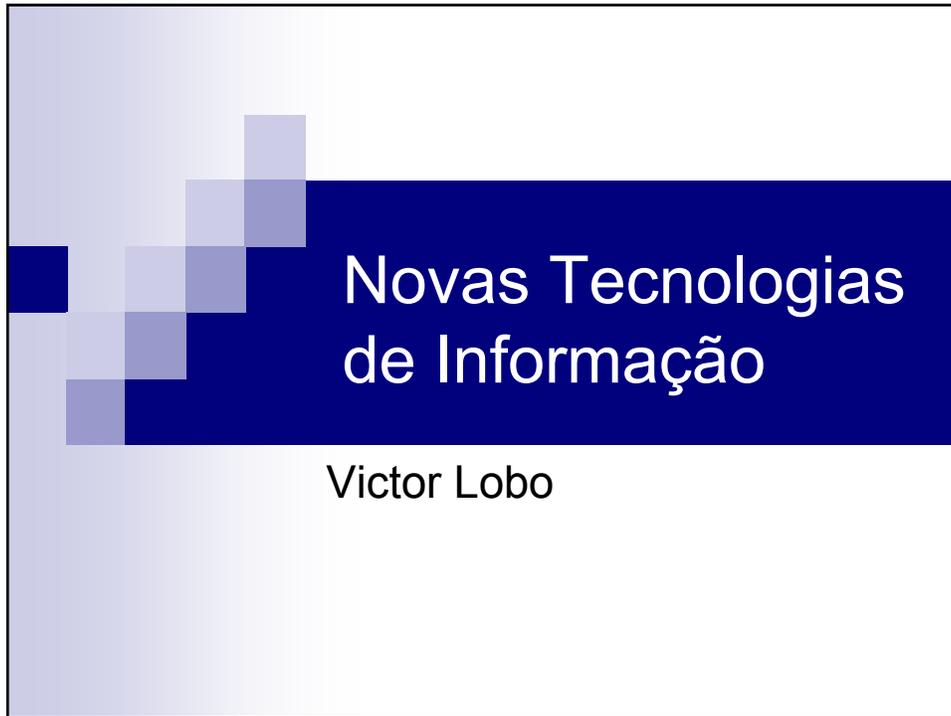


# Cap.1 – Introdução a NTI, DataMining, e Aprendizagem Automática

V 3.0, V.Lobo, EN/ISEGI, 2005



## Objectivos gerais

- Abrir horizontes em temas actuais
- Aprender técnicas de “Business Intelligence”, ou “Sistemas de apoio à decisão”
  - Métodos de DataMining
    - Pesquisa de informação em grandes bases de dados
  - Métodos de Optimização
    - Resolver problemas de pesquisa “complicados”
- Trabalhar sózinho em novas áreas de tecnologia
  - Fazer um trabalho de pesquisa

## Programa

1. **Introdução a Data Mining e Aprendizagem Automática**
2. **Aprendizagem com modelos bayesianos e aprendizagem baseada em protótipos**
3. **Redes Neurais – Perceptrão multicamada (MLP)**
4. **Redes Neurais – Mapas auto-organizados (SOM)**
5. **Redes Neurais – Outros tipos de redes**
6. **Árvores de decisão e regras de associação**
7. **Técnicas de otimização**

## Programa

- **1 - Introdução a Data Mining e Aprendizagem Automática**
  - Objectivos de Data Mining. Data Mining e CRM (Customer Relation Management). Data Mining no ciclos de análise de problemas. Técnicas de visualização de dados. Aprendizagem em problemas de predição e problemas de aglomeração (Clustering). Problemas supervisionados e problemas não supervisionados. Representação de dados. Pré-processamento de dados para Data Mining e Aprendizagem. Generalização e sobre-especificação.

## Programa

### ■ 2 - Aprendizagem com modelos bayesianos e aprendizagem baseada em protótipos

- Noção de modelo bayesiano. Aprendizagem MAP (Maximum a posteriori), e ML (Máxima Verosimilhança). Modelos de custos. Aprendizagem não paramétrica: k-vizinhos e janelas de Parzen.
- Diferentes origens para modelos baseados em protótipos: modelos baseados em memória, modelos baseados em instâncias. Técnicas de CBR (Case Based Reasoning), e IBL (Instance based reasoning). Técnicas de minimização para classificadores baseados em protótipos: Data Condensing, K-Editing, Q-Sets.

## Programa

### ■ 3- Redes Neurais Perceptrão multicamada (MLP)

- Introdução histórica. Inspiração e modelos oriundos da biologia. Perceptrão Simples: modelo, algoritmo de treino, convergência, e limitações. Perceptrão multicamada: modelo e algoritmo de retropropagação do erro (BP). Variantes sobre o algoritmo básico de retropropagação do erro, e outras técnicas para treinar redes multicamada. Vantagens e potenciais problemas das redes multicamada.

## Programa

- **4- Redes Neurais**
  - **Mapas auto-organizados (SOM)**
    - Mapas auto-organizados (SOM): modelo, algoritmo de treino, interpretação de resultados, e variantes.
- **5- Redes Neurais**
  - **Outros tipos de redes neuronais**
    - LVQ, RBF, Redes de Hopfield, outras redes recorrentes, e aprendizagem hebbiana.

## Programa

- **6- Árvores de decisão e regras de associação**
  - Conceitos gerais. Regras para indução de árvores de decisão. CART, ID3, C4.5, ID3. Critérios para escolhas de partições. Pruning de árvores de decisão. Extração de regras de associação. Suporte e confiança de regras de associação.
- **7- Técnicas de optimização**
  - Métodos exactos e suas limitações. Métodos de Monte Carlo. Métodos de gradiente estocásticos. Simulated Annealing e algoritmo de metropolis. Pesquisas Tabu. Algoritmos Genéticos.

## Bibliografia

- Data Mining Techniques, for sales and customer support
  - Berry, M., Linoff, G., John Wiley and Sons, 1997
- Principles of Data Mining
  - Hand, D., Mannila, H., Smyth, P.; MIT Press, 2001
- Machine Learning
  - Mitchell, Tom., McGraw-Hill, 1997
- Apontamentos de F.Baço e V.Lobo

## Avaliação

- Exame
- Apresentação oral e escrita de um tema
  - Soma 0, 1 ou 2 valores à nota do exame
  - Entre 500 e 2000 palavras
  - 10 Minutos para a apresentação oral
- Trabalhos de casa
  - Soma 0, 1 ou 2 valores à nota do exame
  - Analisar problemas com o Enterprise Miner (SAS), e outros programas
  - São dados numa semana para entregar na semana seguinte

## Apresentação de um tema

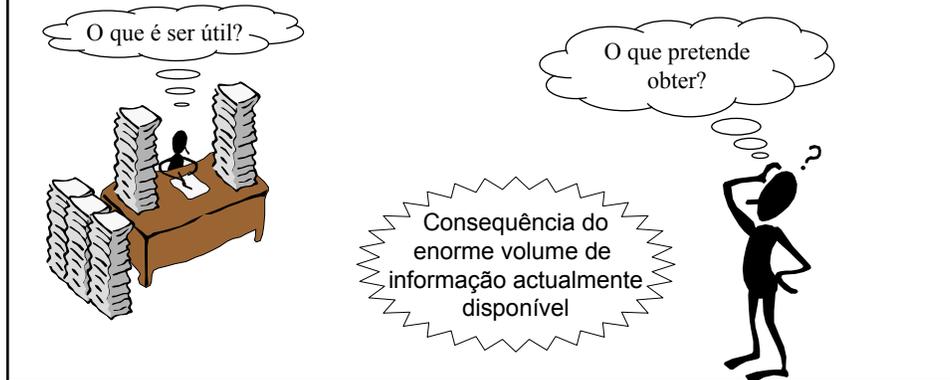
- Objectivo:
  - Demonstrar capacidade para pesquisar/estudar um tema novo, identificando os pontos mais relevantes
  - Demonstrar capacidade de síntese
- Datas
  - Sorteadas na primeira aula
- Temas
  - Escolhidos da lista apresentada



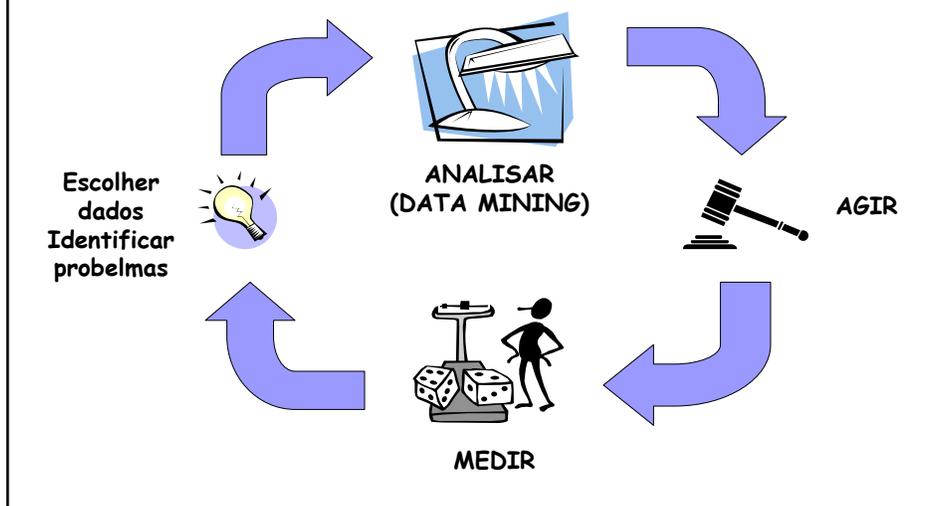
Data Mining

## O que é “Data Mining”?

- “Data Mining” é a pesquisa de informação útil em grande quantidades de dados



## O ciclo de data mining



## Data Mining é

- A utilização de três técnicas diferentes:
  - Bases de dados
  - Estatística
  - **Aprendizagem por máquina.**  
(Machine Learning)
- Para resolver dois tipos de problemas
  - Predição
  - Descobrir novo conhecimento

Vamos estudar tudo isto?



## Predição e novo conhecimento

- Predição
  - é aprender critérios de decisão para ser capaz de classificar casos desconhecidos
- Descobrir novo conhecimento
  - é encontrar padrões desconhecidos existentes nos dados

Gostava de ver exemplos?



## Exemplos

- Detecção de fraudes na utilização de um cartão de crédito
- Deferir, ou não, um pedido de crédito
- Prever os níveis de audiência dos canais de televisão
- Classificar os efeitos hidrofónicos produzidos por diferentes navios
- Analisar as respostas de um inquérito médico
- Escolher clientes a quem direccionar uma campanha de marketing
- Cross-selling, fidelização, etc, etc,



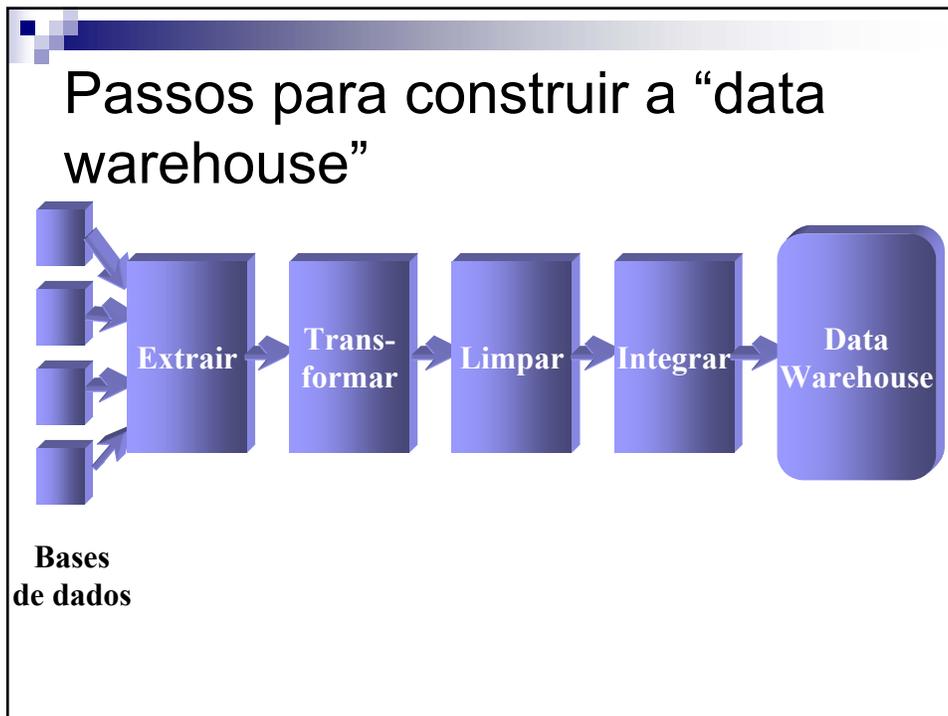
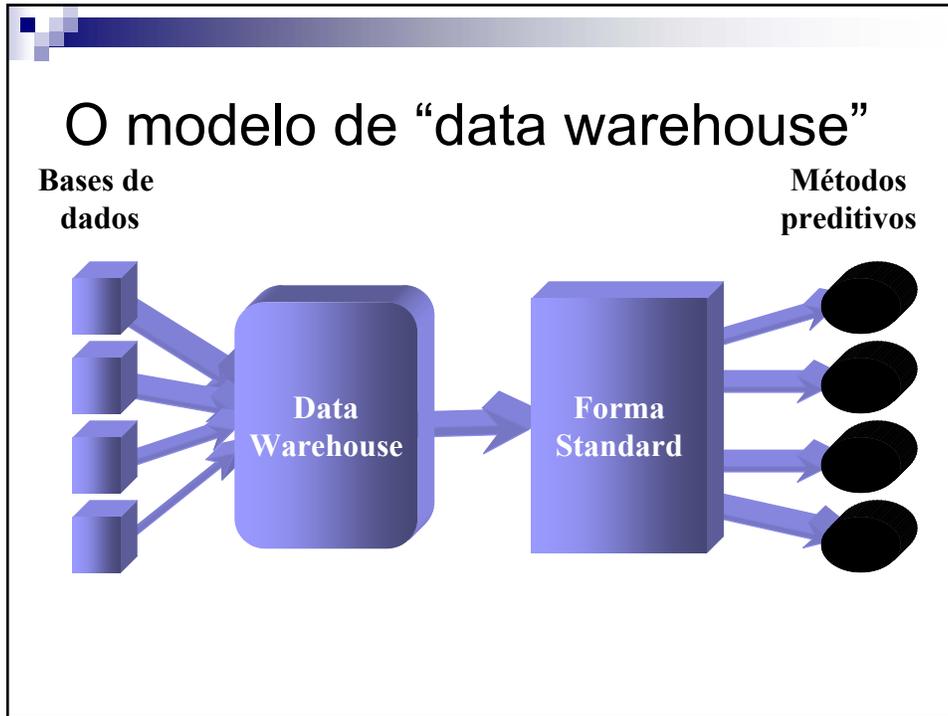
## Como organizar os dados?

- “Data warehouse”
  - É o suporte centralizado de informação importante para a decisão.



# Cap.1 – Introdução a NTI, DataMining, e Aprendizagem Automática

V 3.0, V.Lobo, EN/ISEGI, 2005



## Tipos de problemas

### ■ Predição

- Classificação**
- Regressão



### ■ Descoberta de conhecimento

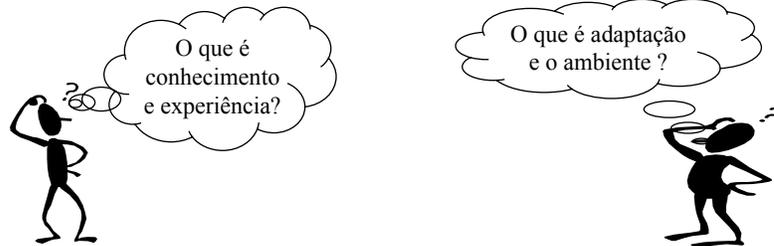
- Detecção de desvios
- Segmentação de bases de dados
- Clustering**
- Regras de associação
- Sumarização
- Visualização
- Pesquisa em texto

## Introdução à aprendizagem

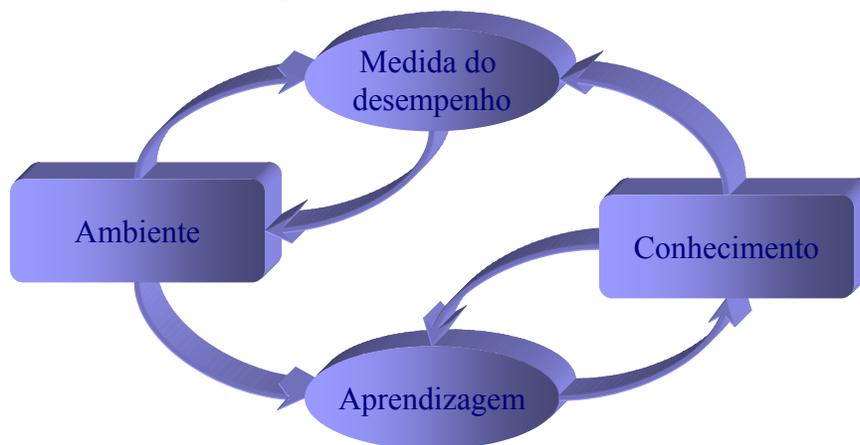
Definições de aprendizagem

## Uma definição

- Aprendizagem é a melhoria do *desempenho* num *ambiente* através da aquisição de *conhecimento* resultante da *experiência* nesse ambiente (Langley, 96)



## Aprendizagem



## Outra definição

- Um computador “**aprende**”, a partir de experiências  $E$  em relação a uma classe de tarefas  $T$  e usando uma medida de desempenho  $P$ , se o seu desempenho para as tarefas  $T$ , medidas por  $P$ , melhora com as experiências  $E$ . (Mitchell, 97)



## Exemplos (1)

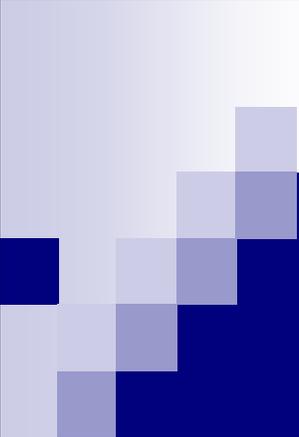
- Jogo das damas
  - **Tarefa T:** *Jogar o jogo das damas.*
  - **Medida de desempenho P:** *percentagem de jogos ganhos contra o adversário.*
  - **Experiência de aprendizagem E:** *jogos jogados contra ele mesmo.*

## Exemplos (2)

- Reconhecimento de palavras manuscritas
  - **Tarefa T:** *reconhecimento e classificação de palavras manuscritas através de imagens.*
  - **Medida de desempenho P:** *percentagem de palavras correctamente classificadas.*
  - **Experiência de aprendizagem E:** *uma base de dados de palavras manuscritas previamente classificadas.*

## Exemplos (3)

- Condução de um veículo numa autoestrada
  - **Tarefa T:** *Condução de um veículo sem condutor numa autoestrada de quatro faixas usando sensores de visão.*
  - **Medida de desempenho P:** *distância média entre troços antes de um erro (quando julgado por um humano).*
  - **Experiência de aprendizagem E:** *uma sequência de imagens e de comandos de condução guardados enquanto se observava um condutor humano.*



## Introdução à aprendizagem

Relações entre o ambiente e o algoritmo de aprendizagem

### Professor/Aluno

- Todo o processo de aprendizagem pode ser caracterizado por um protocolo entre o professor e o aluno.
- O professor pode variar entre o tipo dialogante e o não cooperante.



## Protocolos Professor/Aluno

- Professor nada cooperante
  - Só dá os exemplos => **não supervisionada**
- Professor cooperante
  - Dá exemplos classificados => **supervisionada**
- Professor pouco cooperante
  - Só diz se os resultados estão certos ou errados  
=> **aprendizagem por reforço**
- Professor dialogante - ORÁCULO

## Acesso aos exemplos

- Aprendizagem “offline”
  - Todos os exemplos são apresentados ao mesmo tempo
- Aprendizagem “online”
  - Os exemplos são apresentados um de cada vez
- Aprendizagem mista
  - Uma mistura dos dois casos anteriores

## Dificuldades na aprendizagem

- Adequabilidade da representação do conhecimento à tarefa que se quer aprender
- Informação redundante
  - Quais os atributos importantes para a tarefa?
- Ruído do ambiente
  - Ruído na classificação dos exemplos ou nos valores dos atributos.

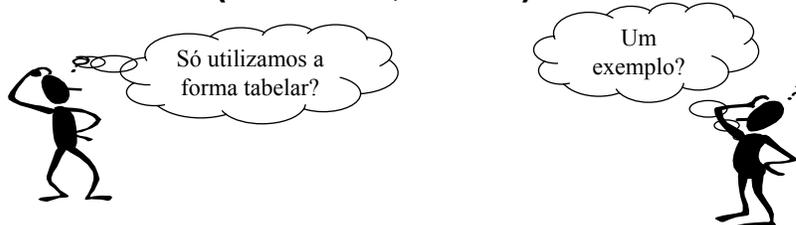
## Introdução à aprendizagem

Formas de representação de conhecimento

## Descrição dos exemplos

- Descrição relacional
  - descrição do mundo dos blocos
- Descrição através de uma tabela de pares de valores

**(Atributo, Valor)**



## Descrição tabelar

Sky	AirTemp	Humid.	Wind	Water	Forecast	Sport
<i>Sunny</i>	<i>Warm</i>	<i>Normal</i>	<i>Strong</i>	<i>Warm</i>	<i>Same</i>	<i>Yes</i>
<i>Sunny</i>	<i>Warm</i>	<i>High</i>	<i>Strong</i>	<i>Warm</i>	<i>Same</i>	<i>Yes</i>
<i>Rainy</i>	<i>Cold</i>	<i>High</i>	<i>Strong</i>	<i>Warm</i>	<i>Change</i>	<i>No</i>
<i>Sunny</i>	<i>Warm</i>	<i>High</i>	<i>Strong</i>	<i>Cool</i>	<i>Change</i>	<i>Yes</i>

## Tipos de atributos

- Booleanos ou binários
  - Só toma dois valores
- Nominais
  - Tomam um conjunto de valores não ordenados
- Ordinais
- Numéricos

## Representação em extensão e em intenção

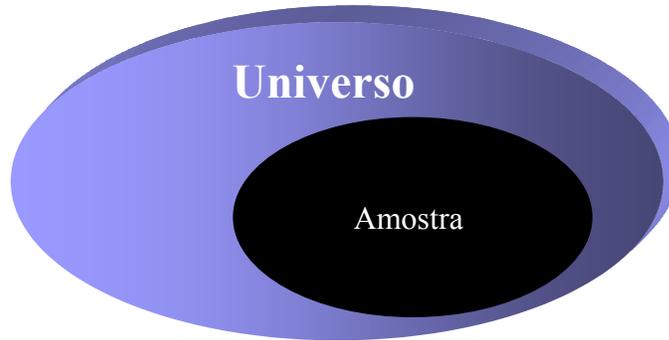
- Uma linguagem de descrição de conceitos
- O que é uma representação em extensão?
- O que é uma representação em intenção?

**Intenção + Interpretador = Extensão**

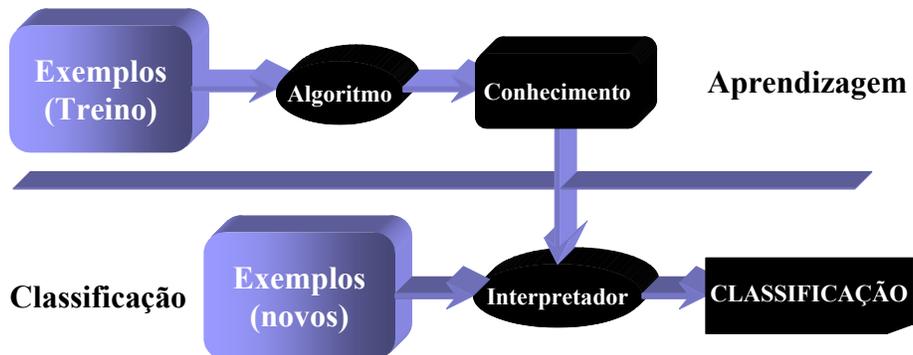
# Cap.1 – Introdução a NTI, DataMining, e Aprendizagem Automática

V 3.0, V.Lobo, EN/ISEGI, 2005

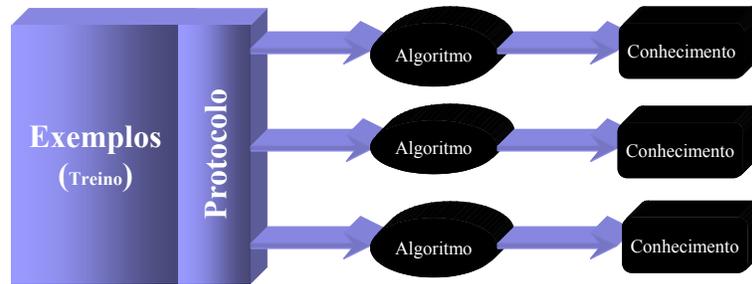
## Os dados



## Fases do processo



## Processo de aprendizagem



## Problemas da aprendizagem

- Exemplos
  - Representação dos exemplos
  - Protocolo professor/aluno
- O algoritmo
- Representação do conhecimento induzido

## Linguagem de descrição

- Linguagens lógicas
  - Combinações lógicas de atributo/valor
- Unidades de limiar
  - Redes neuronais
- Modelos probabilísticos
- Lógica de primeira ordem



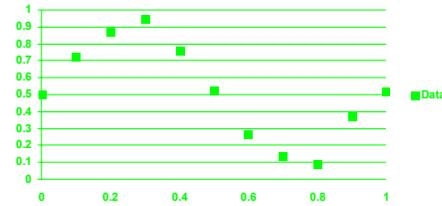
## Algumas definições

- O que é uma descrição mais geral?
- O que é uma descrição mais específica?
- O que é o processo de **generalização**?
- E o processo de especialização?
- O que é o "overfitting"?



## Exemplo de overfitting

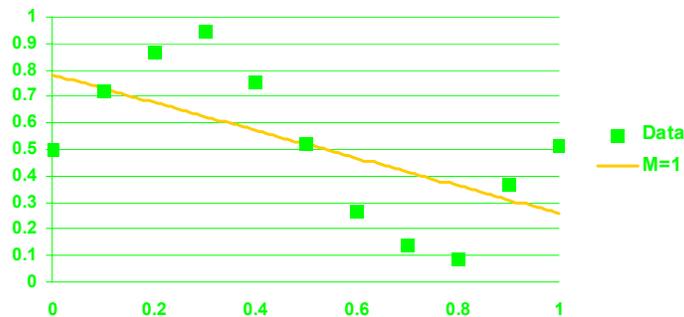
- Seja um conjunto de 11 pontos.
- Encontrar um polinómio de grau M que represente esses 11 pontos.



$$y(x) = \sum_{i=0}^M w_i x^i$$

## Aproximação M = 1

$$y(x) = w_0 + w_1 x$$

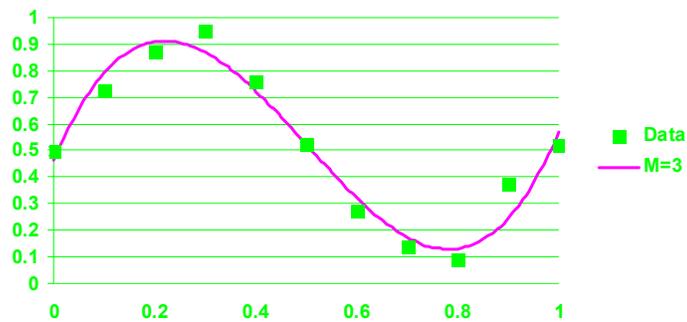


# Cap.1 – Introdução a NTI, DataMining, e Aprendizagem Automática

V 3.0, V.Lobo, EN/ISEGI, 2005

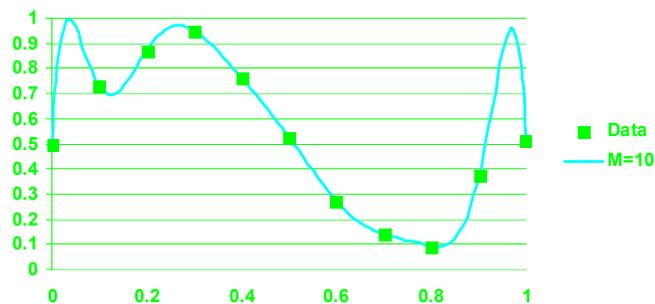
## Aproximação M = 3

$$y(x) = w_0 + w_1x + w_2x^2 + w_3x^3$$



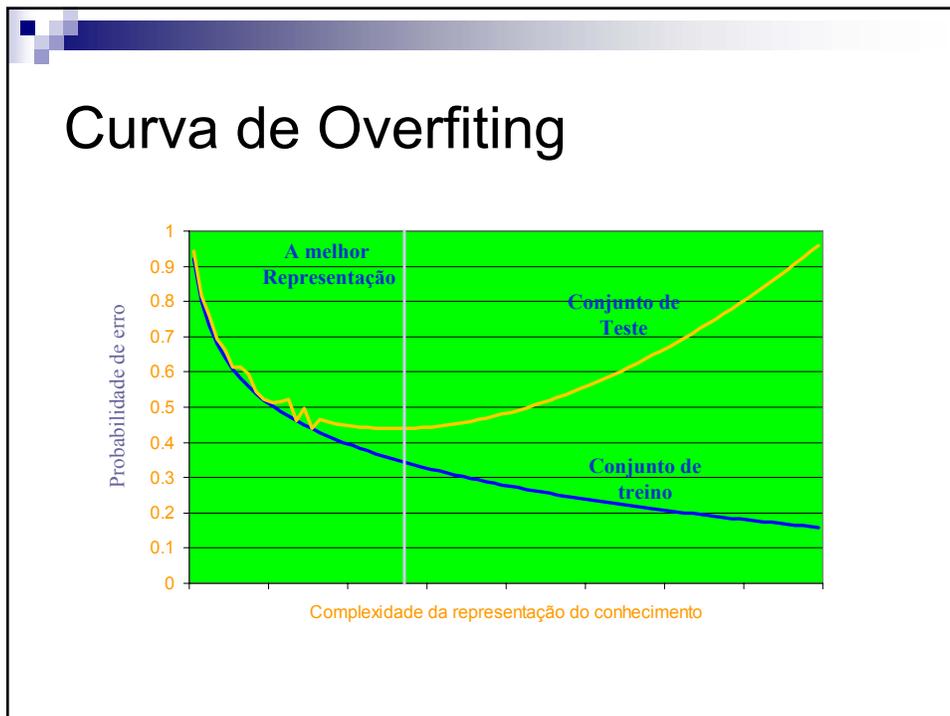
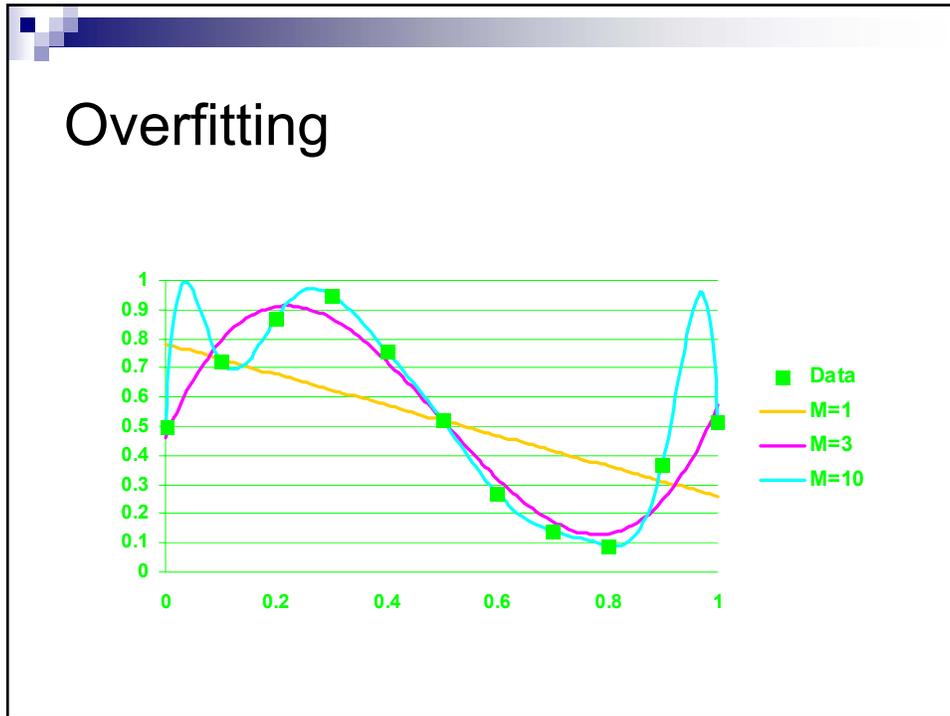
## Aproximação M = 10

$$y(x) = w_0 + w_1x + w_2x^2 + w_3x^3 + w_4x^5 + w_6x^6 + w_7x^7 + w_8x^8 + w_9x^9 + w_{10}x^{10}$$



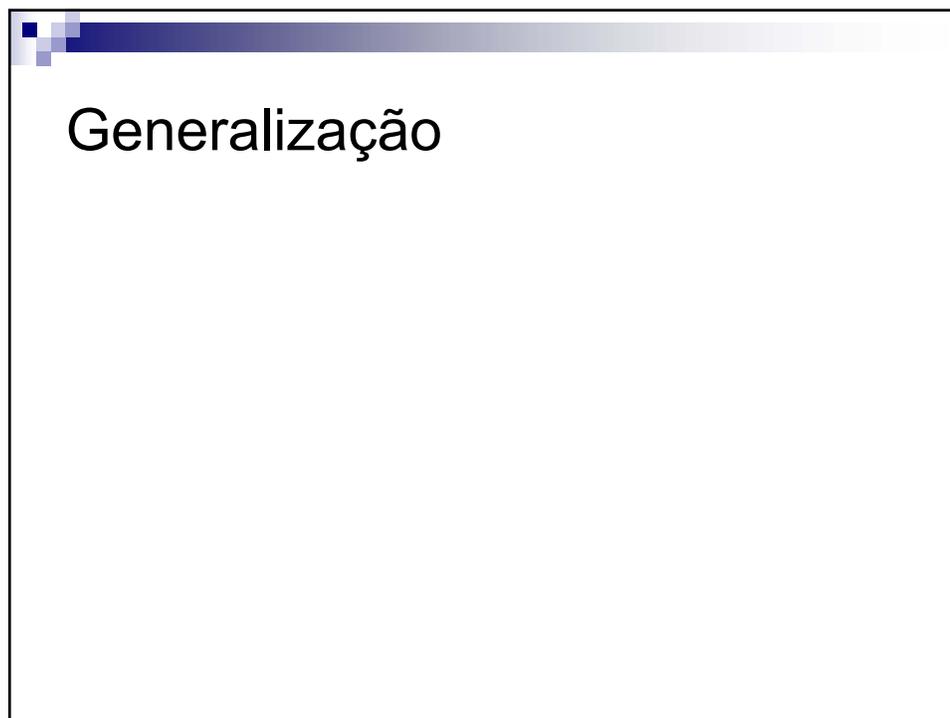
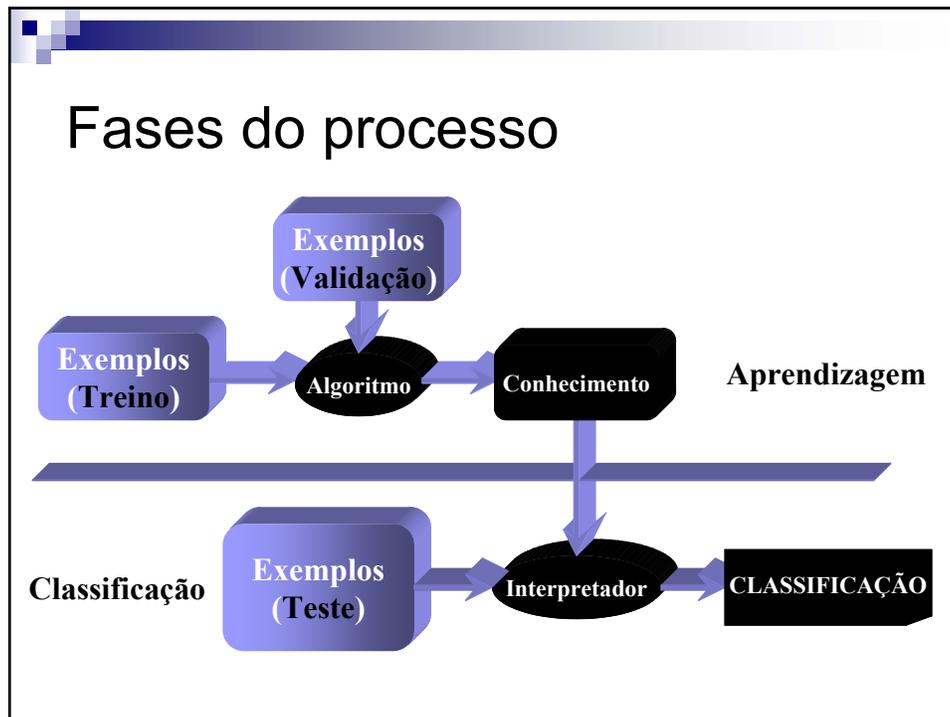
# Cap.1 – Introdução a NTI, DataMining, e Aprendizagem Automática

V 3.0, V.Lobo, EN/ISEGI, 2005



# Cap.1 – Introdução a NTI, DataMining, e Aprendizagem Automática

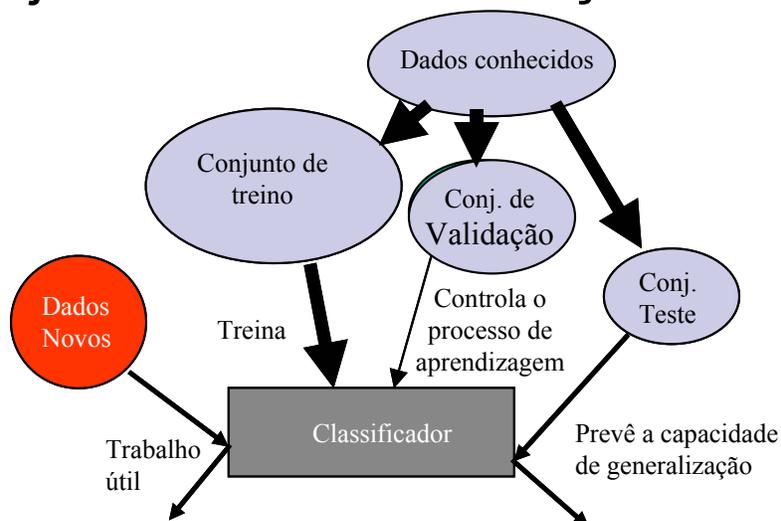
V 3.0, V.Lobo, EN/ISEGI, 2005



## Generalização

- O objectivo não é aprender a agir no conjunto de treino mas sim no universo “desconhecido” !
  - Como preparar para o desconhecido ?
- Manter um conjunto de teste “de reserva”

## Conjunto de treino/validação/teste



## Divisão dos dados

- Conjunto de treino
  - Quanto maior, melhor o classificador obtido
- Conjunto de validação
  - Quanto maior, melhor a estimação do treino óptimo
- Conjunto de teste
  - Quanto maior, melhor a estimação do desempenho do classificador

## Processo de aprendizagem

- A aprendizagem é um processo de optimização
- Algoritmo de optimização
  - Método do gradiente
  - Subir a encosta
  - Guloso
  - Algoritmos genéticos
  - “Simulated annealing”
- Formas de adquirir o conhecimento



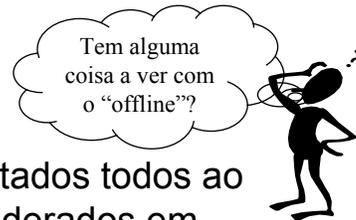
## Formas de adquirir o conhecimento

### ■ Incremental

- Os exemplos são apresentados um de cada vez e a estrutura de representação vai-se alterando

### ■ Não incremental

- Os exemplos são apresentados todos ao mesmo tempo e são considerados em conjunto.

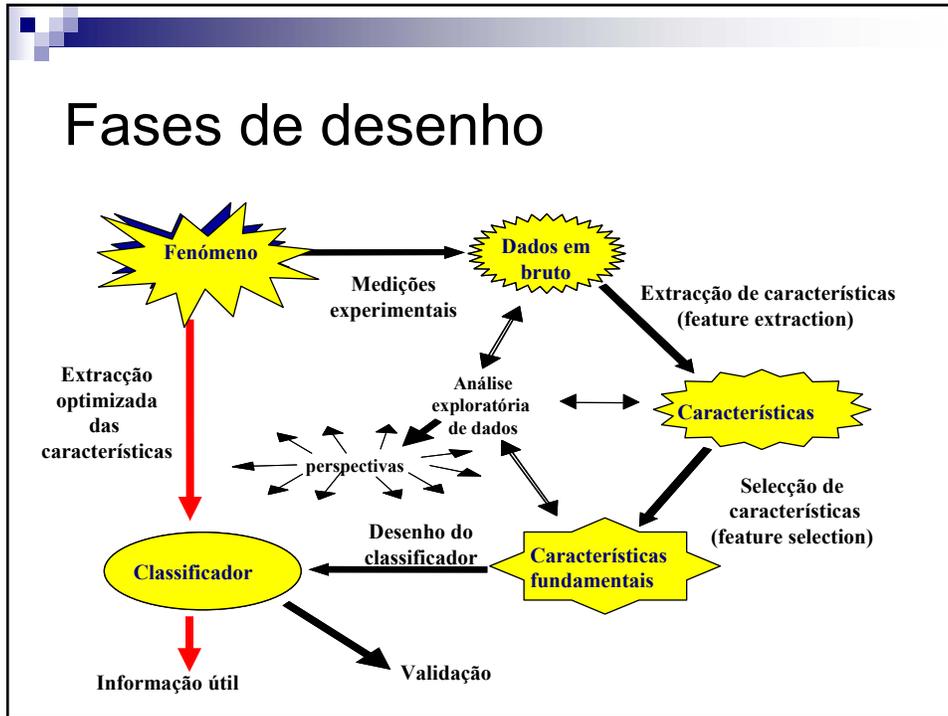


## Introdução à aprendizagem

Projecto de um sistema com aprendizagem

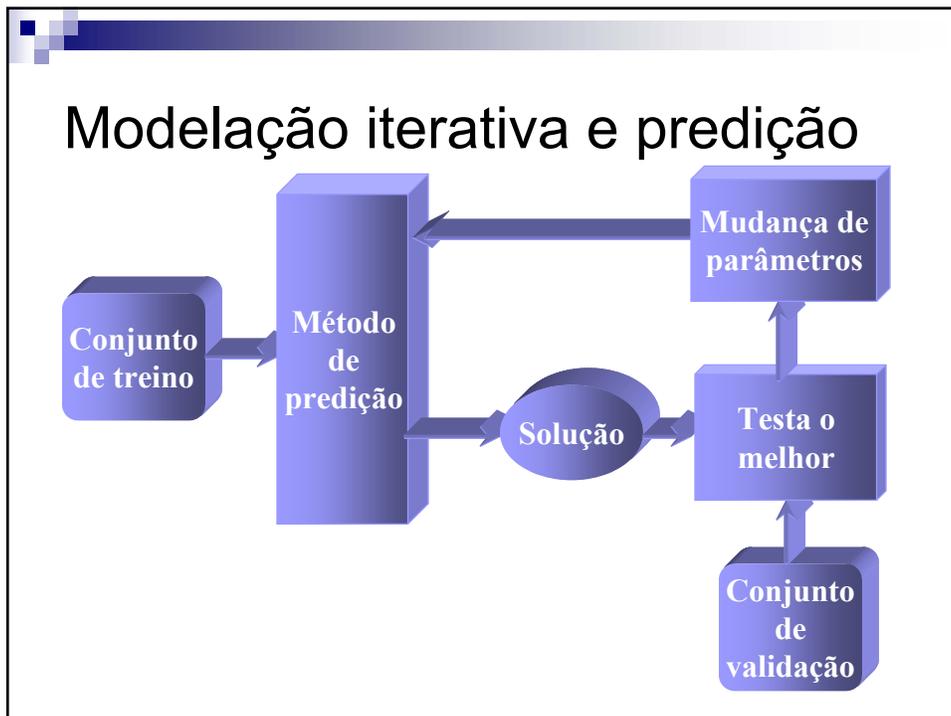
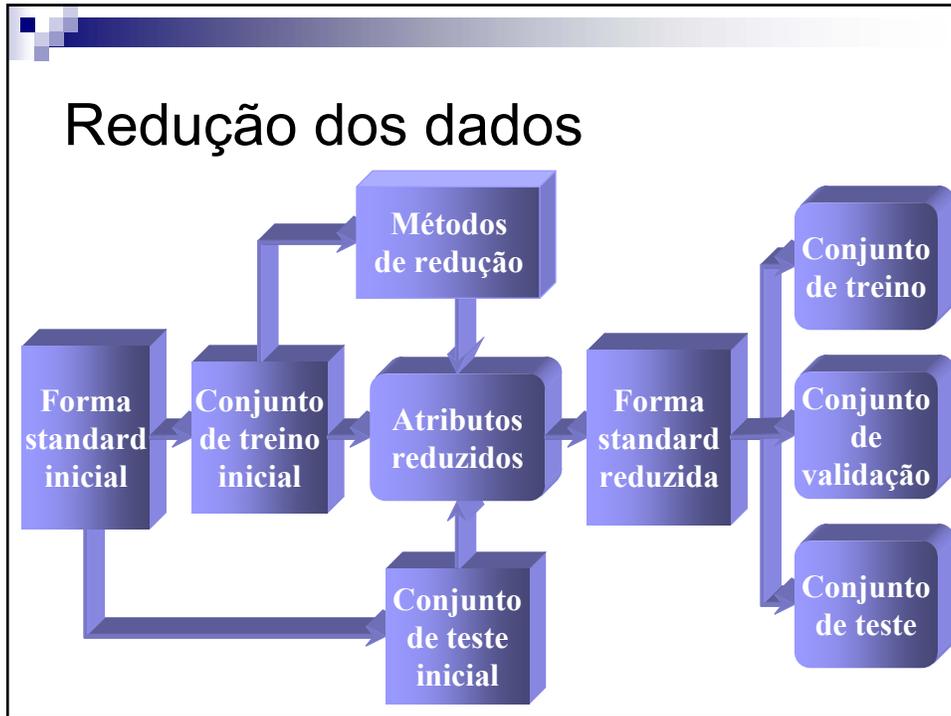
# Cap.1 – Introdução a NTI, DataMining, e Aprendizagem Automática

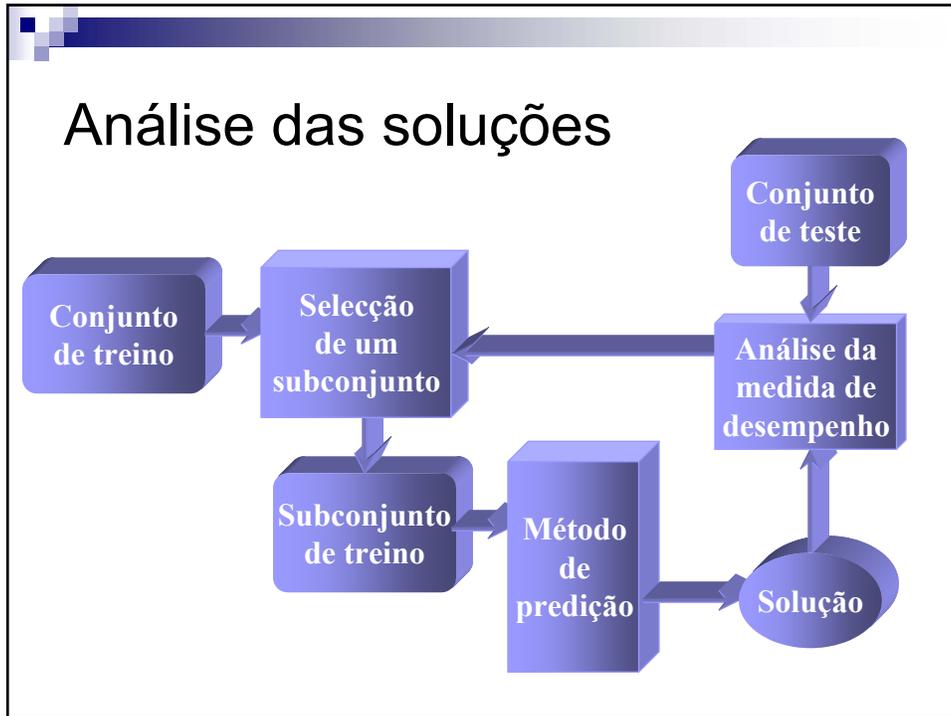
V 3.0, V.Lobo, EN/ISEGI, 2005



# Cap.1 – Introdução a NTI, DataMining, e Aprendizagem Automática

V 3.0, V.Lobo, EN/ISEGI, 2005



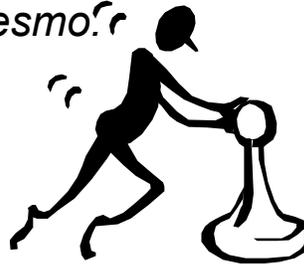


## Introdução à aprendizagem

Um exemplo simples de um problema de aprendizagem

## Jogo das damas

- **Tarefa T:** *Jogar o jogo das damas.*
- **Medida de desempenho P:** *percentagem de jogos ganhos contra o adversário.*
- **Experiência de aprendizagem E:** *jogos jogados contra ele mesmo.*



## Conjunto de treino

- Regras dadas pelo programador
- Um histórico de jogos de damas, quer ao nível de aberturas quer ao nível de finais
- Jogar contra um humano
- O sistema de aprendizagem joga contra ele mesmo

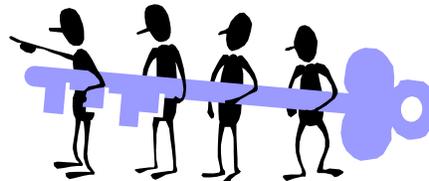


## Problemas a resolver

- “Credit assignment”
  - medida da qualidade das possíveis jogadas.  
Pode ser dada uma sequência de jogadas e só fim se sabe o resultado das mesmas.
- Que tipo de professor existe
  - Dialogante ou cooperante
- Representatividade do conjunto de treino
  - Até que ponto o conjunto de exemplos representa bem todos os tabuleiros possíveis ?

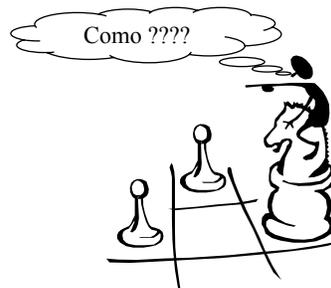
## Passos para a solução

- Qual o tipo de conhecimento a ser aprendido?
- A representação de uma função objectivo.
- Um mecanismo de aprendizagem.



## Conhecimento a ser aprendido

- Dado um tabuleiro pretende-se aprender o melhor movimento a ser feito. Conhece-se o espaço de pesquisa mas desconhece-se a melhor estratégia.



## O melhor movimento

- O melhor movimento pode ser descrito como uma função

$$\text{EscolherMov} : B \rightarrow M$$

em que B é o conjunto dos tabuleiros válidos e M é o conjunto dos movimentos válidos

- Outra função

$$V : B \rightarrow \mathcal{R}$$

em que V é uma medida da qualidade da jogada

$$V(b) = \begin{cases} 100 & \text{ganha} \\ 0 & \text{empata} \\ -100 & \text{perde} \\ V(b') \end{cases}$$

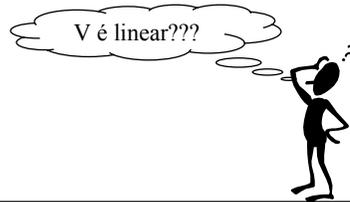
e b' é o melhor tabuleiro que se pode obter a partir de b

## Uma definição operacional de $V$

- $x_1$  n° de peças pretas
  - $x_2$  n° de peças brancas
  - $x_3$  n° de rainhas pretas
  - $x_4$  n° de rainhas brancas
  - $x_5$  n° de peças brancas que podem ser comidas
  - $x_6$  n° de peças pretas que podem ser comidas
- Descrição operacional da função objectivo

$$\hat{V}(b) = w_0 + \sum_{i=1}^6 w_i x_i$$

V é linear???



## Nova formulação

- **Tarefa T:** jogar às damas;
- **Medida de desempenho P:** percentagem de jogos ganhos contra um oponente
- **Experiência de treino:** praticar jogos contra ele próprio.
- **Função objectivo:**  $V : B \rightarrow \mathcal{R}$
- **Representação da função objectivo:**

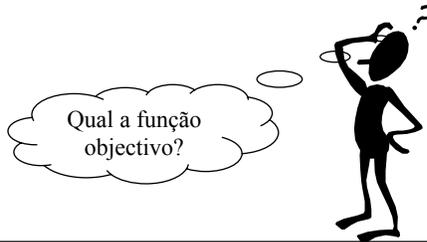
$$\hat{V}(b) = w_0 + \sum_{i=1}^6 w_i x_i$$

## Aprendizagem

- O conjunto de treino  $E$ , é definido por  $(b, V_{\text{treino}}(b))$
- Regra para estimação dos valores de treino

$$V_{\text{treino}}(b) \leftarrow \hat{V}(\text{Sucessor}(b))$$

- Objectivo: a partir do conjunto de treino calcular os  $w_i$



## Função objectivo

- A função objectivo é o erro quadrático
- Pela regra de actualização do erro quadrático

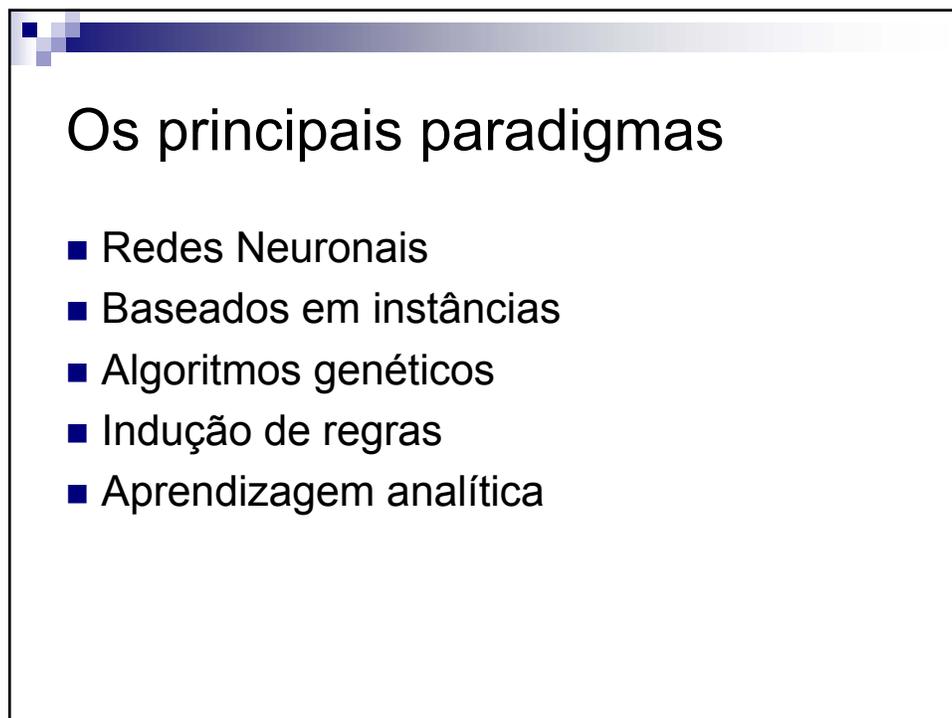
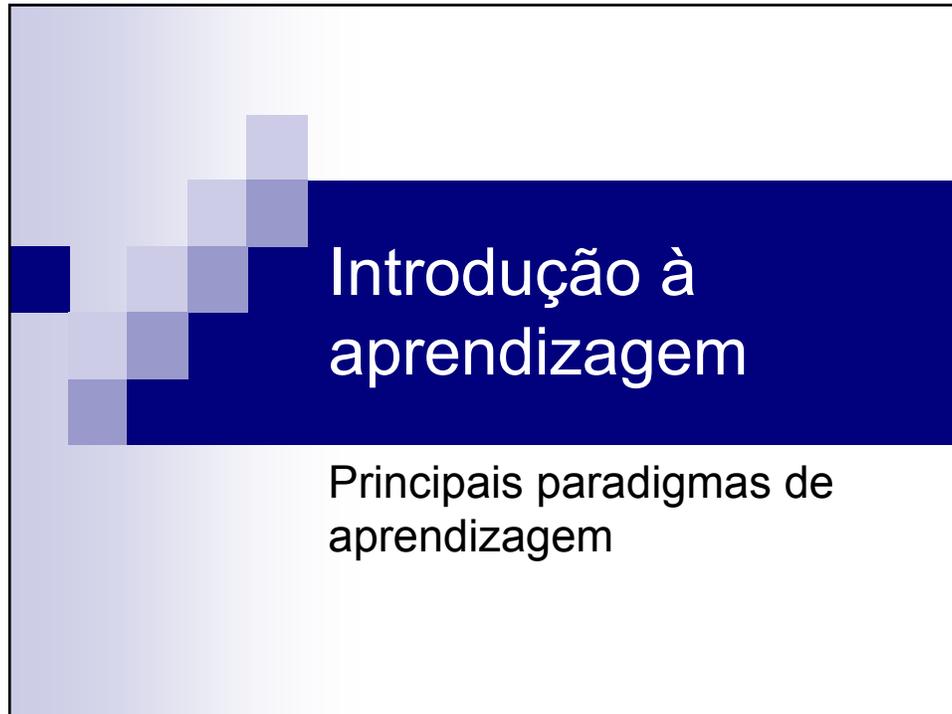
$$E = \sum_{(b, V_{\text{treino}}(b)) \in TS} (V_{\text{treino}}(b) - \hat{V}(b))^2$$

$$\Delta w_i = \eta (V_{\text{train}}(b) - \hat{V}(b)) x_i$$

e pretende-se

$$\min_{w_0, \dots, w_6} E$$





## Alguns pontos para meditar(1)

- Que modelos são mais adequados para um caso específico?
- Que algoritmos de treino são mais adequados para um caso específico?
- Quantos exemplos são necessários? Qual a confiança que podemos ter na medida de desempenho?
- Como pode o conhecimento a priori ajudar o processo de indução?

## Alguns pontos para meditar(2)

- Qual a melhor estratégia para escolher o processo exemplo? Em que medida a estratégia altera o processo de aprendizagem?
- Quais as funções objectivo que se devem escolher para aprender? Poderá esta escolha ser automatizada?
- Como pode o sistema alterar automaticamente a sua representação para melhorar a capacidade de representar e aprender a função objectivo?