoby

(leticia.zoby@udf.edu.br)

Porque DCBD? Ponto de vista comercial

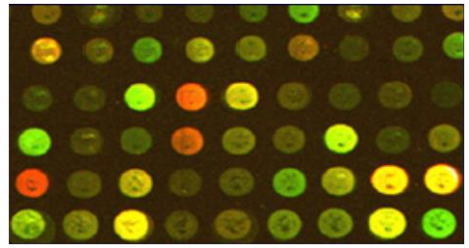
Descoberta de Conhecimento em Bases de Dados

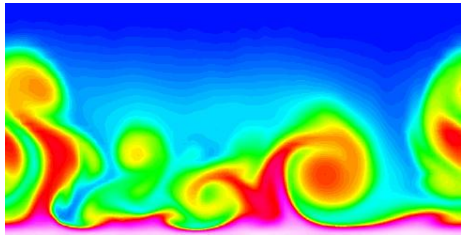
- Enormes quantidades de dados são coletadas e armazenadas
  - Dados da Web, e-commerce
  - Compras em supermercados, lojas de
    - departamentos, entre outros
  - Transações bancárias e de cartões
    - de crédito
- Os computadores se tornaram baratos e mais poderosos
- A pressão competitiva é muito forte

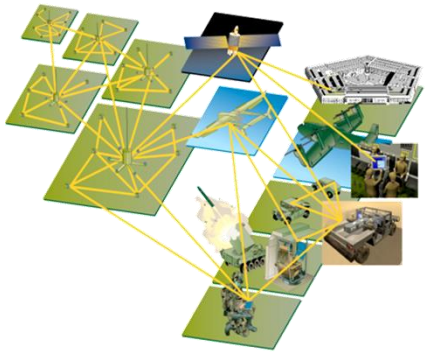
Porque DCBD? Ponto de vista comercial

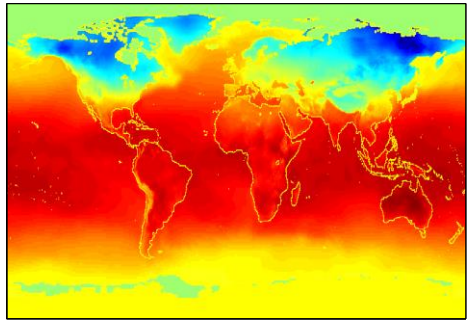
- Dados captados e armazenados em grande velocidade (GB/hora)
  - sensores remotos em satélites
  - telescópios varrendo o firmamento
  - microarrays gerando dados de expressão gênica
  - simulações científicas gerando terabytes de dados

- Técnicas tradicionais de análise são inviáveis para estes dados brutos



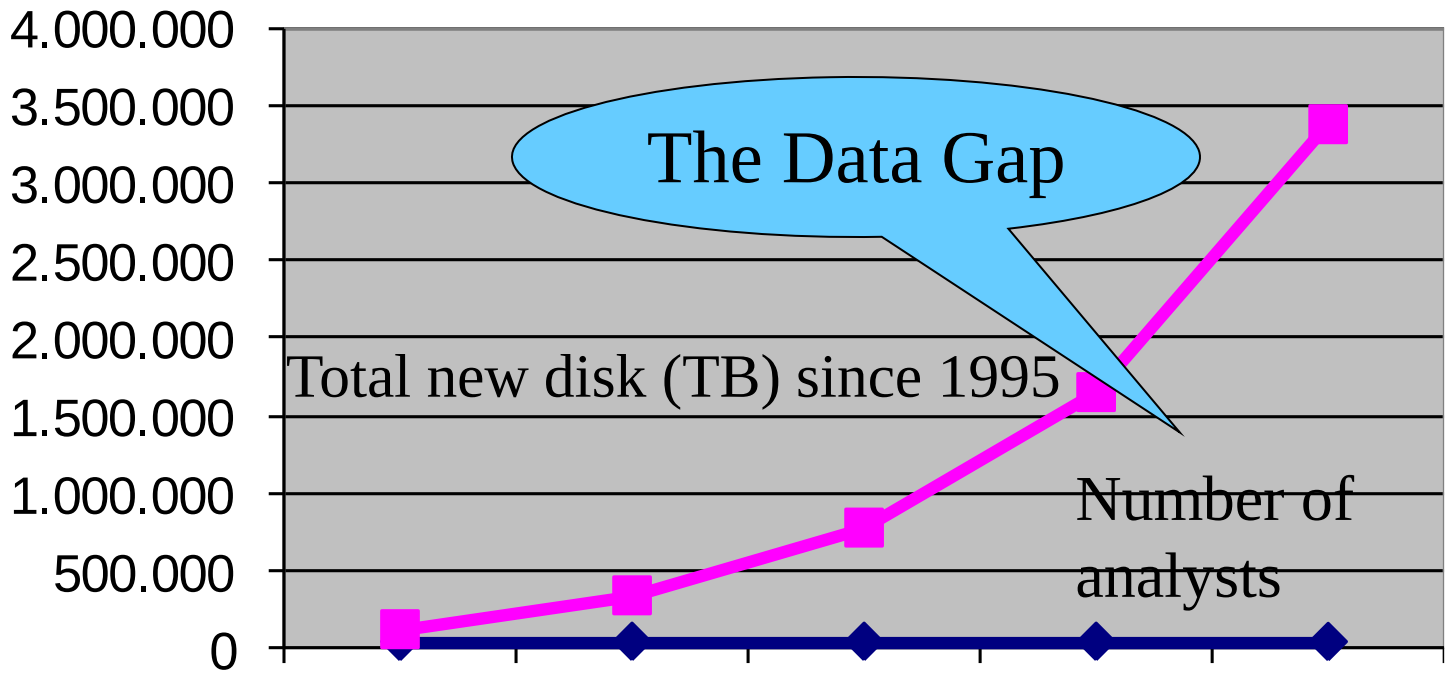






Motivação para minerar grandes bases de dados

- Frequentemente há informação "escondida" nos dados, que não é evidente
- Analistas humanos podem levar semanas para descobrir informação útil
- Muitos dados *nunca* são analisados



From: R. Grossman, C. Kamath, V. Kumar, "Data Mining for Scientific and Engineering Applications", Kluwer, 2001

Observando e Aprendendo

Exemplo: um proprietário de uma pequena loja de vinhos conhece tudo sobre vinhos, por exemplo, o tipo de uva, a região onde a uva foi cultivada, o clima, o solo, a altitude dos parreirais, aroma, sabor, cor, o processo de fabricação. Os clientes gostam de visitar sua loja pois, também, aprendem muito sobre vinhos. Porém, só isto não basta, o proprietário precisa conhecê-los, como por exemplo, qual o tipo de vinho que o cliente gosta? Qual o poder aquisitivo? Assim, ele poderá dar um atendimento diferenciado (um a um) aos clientes. Temos, portanto, duas necessidades:

conhecimento e aprendizado

Uma pequena loja $\Rightarrow$ poucos clientes $\Rightarrow$ atendimento personalizado

Uma grande empresa $\Rightarrow$ milhares de clientes $\Rightarrow$ dificuldade em dar um atendimento dedicado

Observando e Aprendendo

Qual a tendência nos dias atuais?

Ter clientes leais, através de um relacionamento pessoal, *um-para-um*, entre a empresa e o cliente

Dentro desta tendência, as empresas desejam identificar os clientes cujos valores e necessidades sejam compatíveis com o uso prolongado de seus produtos, e nos quais é válido o risco de investir em promoções com descontos, pacotes, brindes e outras formas de criar essa relação pessoal.

Esta mudança de foco requer mudanças em toda a empresa, mas principalmente nos setores de marketing, vendas e atendimento ao cliente.

Tipos de descobertas (tarefas de DCBD)

- Os dois principais objetivos de alto nível da DCBD são a descrição e a predição
  - a descrição se concentra em encontrar padrões que descrevem os dados de forma compreensível para o usuário;
  - a predição envolve usar valores conhecidos de campos ou variáveis para predizer o valor desconhecido ou futuro de variáveis de interesse.

Para isso, utiliza-se vários tipos de descoberta ou tarefas de DCBD

Principais Tarefas de DCBD

- Descoberta de regras de associação [Descritiva]
- Descoberta de padrões sequenciais [Descritiva]
- Classificação [Preditiva]
- Regressão [Preditiva]
- Clustering [Descritiva]
- Detecção de desvios [Preditiva]

Associação

Descoberta de regras de associação

- Também denominada de Descoberta de Associação ou Descoberta de Regras Associativas
- “Consiste em encontrar conjuntos de itens que ocorram simultaneamente de forma frequente em um banco de dados.”

Descoberta de regras de associação

- Exemplos:
  - determinados procedimentos médicos aparecem sempre juntos
  - determinados procedimento médicos aparecem mais associados a homens e outros a mulheres
  - Estratégias de vendas:
    - Promoções entre produtos
    - Rearranjo da disposição dos produtos em prateleiras e gôndolas

Descoberta de regras de associação

- Exemplos:
  - Sei que quem compra o produto A também compra o B.

Descoberta de regras de associação

- "Encontrar produtos que sejam frequentemente vendidos de forma conjunta."

Tabela 1 – Relação das vendas de um minimercado em um período

Transação	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

Tabela 2 –Formato Cesta da relação das venas da Tabela 1.

Transação	Item
1	Café
1	Pão
1	Manteiga
2	Leite
2	Cerveja
2	Pão
2	Manteiga
3	Café
3	Pão
3	Manteiga
4	Café
4	Leite
4	Pão
4	Manteiga
5	Cerveja
6	Manteiga
7	Pão
8	Feijão
9	Arroz
9	Feijão
10	Arroz

Descoberta de regras de associação

- Regras de Associação – Algumas Definições:

Def: **Transação:** Elemento de ligação existente em cada ocorrência de itens no conjunto de dados.

Def: **Regra de Associação:** $X \rightarrow Y$, onde X e Y são **itemsets** (conjuntos de itens) tais que $X \cap Y = \varnothing$.

Def: **K-Itemset** é um itemset contendo exatamente k itens

Descoberta de regras de associação

- Regras de Associação – Formalização:

  "Consiste em encontrar **regras de associação frequentes e válidas** em um conjunto de dados, a partir da especificação dos parâmetros de suporte e confiança mínimos."

- Exemplo de Regras de Associação:

  {Leite} → {Pão}

  {Pão, Manteiga} → {Café}

Descoberta de regras de associação

- Estrutura comum: Métricas de avaliação das regras

- **Suporte mínimo**
  - Fração das transações que contêm X e Y
  - Identificação dos conjuntos de itens frequentes:

$$|X \cup Y| / |D| >= \text{MinSup}$$ (**Suporte Mínimo**)

Maior custo computacional

- **Confiança (c)**
  - Mede a freqüência com que Y aparece nas transações que contêm X
  - Identificação, dentre os conjuntos de itens frequentes, quais as regras válidas:

$$|X \cup Y| / |X| >= \text{MinConf}$$ (**Confiança Mínima**)

Descoberta de regras de associação

- Exemplo: Considere o seguinte Conjunto de Dados (Tabela 1)

Transação	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

- Exemplo:
  - Algumas Regras Descobertas:

    - Regra: SE (café) ENTÃO (pão).
    - Regra: SE (café) ENTÃO (manteiga).
    - Regra: SE (pão) ENTÃO (manteiga).
    - Regra: SE (manteiga) ENTÃO (pão).
    - Regra: SE (café E pão) ENTÃO (manteiga).
    - Regra: SE (café E manteiga) ENTÃO (pão).
    - Regra: SE (café) ENTÃO (manteiga E pão).

Descoberta de regras de associação

- Exemplo: Regras de Associação – Como obtê-las?

Fase I: Definir os valores de suporte e confiança mínimos:

MinSup = 0,3

MinConf = 0,8

Descoberta de regras de associação

- Exemplo: Regras de Associação – Como obtê-las?

Fase II: Identificar os conjuntos de itens frequentes:

1ª Iteração:

1 - Itemsets	Suportes
Leite	0,2
Café	0,3
Cerveja	0,2
Pão	0,5
Manteiga	0,5
Arroz	0,2
Feijão	0,2

Descoberta de regras de associação

- Exemplo: Regras de Associação – Como obtê-las?

Fase II: Identificar os conjuntos de itens frequentes:

1ª Iteração:

1 - Itemsets	Suportes
Leite	0,2
Café	**0,3**
Cerveja	0,2
Pão	**0,5**
Manteiga	**0,5**
Arroz	0,2
Feijão	0,2

Descoberta de regras de associação

- Exemplo: Regras de Associação – Como obtê-las?

Fase II: Identificar os conjuntos de itens frequentes:

2ª Iteração: Combinar os 1-itemsets identificados anteriormente

2 - Itemsets	Suportes
Café , Pão	0,3
Café , Manteiga	0,3
Pão , Manteiga	0,4

Descoberta de regras de associação

- Exemplo: Regras de Associação – Como obtê-las?

Fase II: Identificar os conjuntos de itens frequentes:

2ª Iteração: Combinar os 1-itemsets identificados anteriormente

2 - Itemsets	Suportes
Café , Pão	**0,3**
Café , Manteiga	**0,3**
Pão , Manteiga	**0,4**

Descoberta de regras de associação

- Exemplo: Regras de Associação – Como obtê-las?

Fase II: Identificar os conjuntos de itens frequentes:

3ª Iteração: Combinar os 2-itemsets identificados anteriormente

3 - Itemsets	Suportes
Café , Pão , Manteiga	0,3

Descoberta de regras de associação

- Exemplo: Regras de Associação – Como obtê-las?

Fase II: Identificar os conjuntos de itens frequentes:

3ª Iteração: Combinar os 2-itemsets identificados anteriormente

3 - Itemsets	Suportes
Café , Pão , Manteiga	**0,3**

Descoberta de regras de associação

- Exemplo: Regras de Associação – Como obtê-las?

Fase II: Identificar os conjuntos de itens frequentes:

Lista de **todos** os k-itemsets freqüentes obtidos ($K \geq 2$)

- Café e Pão,
- Café e Manteiga,
- Pão e Manteiga,
- Café e Pão e Manteiga

Descoberta de regras de associação

- Exemplo: Regras de Associação – Como obtê-las?

Fase III: Identificação das Regras Válidas:

- **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.

Descoberta de regras de associação

- Exemplo: Regras de Associação – Como obtê-las?

Fase III: Identificação das Regras Válidas:

- **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.

Finalmente, seleciona-se regras com Conf. maior ou igual ao valor mínimo especificado pelo usuário (**MinConf = 0,8**).

Descoberta de regras de associação

- Exemplo: Regras de Associação – Regras Obtidas no Exemplo

SE café ENTÃO pão.
SE café ENTÃO manteiga.
SE manteiga ENTÃO pão.
SE pão ENTÃO manteiga.
SE café,pão ENTÃO manteiga.
SE café, manteiga ENTÃO pão.
SE café ENTÃO pão, manteiga.

Descoberta de regras de associação

- Exemplo de Algoritmos
  - **Apriori**
  - **DHP – Direct Hashing and Pruning**
  - **Partition**
  - **DIC – Dynamic Itemset Counting**

Padrões sequenciais
Descoberta de Sequências

Padrões sequenciais

- Detecção de dependências temporais entre eventos.
- Extensão da Mineração de Associações: aspecto “temporal”.

- Definição:
  - Dado um conjunto de *objetos*, com cada objeto associado com a sua *linha de eventos*, encontre regras com forte dependência sequencial entre diferentes eventos.

(A B) (C) ⟶ (D E)

Padrões sequenciais

- Exemplos:
  - Histórico de acessos a páginas de um site pelos usuários da web.
  - Histórico de itens comprados por consumidores ao longo de um período
  - Determinado procedimento médico sempre precede outro
  - Turistas que visitam o museu do Louvre depois visitam a Notre Dame

Padrões sequenciais

- Definições Relevantes:

Def: **Sequência:** Lista ordenada de Itemsets. Caracterizada por objeto, rótulo temporal e eventos. Cada registro armazena ocorrências de eventos sobre um objeto em um instante de tempo particular. Notação: $<s_1 s_2 ... s_n>$, onde s_j é um itemset.

Exemplo:

Consumidores $\Leftrightarrow$ objetos

itens comprados $\Leftrightarrow$ eventos

Def: O itemset s_j é também chamado de **elemento da sequência**. Cada elemento de uma sequência é denotado por $(x_1, x_2, ..., x_m)$, onde x_j é um item ou evento.

Padrões sequenciais

- Definições Relevantes:

Def: Uma sequência $<a_1 a_2 ... a_n>$ é uma **subsequência (ou especialização)** de outra sequência $<b_1 b_2 ... b_n>$ se existirem inteiros $i_1 < i_2 < ... < i_n$ tais que $a_1 \subseteq b_{i1}$, $a_2 \subseteq b_{i2}$, ...e $a_n \subseteq b_{in}$.

Exemplo:

< (3) (4, 5) (8) > é uma subsequência de < (7) (3, 8) (9) (4, 5, 6) (8) >, pois (3) $\subseteq$ (3, 8), (4, 5) $\subseteq$ (4, 5, 6) e (8) $\subseteq$ (8).

No entanto, a sequência < (3) (5) > não é uma subsequência de < (3, 5) > e vice versa.

Padrões sequenciais

- Definições Relevantes:

Def: O **suporte (ou frequência)** de uma sequência α refere-se ao número total de objetos que contêm α.

Def: Dado um limiar definido pelo usuário, denominado **suporte mínimo**, diz-se que uma **sequência** é **frequente** se esta ocorrer mais do que o suporte mínimo.

Def: Uma **k-sequência** é uma sequência com exatamente k elementos.

Padrões sequenciais

- **Algoritmos específicos:**
  - GSP – Generalized Sequential Patterns
  - MSDD – Multi Stream Dependency Detection
  - SPADE – Sequential Pattern Discovery using Equivalence Classes

- **Exemplos:**
  - Em transações de vendas:
    - Lojas calçados femininos: {cinto, bolsa} → {sapato}
    - Loja de artigos esportivos: {tenis} {raquete, bolas} → {moleton}
  - Marketing
  - Reestruturação de web sites

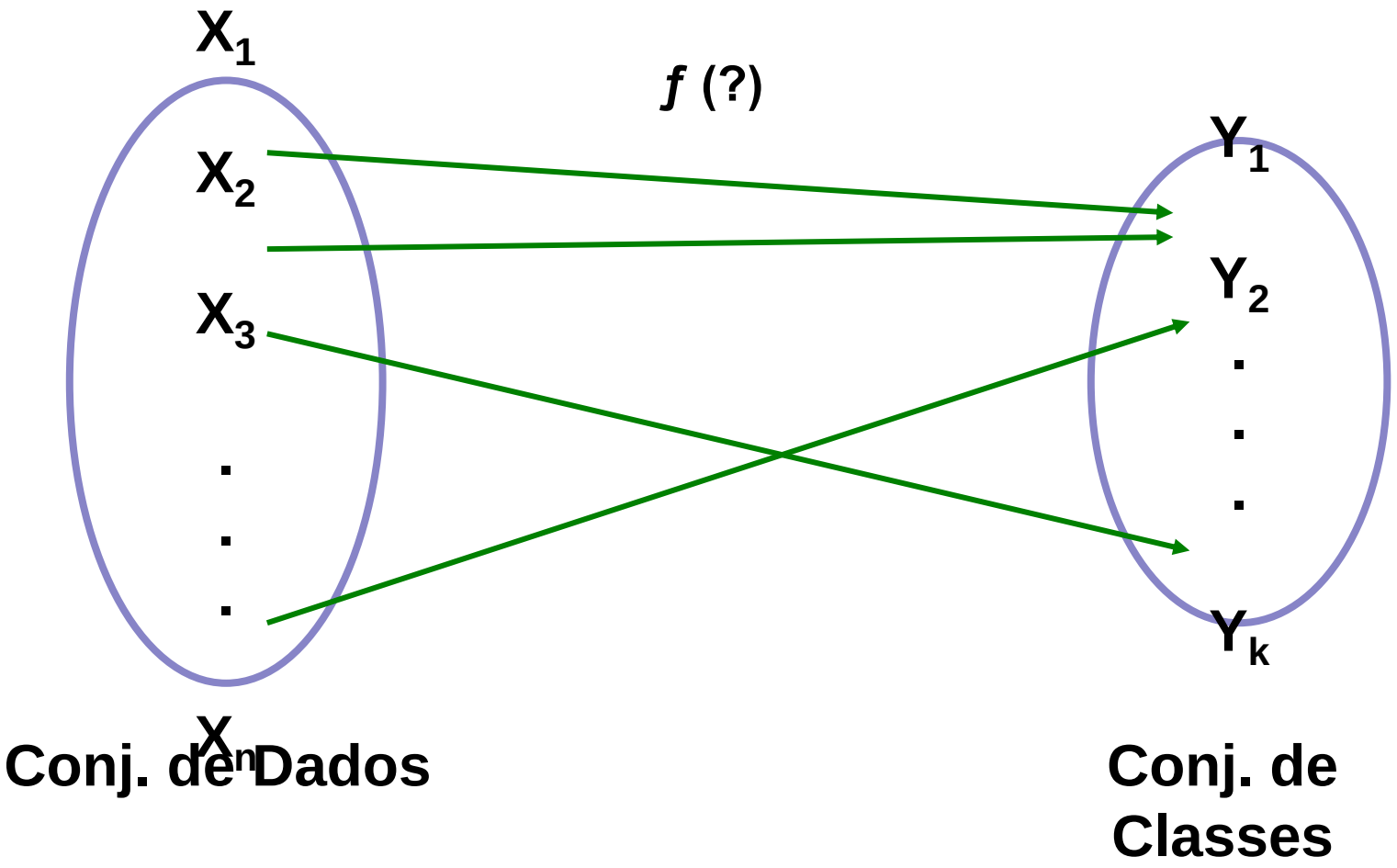
Classificação

Classificação: exemplos

- O professor classifica o desempenho do aluno em A, B, C, D ou E
- Identificar um cogumelo como sendo venenoso ou não
- Reconhecimento de caracteres

Classificação: definição

- Caracterização do Problema:



Classificação: definição

- Nos casos em que a imagem de ***f*** é formada por rótulos de classes, a tarefa de inferência indutiva é denominada classificação e toda hipótese ***h*** chamada de classificador.
- A identificação da função ***h*** consiste de um processo de busca no espaço de hipóteses ***H***, pela função que mais se aproxime da função original ***f***. Este processo é denominado ***aprendizado*** (Russell e Norvig, 1995).
- Todo algoritmo que possa ser utilizado na execução do processo de aprendizado é chamado ***algoritmo de aprendizado***.

Classificação: definição

- O conjunto de todas as hipóteses que podem ser obtidas por um algoritmo de aprendizado $\boldsymbol{L}$ é representado por $\boldsymbol{H_L}$. Cada hipótese pertencente ao $\boldsymbol{H_L}$ é representada por $\boldsymbol{h_L}$.
- Acurácia da hipótese ***h:*** qualidade ou precisão de ***h*** em mapear corretamente cada vetor de entradas ***x*** em ***f(x)***.

$$Acc(h) = 1 - Err(h)$$

$$Err(h) = \frac{1}{n}\sum_{i=1}^{n} \| y_i \neq h(i) \|$$

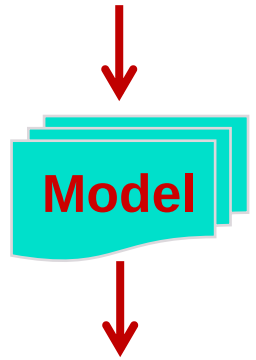
Classificação: definição

- Dada uma coleção de registros (*conjunto de treinamento*)
  - Cada registro contém um conjunto de *atributos*, e um dos atributos é a *classe*.

- Encontre um *modelo* para o atributo classe como uma função dos valores dos outros atributos

- Objetivo: definir a classe para novos registros tão acuradamente quanto possível.

Id	Casa própria	EstCivil	Rendim.	Mau Pagador
1	S	Solteiro	125K	NÃO
2	N	Casado	100K	NÃO
3	N	Solteiro	70K	NÃO
4	S	Casado	120K	NÃO
5	N	Divorc.	95K	SIM
6	N	Casado	60K	NÃO
7	S	Divorc.	220K	NÃO
8	N	Solteiro	85K	SIM



Casa própria	EstCivil	Rendim.	Mau Pagador
N	Solteiro	75K	?
S	Casado	50K	?
N	Casado	150K	?
S	Divorciado	90K	?

Classificação

- Exemplos de técnicas tradicionais:

  - REDES NEURAIS → BACKPROPAGATION
  - ÁRVORES DE DECISÃO → ID3, C4.5
  - ALGORITMOS GENÉTICOS → RULE EVOLVER
  - ESTATÍSTICA → CLASSIFICADORES BAYESIANOS
  - BASEADAS EM INSTÂNCIA → K-NN

Classificação

- Exemplo de aplicação

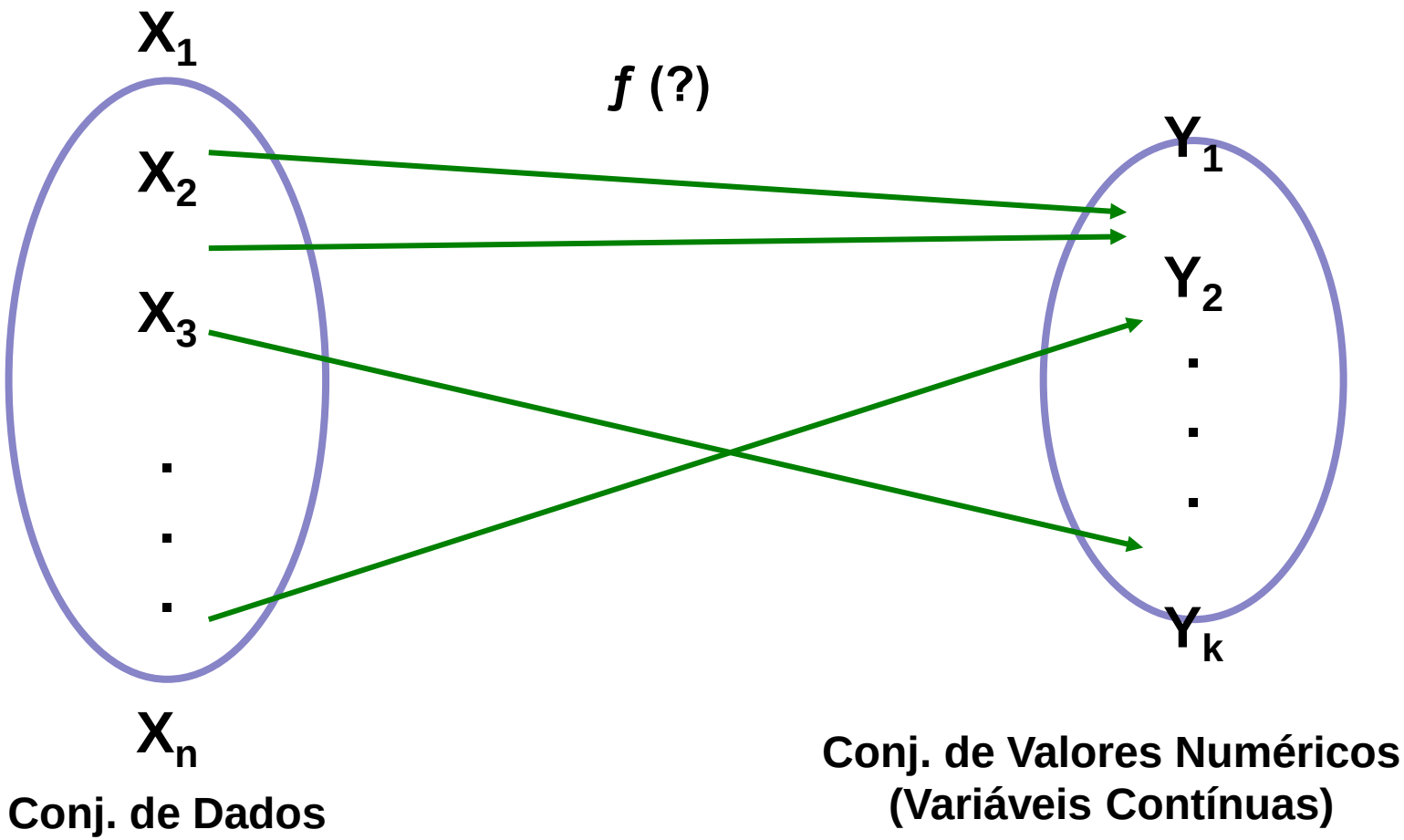
Sexo	País	Idade	Comprar
M	França	25	Sim
M	Inglaterra	21	Sim
F	França	23	Sim
F	Inglaterra	34	Sim
F	França	30	Não
M	Alemanha	21	Não
M	Alemanha	20	Não
F	Alemanha	18	Não
F	França	34	Não
M	França	55	Não

- Algumas Regras:
  - **Se (País = Alemanha) Então Comprar = Não**
  - **Se (País = Inglaterra) Então Comprar = Sim**
  - **Se (País = França e Idade ≤ 25) Então Comprar = Sim**
  - **Se (País = França e Idade > 25) Então Comprar = Não**

Regressão

Regressão: definição

- Caracterização do Problema (análogo à Classificação):



Regressão

- Prediz o valor de uma variável contínua baseado no valor de outras variáveis, considerando um modelo de dependência linear ou não linear.
- Bastante estudado em estatística e redes neurais
- Exemplos:
  - Previsão da quantidade de vendas de um novo produto baseado nos gastos com propaganda
  - Previsão da velocidade do vento em função da temperatura, humidade, pressão atmosférica, etc.
  - Previsão da evolução do índice de bolsa de valores.

Regressão

- Regressão Linear: Formalização

Em sua forma mais simples: Regressão Linear Bivariada

Possui duas variáveis:

- X → variável independente
- Y → variável dependente (função linear da variável X)

Objetivo: Definir valores adequados para os parâmetros α e β (coeficientes de regressão linear) da função:

$$Y = \alpha + \beta X$$

Regressão

- Regressão Linear: Formalização

Objetivo da Regressão Linear Bivariada: Definir valores adequados para os parâmetros α e β (coeficientes de regressão linear) da função:

$$Y = \alpha + \beta X$$

Ex. de algoritmo: Método dos Mínimos Quadrados (MMQ)

MMQ busca minimizar o erro entre os dados reais e os dados estimados pela função.

Regressão

- Regressão Linear: Formalização

  Método dos Mínimos Quadrados (MMQ)

  Busca minimizar o erro entre os dados reais e os dados estimados pela função $Y = \alpha + \beta X$

  Sejam *n* amostras dos dados: $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$

  Estimativa dos coeficientes pelo MMQ:

$$\beta = \frac{\sum_{i=1}^{n}(x_i - x')(y_i - y')}{\sum_{i=1}^{n}(x_i - x')^2} \qquad \alpha = y' - \beta x'$$

  x' e *y'* são as médias dos valores dos atributos X e Y

Regressão

- Regressão Linear: Formalização

Método dos Mínimos Quadrados (MMQ)

Exemplo de Aplicação:

Dados dos funcionários de uma empresa fictícia

X (experiência em anos)	Y (salário anual em R$ 1.000)
03	30
08	57
09	64
13	72
03	36
06	43
11	59
21	90
01	20
16	83

$$x' = 9,1 \text{ e } y' = 55,4$$

$$\beta = \frac{(3-9,1)(30-55,4)+(8-9,1)(57-55,4)+\ldots+(16-9,1)(83-55,4)}{(3-9,1)^2+(8-9,1)^2+\ldots+(16-9,1)^2} = 3,7$$

$$\alpha = 55,4-(3,7)(9,1) = 21,7$$

$$Y = 21,7 + 3,7*X$$

Regressão

- Regressão Não-Linear: Formalização

Existem muitos problemas onde os dados não apresentam dependência linear entre si. Nesses casos, podem ser aplicadas técnicas de ***Regressão Não Linear***.

Por exemplo: a ***Regressão Polinomial*** (consiste em adicionar ao modelo linear termos polinomiais com grau maior que 1).

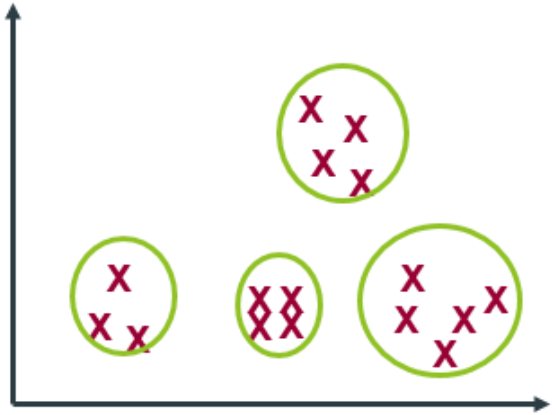
Conversão do modelo não-linear em linear por meio de transformações das variáveis.

Problema linear, aplica-se o MMQ.

Clusterização (Clustering)/ Agrupamentos

Clusterização/ Agrupamento

- Separação dos registros em n “clusters”
- Maximizar/Minimizar similaridade intra/inter cluster



- Def: Cluster: Grupo de registros de um conjunto de dados que compartilham propriedades que os tornam similares entre si.
- Def: Clusterização: Processo de particionamento de uma base de dados em conjuntos em que o objetivo é maximizar a similaridade intra-cluster e minimizar a similaridade inter-cluster.

Clusterização/ Agrupamento

- Dado um conjunto de dados, cada um com um conjunto de atributos, e uma medida de similaridade entre eles, encontre *clusters* (grupos) tais que:
  - Dados de um grupo são mais similares entre si que com dados de outros grupos
  - Dados de grupos diferentes são menos similares entre si.
- Medidas de similaridade:
  - Distância Euclidiana, para atributos contínuos
  - Outras medidas específicas do problema.

Clusterização/ Agrupamento

- Formalização:

  Sejam:

  - n pontos de dados x^1, x^2, ..., x^n tais que cada ponto pertença a um espaço k dimensional R^k
  - d: R^k x $R^k \rightarrow$ R, uma distância entre pontos de R^k

  O processo de Clusterização consiste em encontrar mj pontos (centróides dos clusters), j=1,…,r que minimizem a função

$$\frac{1}{n}\sum_{i=1}^{n}(\min_{j} d^2(X_i, m_j))$$

Clusterização/ Agrupamento

- Estrutura Comum:
- **Inicialização**: Seleção de um conjunto com k centroides de clusters iniciais no espaço de dados. Esta seleção pode ser aleatória ou de acordo com alguma heurística.
- **Cálculo da Distância**: Calcula a distância euclidiana de cada ponto ou padrão ao centroide de cada cluster. Atribui cada ponto ao cluster cuja distância do ponto ao centroide do cluster seja mínima.
- **Recálculo dos Centroides**: Recalcula o centroide de cada cluster pela média dos pontos de dados atribuídos ao respectivo cluster.
- **Condição de Convergência**: Repete os passos 2 e 3 até que o critério de convergência tenha sido atingido. Em geral, considera-se um valor de tolerância do erro quadrado médio global abaixo do qual a distribuição dos pontos de dados pelos clusters é considerada satisfatória.

Clusterização/ Agrupamento

- Técnicas Tradicionais
  - Redes Neurais
  - Algoritmos Genéticos
  - Estatística

- Algoritmos Específicos
  - K-Means
  - Fuzzy K-Means
  - K-Modes
  - K-Medoids
  - K-Prototypes

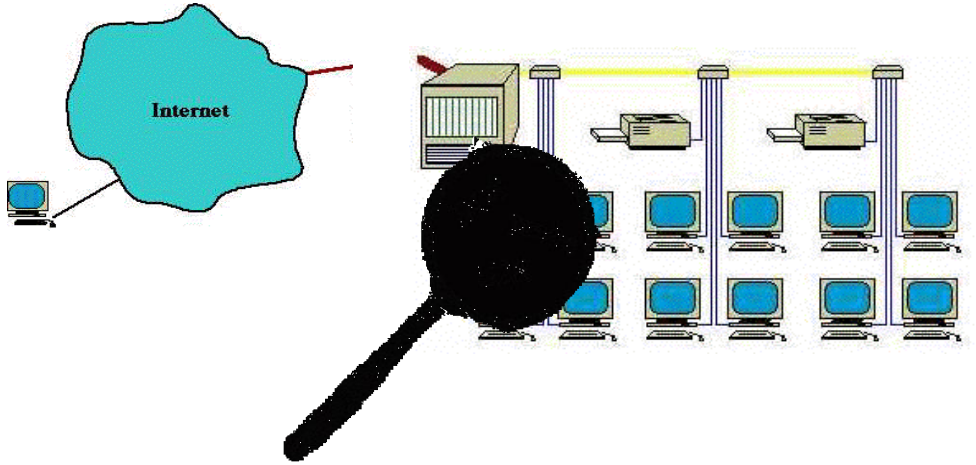
Clusterização/ Agrupamento

- Exemplo de Aplicação:
  - **Marketing direto**
  - **Segmentação de clientes**
  - **Mineração de sub-estruturas em imagens**

Detecção de desvios

Detecção de desvios

- Determinar desvios significativos do comportamento normal
- Aplicações:
  - Detecção de fraudes em cartões de crédito
  - Detecção de invasão em redes de computadores
  - Detecção de eventos através de mensagens do Twitter



Detecção de desvios

Pode ser on-line ou off-line:

- Detecção de Desvios On-line:
  - Mecanismos computacionais ativos devem monitorar a base de dados a fim de identificar a entrada de novos valores que sejam espúrios ou *outlines*. Somente novos dados são analisados.

- Detecção de Desvios Off-line
  - O Banco de Dados é integralmente analisado na busca por *outlines*. Durante o processo de análise, não são incluídos novos dados na base.

Referências

- GOLDSCHMIDT, R.; PASSOS, E. e BEZERRA, E.. **DataMining. Conceitos, Técnicas, Algoritmos, Orientações e Aplicações.** 2a. ed. Elsevier Editora, 2015.
- TAN, PN; STEINBACH, M. e KUMAR, V.; KARPATNE, A. **Introduction to Data Mining**. 2a. ed. Person Editora, 2018.
- Slides do prof. José Leomar Todesco (UFSC)