



# **Mini Curso**

## **Inteligência Artificial e Aplicações para Sistemas de Potência**

**Curitiba, 08 de novembro de 2009.**

Slide 1



## Introdução à Inteligência Artificial

Parte 1 - **Dr. Eng. Milton Pires Ramos/TECPAR**

- Fundamentos, conceitos, histórico e evolução;
- Subáreas e domínios de aplicação;
- Engenharia do Conhecimento;

Parte 2 - **Prof. Dr. Julio Cesar Nievola/PUCPR**

- Redes Neurais Artificiais;
- Computação Evolucionária;
- Mineração de Dados.

Curitiba, Brazil November 8 - 12, 2009 1

Slide 2




## Inteligência Artificial



Curitiba, Brazil November 8 - 12, 2009 2

Slide 3



## ISAP 2009

### O que é INTELIGÊNCIA ??

- Não existe uma definição geral e completa para o que seja inteligência;
- É possível avaliar se algum sistema (natural ou artificial) é ou não inteligente;
- É possível determinar atributos para que um sistema seja considerado inteligente.

Curitiba, Brazil November 8 - 12, 2009 3

Slide 4



## ISAP 2009

### Inteligência Artificial:

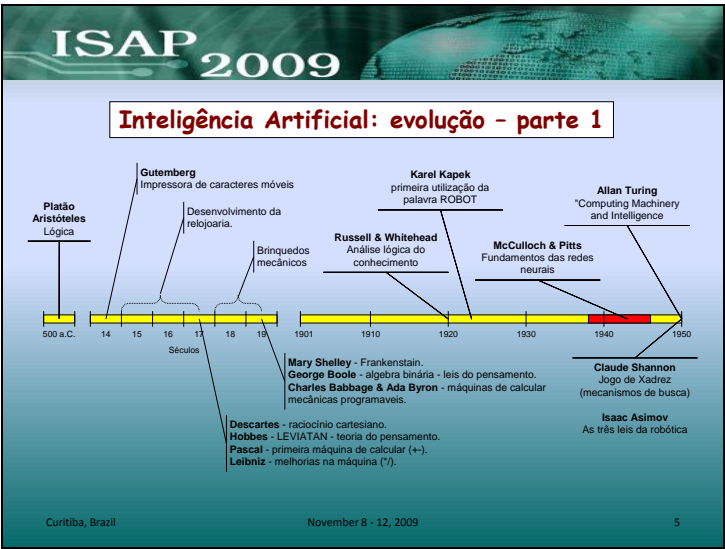
**Inteligência Artificial** - parte das Ciências da Computação que busca simular ou emular o comportamento humano inteligente em termos de processos computacionais. *[Schalkoff, 1990]*

**Inteligência Artificial** – disciplina com duas vertentes:

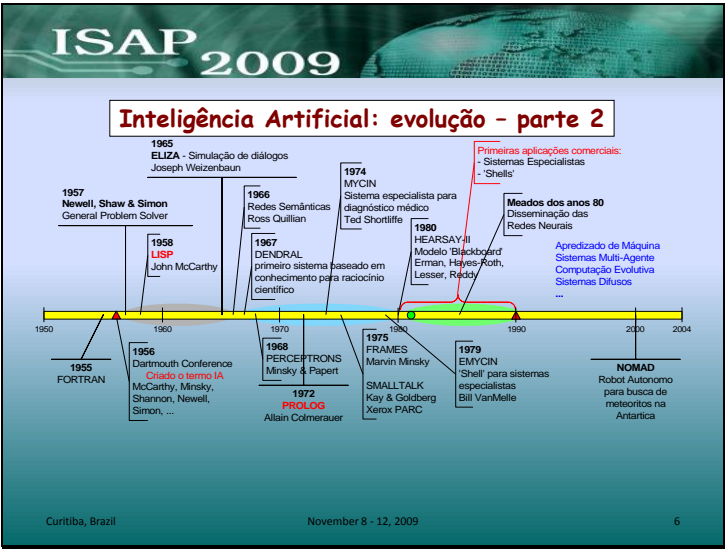
- **Científica** – voltada a entender os requisitos e mecanismos para a inteligência, considerando os seus vários tipos, em humanos, outros animais, computadores e robôs;
- **de Engenharia** – voltada à aplicação deste conhecimento no projeto e construção de novos tipos de máquinas e ferramentas de software mais eficientes. *[Sloman, 2001]*

Curitiba, Brazil November 8 - 12, 2009 4

Slide 5



Slide 6



Slide 7

ISAP 2009

**IA: diferenças**

<b>Informática convencional</b>	<b>Inteligência Artificial</b>
numérica	simbólica
procedural	declarativa
algorítmica	heurística

Curitiba, Brazil November 8 - 12, 2009 7

Slide 8

ISAP 2009

**IA: subáreas e aplicações**

- Engenharia do conhecimento;
- Processamento de linguagem natural;
- Aprendizado automático;
- Mineração de dados;
- Sistemas difusos;
- Planejamento automático;
- Computação evolucionária;
- Robótica & visão artificial;
- Redes neurais artificiais;

- . . . < lista não exaustiva !! >

Curitiba, Brazil November 8 - 12, 2009 8

Slide 9

ISAP 2009

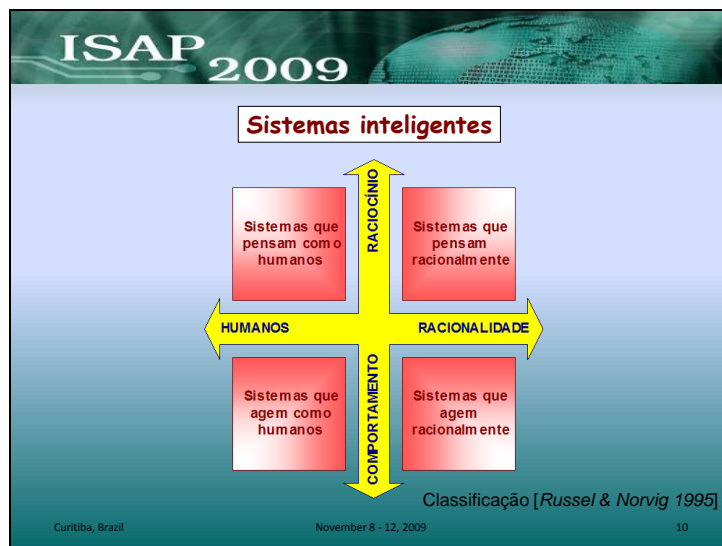
**IA: subáreas e aplicações**

- Engenharia, robótica, matemática
- Aero-espacial, militar;
- Indústria, energia, telecomunicações;
- Arquitetura, direito, comércio, finanças, bolsa de valores;
- Medicina, biologia (*biologia molecular* - *bioinformática*);
- Educação, jogos/entretenimento, literatura;
- Gestão da informação, interface humano/máquina;


< lista não exaustiva !! >

Curitiba, Brazil November 8 - 12, 2009 9

Slide 10



Slide 11



## ISAP 2009

### IA: estado da arte

- Sistemas Híbridos (neuro-fuzzy , neuro-cognitivos, etc);
- Sistemas Multi-Agente (IAD);
- Swarm – Enxame de partículas
- Rough Sets – Conjuntos aproximados

Curitiba, Brazil November 8 - 12, 2009 11

Slide 12



## ISAP 2009

### IA: referências

Luger, G.F. **Inteligência Artificial: estruturas e estratégias para a solução de problemas complexos** (4 edição). Bookman, 2004.

Russell, S., Norvig, P. **Artificial Intelligence: A modern approach**. Prentice Hall, 1995. (2a. edição 2002) (1a. edição em português 2004)

Bittencourt, G. **Inteligência Artificial: Ferramentas e teorias** (2a. edição). Editora da UFSC, Florianópolis, 2001.

Rezende, S.O. **Sistemas Inteligentes – Fundamentos e aplicações**. Manole, São Paulo, 2003.


Schalkoff, R.J. **Artificial Intelligence: An engineering approach**. McGraw-Hill, Singapore, 1990.

IEEE-CS <http://www.computer.org/portal/web/intelligent/home>

AITopics <http://www.aaai.org/AITopics/pwwiki.phpAITopics/HomePage>

Curitiba, Brazil November 8 - 12, 2009 12

Slide 13



**ISAP 2009**

**ENGENHARIA DO  
CONHECIMENTO**

**Sistemas Especialistas**

Curitiba, Brazil November 8 - 12, 2009 13

Slide 14



**Engenharia do Conhecimento**

**Engenharia do conhecimento** - termo usado para descrever todo o processo de desenvolvimento e construção de sistemas baseados em conhecimento.

*[McGraw & Harbison-Briggs, 1989]*

- Sistemas baseados no conhecimento;
- Sistemas especialistas;
- Ontologias;
- Sistemas de gestão do conhecimento.

Curitiba, Brazil November 8 - 12, 2009 14



Slide 15

**ISAP 2009**

### O que é Conhecimento ?

*“Conhecimento é como uma máquina na cabeça de alguém que recebe dados e informações em uma ponta, analisa e trata estas entradas com relação a todo o conhecimento e experiências já acumulados, fazendo brotar na outra ponta decisões e ações que vão provavelmente gerar mais dados e informações.”*

Knowledge is the {ability skill expertise} to {manipulate transform create} {data information ideas} to {perform skillfully make decisions solve problems}


Milton, N.R. Knowledge Acquisition in Practice: A Step-by-step Guide. Springer, July 2007.

Curitiba, Brazil November 8 - 12, 2009 15

Slide 16



Slide 17



### Sistemas Especialistas


Sistema desenvolvido, a partir do **conhecimento** de um especialista humano, com o objetivo de apresentar a mesma **performance** desse especialista na **solução de problemas** em um domínio específico.

Características ideais de um SE:

- Conhecimento específico do domínio;
- Técnicas de busca;
- Análise heurística;
- Processamento simbólico;
- Capacidade de explicar seu raciocínio.

Curitiba, Brazil November 8 - 12, 2009 17

Slide 18




### SE's - principais aplicações

- **Sistemas de interpretação:** identifica objetos a partir de conjuntos de observações: compreensão de fala, análise de imagens, interpretação geológica.
- **Sistemas de diagnóstico:** deduz possíveis problemas a partir de observações ou sintomas; ex.: diagnósticos médicos, mecânicos, de processos.
- **Sistemas de projeto:** desenvolve configurações de objetos que satisfazem determinados requisitos ou restrições; ex.: projeto de circuitos digitais, projeto de edifícios.
- **Sistemas de monitoração:** comparam observações do comportamento de sistemas com características consideradas necessárias para alcançar objetivos; ex.: monitoração de rede de distribuição elétrica, controle de tráfego aéreo.
- **Sistemas de controle:** governam de forma adaptativa o comportamento de um sistema; ex: robôs, sistemas de produção.

Curitiba, Brazil November 8 - 12, 2009 18

## Slide 19




### SE's - que problemas tratar ?

- problemas complexos cuja solução depende fortemente do conhecimento especializado e da heurística de um especialista humano;
- problemas de base lógica – o especialista no domínio é capaz de formalizar a sua tomada de decisão e justificá-la;
- a solução é valiosa para a organização seja pelo retorno que a distribuição do conhecimento pode trazer, seja pelo valor estratégico da preservação do conhecimento e da experiência, ambos caros e difíceis de formar;
- existem especialistas disponíveis e motivados.

Curitiba, Brazil November 8 - 12, 2009 19

## Slide 20



### SE's - vantagens

As grandes e principais vantagens no desenvolvimento de Sistemas Especialistas são a **preservação e distribuição de conhecimento** caro e difícil de formar, ou seja, de grande valor estratégico para a empresa.

Ainda outra vantagem, normalmente não contabilizada, diz respeito à contribuição para a gestão do conhecimento da organização ou empresa, que acontece durante o desenvolvimento de um Sistema Especialista, simplesmente **pela obrigatoriedade de reunir os especialistas da empresa e outros profissionais envolvidos, para de comum acordo definir conceitos, procedimentos e esclarecer pontos de vista e experiências**

Curitiba, Brazil November 8 - 12, 2009 20

Slide 21

ISAP2009

Sistemas Especialistas

TABLE 1.1 Early Expert Systems

System	Date	Author	Subject
Dendral	1965	Stanford	Infers information about chemical structures
Macsyma	1965	MIT	Performs complex mathematical analysis
Hearsay	1965	Carnegie-Mellon	Natural-language interpretation for subset language
Age	1973	Stanford	Expert-system-generation tool
Mycin	1972	Stanford	Diagnosis of blood disease
Teiresias	1972	Stanford	Knowledge transformation tool
Prospector	1972	Stanford Res. Inst.	Mineral exploration and identification tool
Rosie	1978	Rand	Expert-system-building tool
OPSS	1974	Carnegie-Mellon	Expert-system-building tool
R1	1978	Carnegie-Mellon	Configurator for DEC computer equipment
Caduceus	1975	Univ. of Pittsburgh	Diagnostic tool for internal medicine

Curitiba, Brazil

November 8 - 12, 2009

21

Slide 22

ISAP2009

Busca em um Espaço de Estados

busca em largura

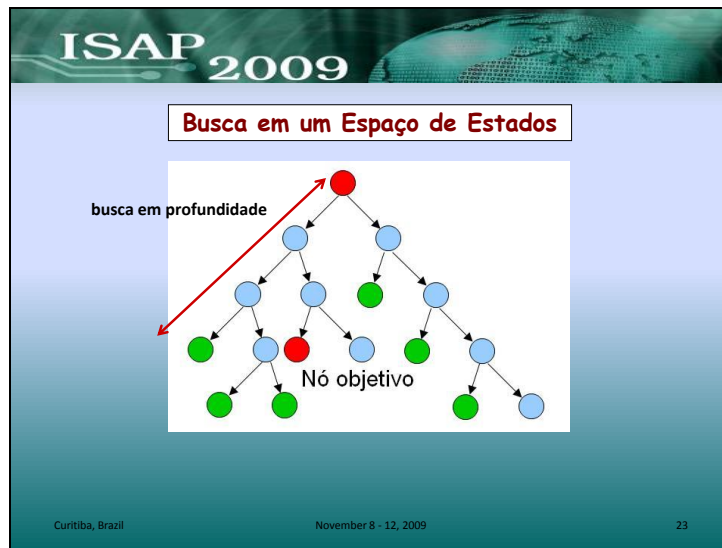
Nó objetivo

Curitiba, Brazil

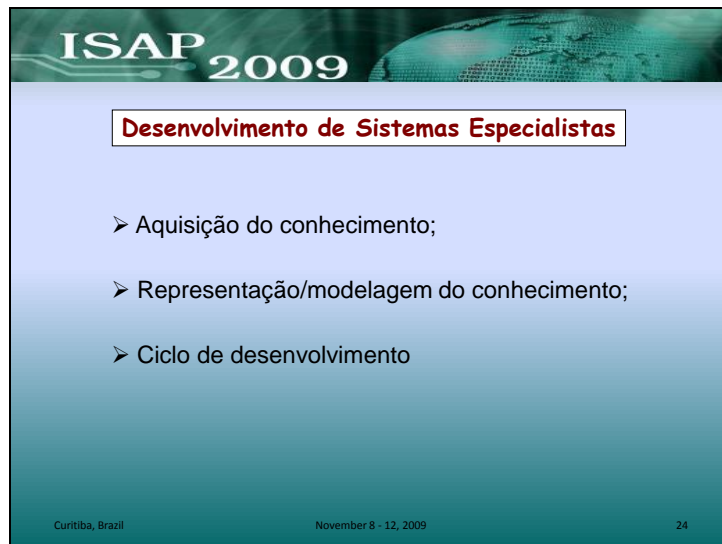
November 8 - 12, 2009

22


Slide 23



Slide 24



Slide 25



### Aquisição do Conhecimento

Processo complexo e longo de extração do conhecimento (experiência) de um especialista humano em determinado domínio.  
(crítico para o sucesso do projeto !!)

- **Imersão na literatura** – leitura de material básico sobre o domínio (fase inicial);
- **Entrevistas não-estruturadas** – processo informal, aparentemente sem um objetivo específico definido;
- **Acompanhamento de casos** – acompanhar o especialista em sua atividade rotineira;
- **Entrevistas estruturadas** – processo formal com objetivos bem claros definidos anteriormente.

Curitiba, Brazil November 8 - 12, 2009 25

Slide 26



### Representação do Conhecimento

#### REGRAS DE PRODUÇÃO

**SE** < premissa 1 >  
**E** < premissa 2 >  
**ENTÃO** < conclusão A >

- - Um dos primeiros e mais tradicionais modelos de representação do conhecimento;
- - Bom nível de representação, simples, de fácil aprendizagem, porém pouco flexível.
- - Modelo mais usado na construção de sistemas especialistas.

Curitiba, Brazil November 8 - 12, 2009 26

Slide 27

ISAP2009

Inputs

{

VazaoInj :Vazao total da agua de injecao

Particulas :Numero de particulas em suspensao

O2galv :Teor de oxigenio [ON]-line medido por par galvanico

corrosao medida por LPR

corrosao medida por resstencia eletrica

oxigenio [ON]-line medido por membrana

a carga nos filtros

a injecao de sequestrante de oxigenio [ON]-[OFF]

}

If VazaoInj >= [VazaoMin]

Then Planta operando

If VazaoInj < [VazaoMin]

Then PLANTA FORA DE OPERACAO

If Planta operando

and PV12 = [OFF]

Then Planta injetando

If Planta operando

and Planta com problemas de corrosao

and Desaeradora dentro dos parametros operacionais

and BseqO2 = [OFF]

Then Sequestrante de Oxigenio Alarme amarelo

and BOMBA DE INJECAO DE SEQUESTRANTE DESLIGADA

and RELIGAR bomba de injecao de sequestrante de Oxigenio

LABEL rec.injex.oxigenio

"RELIGAR bomba de injecao de sequestrante de Oxigenio"

Curitiba, Brazil

November 8 - 12, 2009

27

Slide 28

ISAP2009

Condição

SE a situação é "Conclusão da Operação"

E a operação não é "Enrolamento de forma"

E a operação não é "Alivio de Tensões"

E a operação não é "Retífica Lateral"

E a operação não é "Inspeção Final"

E a medida de OLD foi realizada

E old < OLD\_estim\_naOP - [Tol\_OLD]

Ação

ENTÃO "Ate o momento o anel apresenta valor abaixo do limite especificado (OLD)."

Condição

SE a situação é "Conclusão da Operação"

E a operação atual é "Retífica de Topo Bruto"

E "Ate o momento o anel apresenta valor abaixo do limite especificado (OLD)."

Ação

ENTÃO "Aumentar o numero passe da Retífica Lateral (reduzindo a remocao por passe)."

Hipóteses

OLD\_estim\_naOP = (citt\_old + Def\_estimada\_acumulada + Soma\_Delta\_Radial\_Acum\_Estimado)

Projeção do valor atual de OLD a partir do citt\_OLD e das deformações esperadas, incluindo variações estimadas de Radial.

Def\_estimada\_acumulada Somatório das deformações estimadas na rota até a operação atual, excluindo-se as variações na Radial.

Soma\_Delta\_Radial\_Acum\_Estimado Somatório das variações de Radial estimada (DRadExt) até a operação atual.

Curitiba, Brazil

November 8 - 12, 2009

28

Slide 29

ISAP 2009

Representação do Conhecimento

REDES SEMÂNTICAS

Redes semânticas apresentam relações entre elementos em um domínio. Seus elementos básicos são nós e arcos:

- **nós** – representam os elementos do domínio;
- **arcos** – representam as relações entre estes elementos.

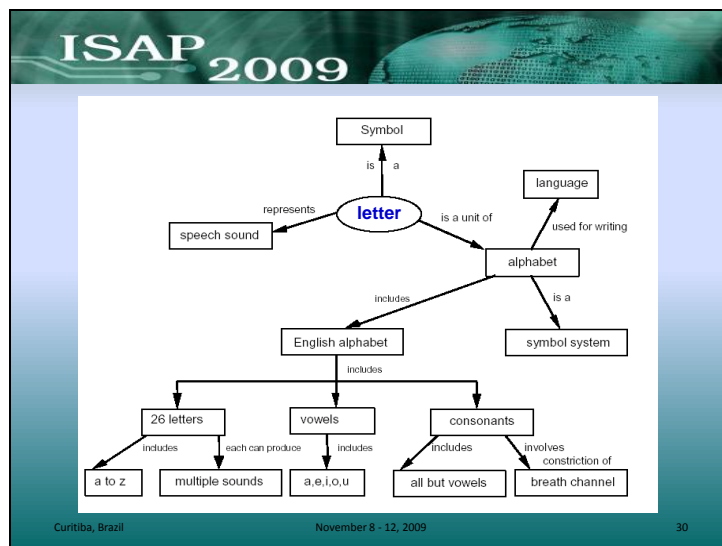
Quillian R. **Semantic Memory**. In M. Minsky (ed) **Semantic Processing**. MIT Press, Cambridge, MA, 1968.

Curitiba, Brazil

November 8 - 12, 2009

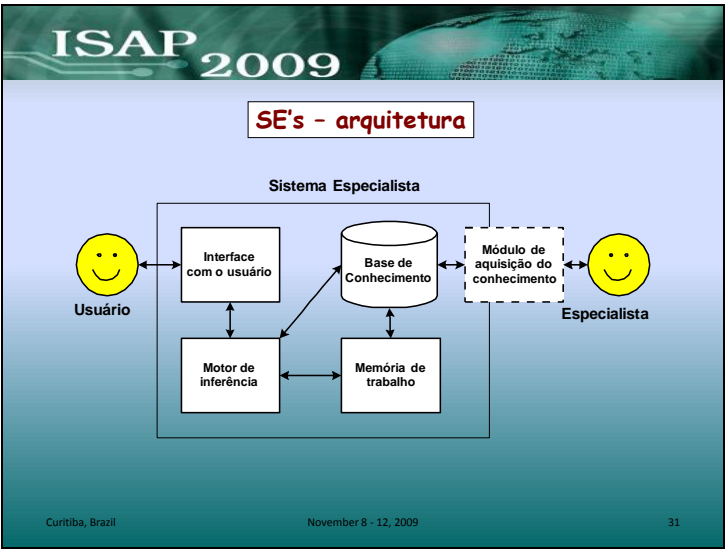
29

Slide 30

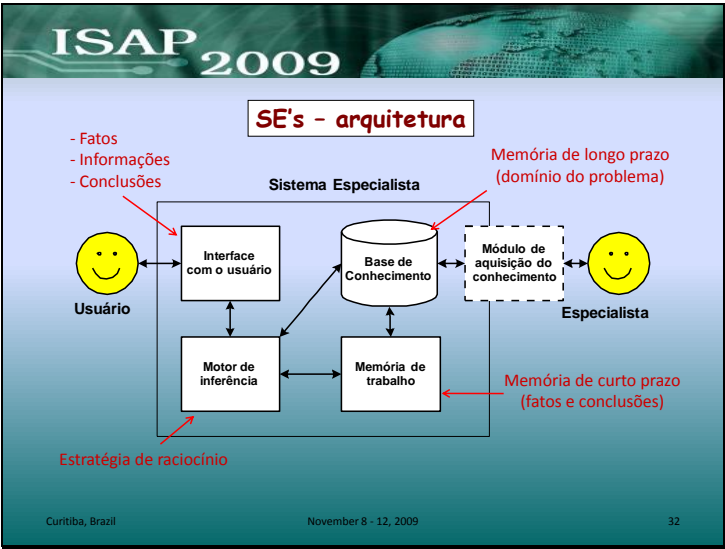




Slide 31



Slide 32



Slide 33

**ISAP 2009**

**Estratégia de inferência**

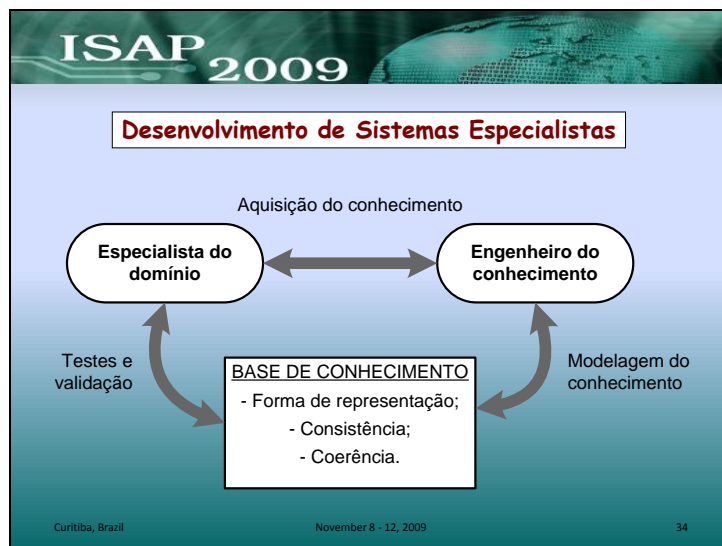
**Raciocínio para trás x Raciocínio para frente**

**Backward chaining x Forward chaining**

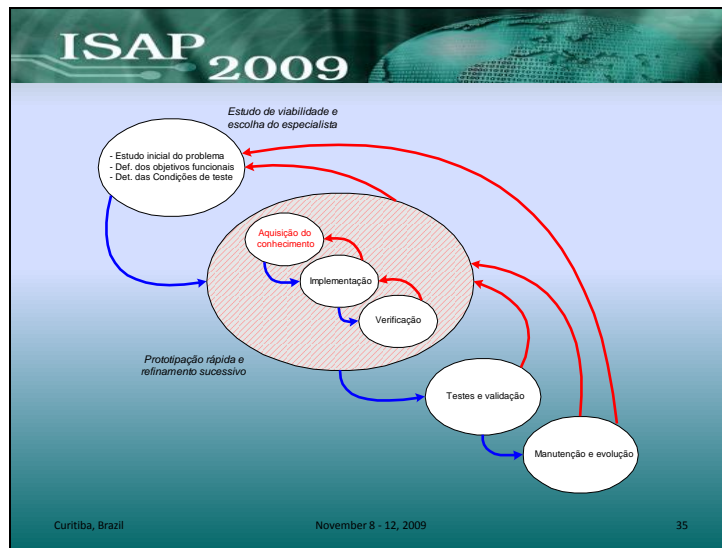
<p><u>encadeamento regressivo</u>, ou dirigido por hipóteses, ou para trás, onde o objetivo a ser alcançado é conhecido e o sistema tenta disparar somente as regras que podem alcançar o objetivo.</p> <p>Sistemas de diagnóstico</p>	<p><u>encadeamento progressivo</u>, ou dirigido por dados, ou para frente, onde o sistema dispara todas as regras aplicáveis a partir do conjunto inicial de fatos conhecidos;</p> <p>Sistemas de apoio a projetos</p>
--	--

Curitiba, Brazil November 8 - 12, 2009 33


Slide 34



Slide 35



Slide 36



**Estudo de caso - PETROBRAS**

**Sistemas Especialistas para monitoramento de processos e controle da deterioração de equipamentos**

**ST-Monitor** - Sistema especialista para o monitoramento e controle da corrosão em sistemas de topo de unidades de destilação de petróleo (REPAR, 1994/2000/2007).

Slide 37

ISAP 2009

### Descrição do problema

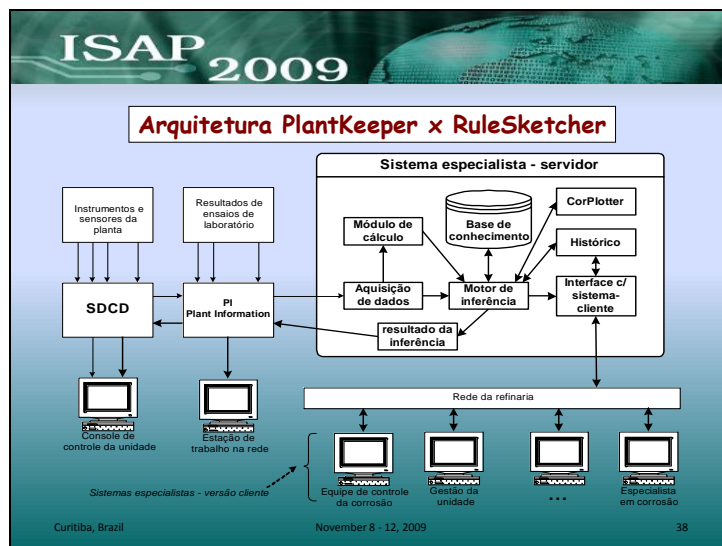
- Processo crítico, complexo e dinâmico;
- Requer a atenção de diferentes especialistas;
- Acúmulo considerável de informações;
- Solução depende fortemente da heurística de especialistas humanos e da disponibilidade de informações do processo.

Solução proposta:

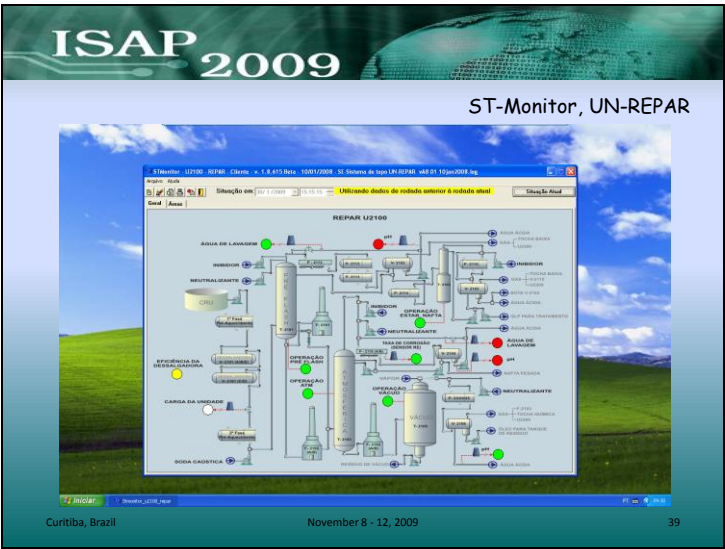
- ✓ Sistema especialista construído a partir do conhecimento dos especialistas da Petrobras, realizando análise periódica dos dados da planta, diagnosticando problemas relativos à corrosão e sugerindo medidas corretivas.
- ✓ Arquitetura cliente/servidor de monitoramento constante e contínuo.

Curitiba, Brazil November 8 - 12, 2009 37

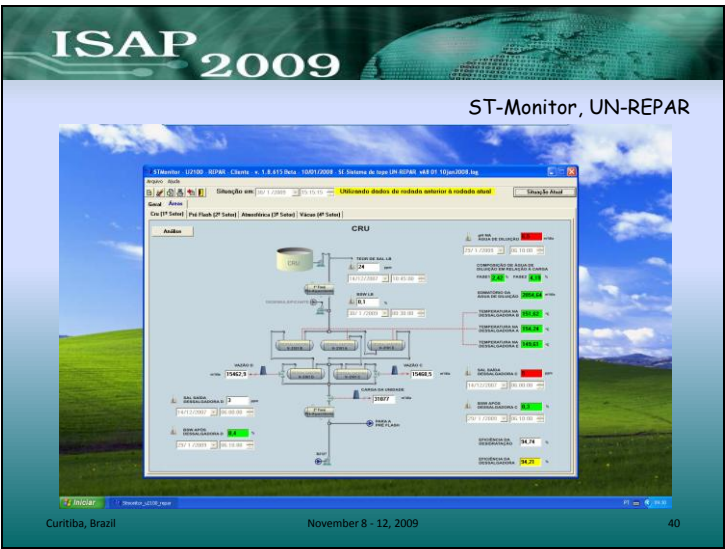
Slide 38



Slide 39



Slide 40



Slide 41

Slide 42

The screenshot displays the ST-Monitor software interface, which is used for monitoring and controlling industrial processes. The main window shows a detailed process flow diagram (PFD) of a refrigeration system. The diagram includes a compressor, condenser, evaporator, and expansion valve, connected by pipes. Various data points are displayed, such as pressures, temperatures, and flow rates. The interface is in Portuguese and includes a menu bar, toolbars, and a status bar.

Key elements visible in the interface include:

- Menu Bar:** File, Edit, View, Tools, Help.
- Toolbars:** Standard Windows toolbars and a specialized toolbar for the software.
- Status Bar:** Displays the current file name and version.
- Process Flow Diagram (PFD):** Shows the refrigeration cycle with various components and data points.
  - Compressor:** Labeled "VÁZÃO DO REFRIGERANTE SUZERTE" with a value of 255.18 kg/h.
  - Condenser:** Labeled "VÁZÃO DO REFRIGERANTE" with a value of 242.14 kg/h.
  - Evaporator:** Labeled "VÁZÃO DO REFRIGERANTE FIM" with a value of 246.19 kg/h.
  - Expansion Valve:** Labeled "VÁZÃO DO REFRIGERANTE INÍCIO" with a value of 246.19 kg/h.
  - Other components:** "OPERAÇÃO FIM", "DA PRÉ FLAM", "PARA O VÁCUO", "RENDIMENTO", "TEMPO DE VIDA RESTANTE (ANOS)", "QUALIDADE DE INSTALAÇÃO".
- Data Table:** A table on the right side of the interface provides detailed data for various parameters.
 

DATA	VALOR	UNIDADE
30/01/2009 15:15:13	249.17	
30/01/2009 15:00:02	244.75	
30/01/2009 14:00:00	252.34	
30/01/2009 13:00:17	251.6	
30/01/2009 12:00:12	250.7	

Slide 43

Slide 44

# ISAP 2009

## ST-Monitor, UN-REPAR

Curitiba, Brazil

November 8 - 12, 2009

44

Slide 45

# ISAP 2009

## Parceria TECPAR / Petrobras

---

**SES** – Sistema especialista em soldagem, para apoio à decisão na qualificação de procedimentos.  
(1ª versão 1990-1991; 2ª versão 1995-1997) – TECPAR e PETROBRAS/REPAR.

---

**MONITOR** – Sistema especialista para o monitoramento e controle da corrosão em unidades de destilação do petróleo.  
(1ª versão 1992-1993; 2ª versão 1998-2000) – TECPAR e PETROBRAS/REPAR.

---

**FccMonitor** – Sistema especialista para o monitoramento e controle da corrosão em unidades de craqueamento catalítico fluido (UCCF).  
(1ª versão 1999-2000; 2ª versão 2001-2002) – TECPAR, Metaldata e PETROBRAS/RLAM.

---

**AmineX** - Sistema especialista para o monitoramento e controle da corrosão em unidades de tratamento de gases.  
(1ª versão 1999-2000; 2ª versão 2001-2002) – TECPAR, Metaldata e PETROBRAS/RLAM.

---

**InjeX** - Sistema especialista para controle da qualidade da água de injeção em plataformas "offshore".  
(1ª versão 2001- previsão Dezembro 2002) – TECPAR, Metaldata e PETROBRAS/CENPES.

Curitiba, Brazil November 8 - 12, 2009 45

Slide 46

# ISAP 2009



**Dr. Eng. Milton Pires Ramos**  
*milton.ramos@tecpa.br*

DIA – Divisão de Inteligência Artificial  
TECPAR - Instituto de Tecnologia do Paraná  
Curitiba Paraná Brasil

Curitiba, Brazil November 8 - 12, 2009 46



Slide 47

ISAP 2009

## TECPAR - Instituto de Tecnologia do Paraná

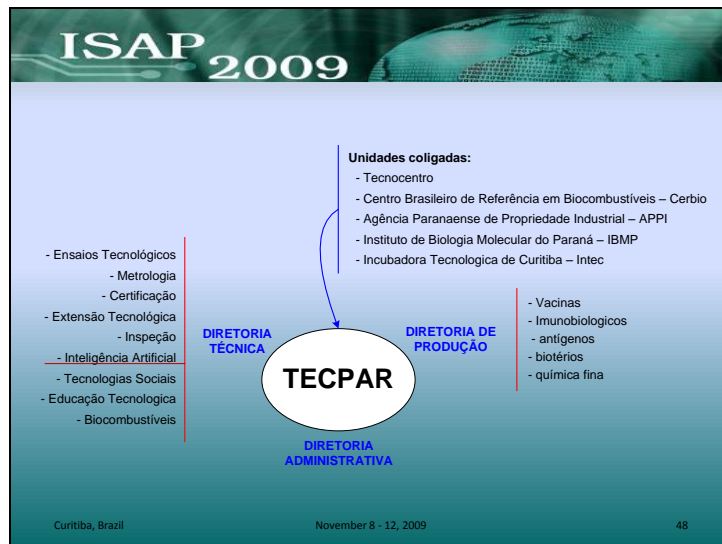


Empresa pública vinculada à Secretaria de Estado da Ciência, Tecnologia e Ensino Superior. É uma instituição de pesquisa, desenvolvimento, produção e prestação de serviços. ( est. 1940).

**Missão:**  
**"CONTRIBUIR COM SOLUÇÕES INOVADORAS PARA O PROGRESSO TÉCNICO E MELHORIA DA QUALIDADE DE VIDA DOS BRASILEIROS".**

Curitiba, Brazil
November 8 - 12, 2009
47

Slide 48



Slide 49

ISAP 2009

### Divisão de Inteligência Artificial



**Objetivo:**  
*"Desenvolver projetos de P&D em Inteligência Artificial, orientados para aplicações industriais, tecnológicas e científicas complexas".*

Curitiba, Brazil November 8 - 12, 2009 49

Slide 50

ISAP 2009

### Eixos de P&D

- Engenharia do Conhecimento;
- Aprendizado automático;
- Inteligência Artificial Distribuída / Sistemas Multi-Agente;
- Sistemas Colaborativos;
- Inteligência Artificial aplicada;
- Tecnologias de apoio à pesquisa em Inteligência Artificial.

Curitiba, Brazil November 8 - 12, 2009 50

# Redes Neurais Artificiais, Computação Evolucionária e Mineração de Dados

Prof. Júlio Cesar Nievola  
PPGla – PUCPR

[www.ppgia.pucpr.br/~nievola](http://www.ppgia.pucpr.br/~nievola)

nievola@ppgia.pucpr.br

## Agenda

- Redes Neurais Artificiais
- Computação Evolucionária
- Mineração de Dados
- Técnicas baseadas em Exames
- Procedimentos de Avaliação
- Referências

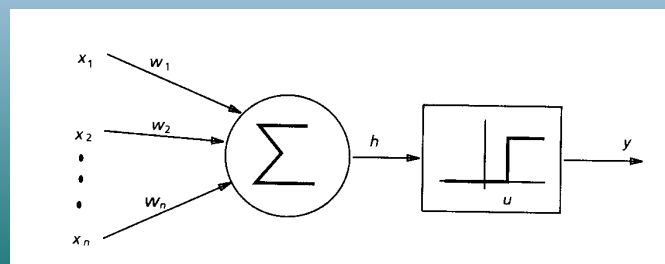
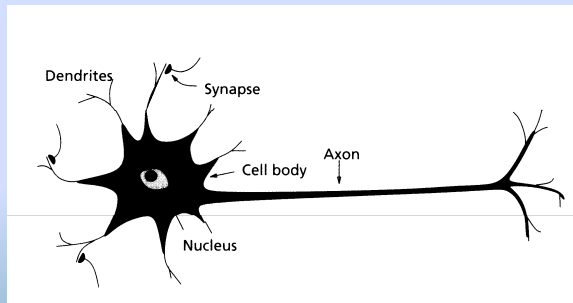
# Redes Neurais Artificiais



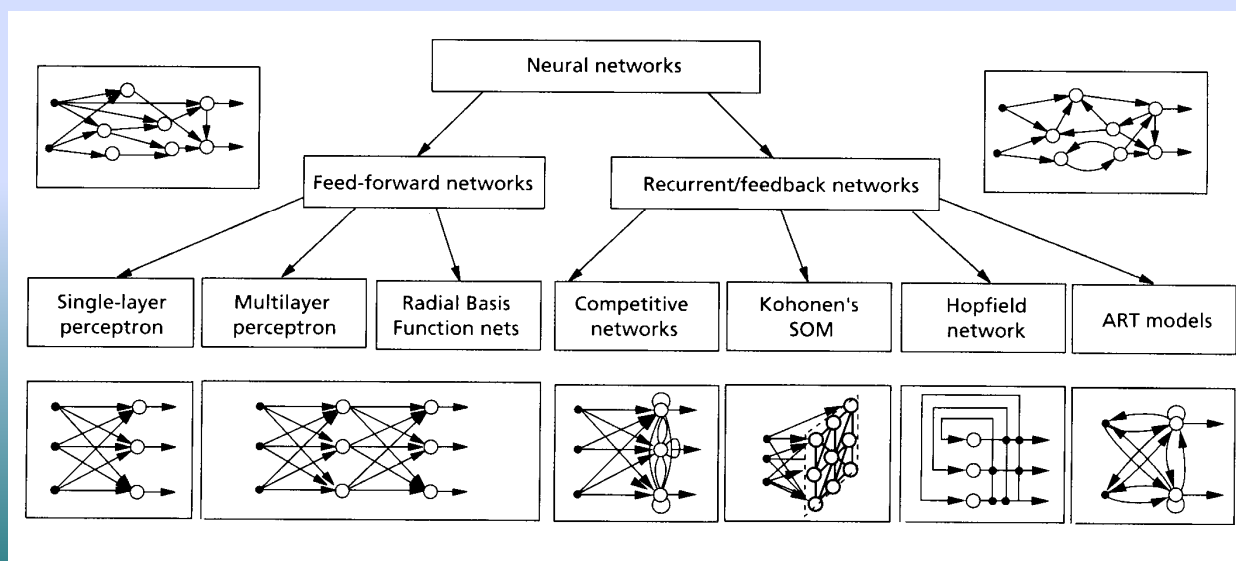
## Histórico de RNAs

- Reunião no Dartmouth College
- Paradigmas básicos:
  - Simbólico
  - Conexionista
- Perceptron (uma camada de pesos ajustáveis)
- Descrédito a partir do final da década 60
- Impulso a partir da década de 80
- Sistemas Híbridos

# Célula Biológica e Modelo



# Taxonomia



## Elementos de uma RNA

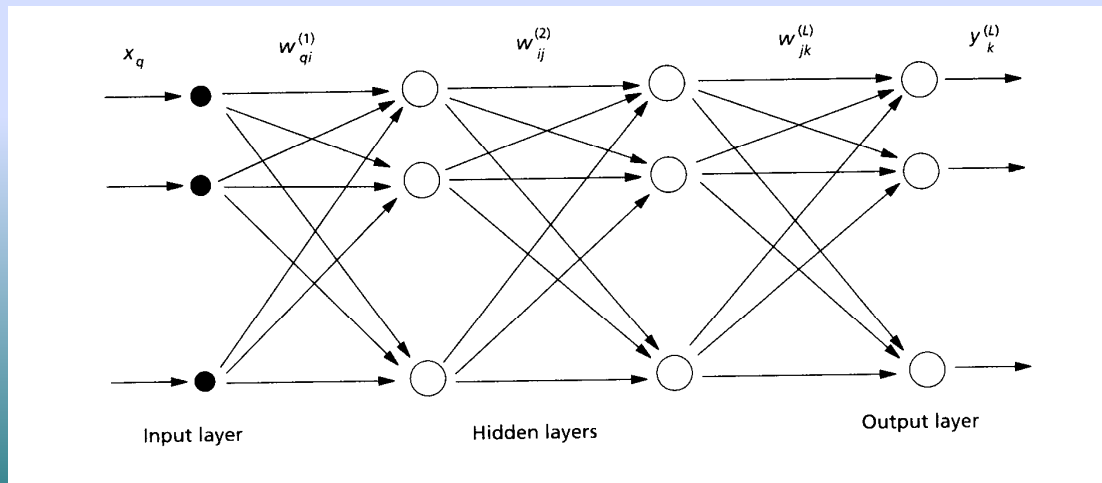
- Topologia organizada (geometria) de elementos de processamento interconectados
- Método de codificação da informação
- Método de aprendizagem (atualização dos pesos entre neurônios)
- Método de recuperação da informação

## Metodologia

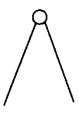
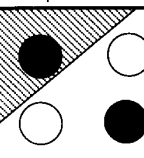
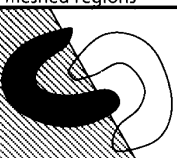
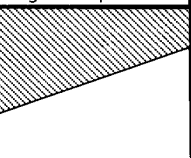
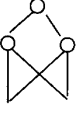
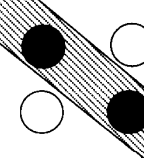
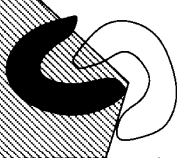
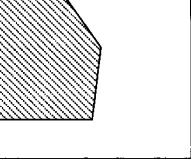

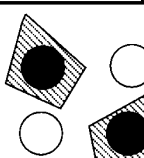
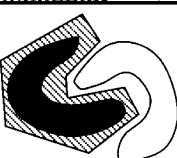
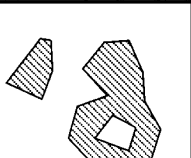
- Definir o problema
- Escolher informação
  - Obter dados
  - Criar arquivos rede
- Treinar a rede
- Testar a rede
- Uso em campo



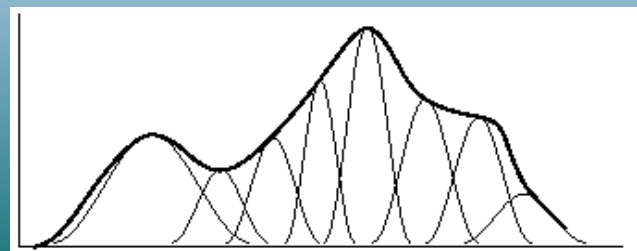
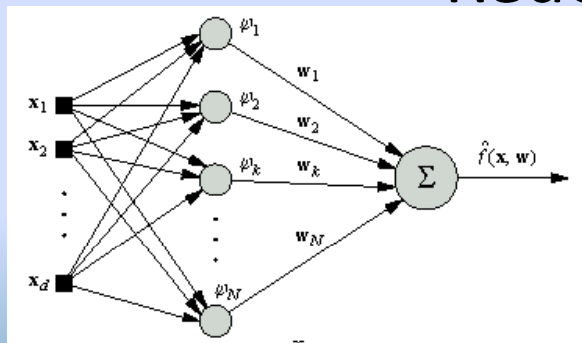
# Perceptron Multi-Camadas – MLP



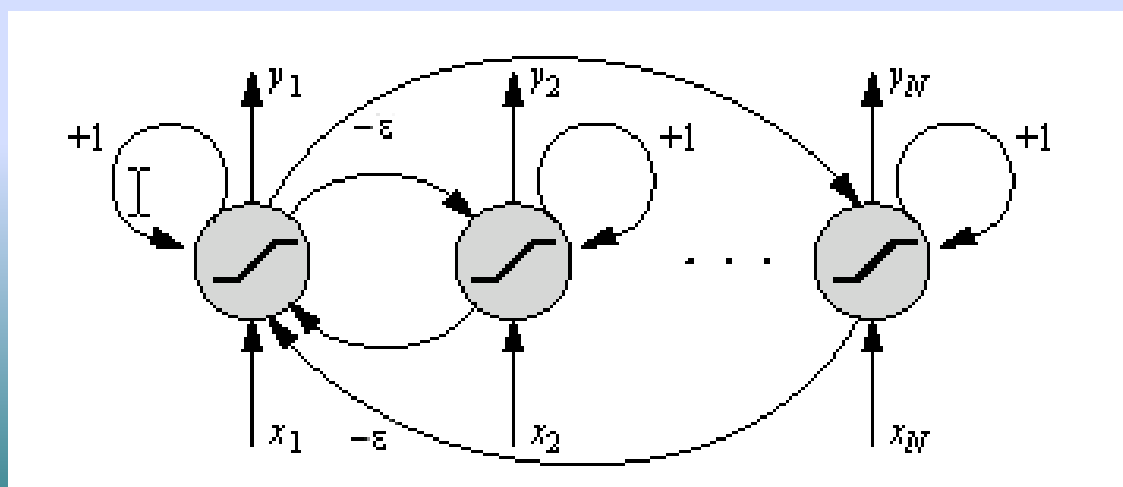
## MLP com retro-propagação

Structure	Description of decision regions	Exclusive-OR problem	Classes with meshed regions	General region shapes
 Single layer	Half plane bounded by hyperplane			
 Two layer	Arbitrary (complexity limited by number of hidden units)			
 Three layer	Arbitrary (complexity limited by number of hidden units)			

## Redes RBF

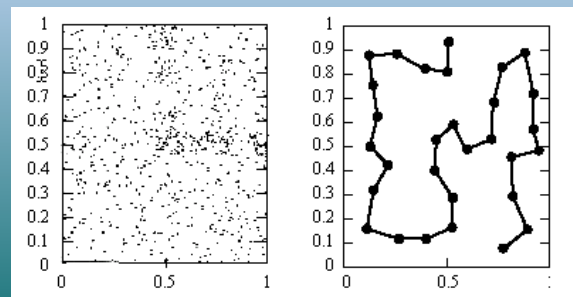
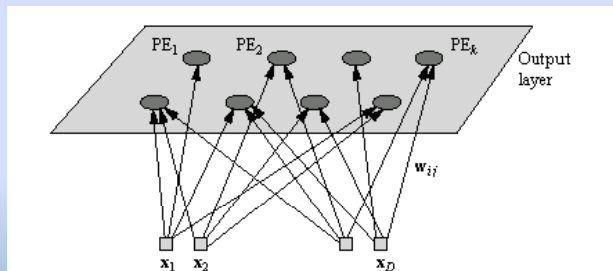


## Rede Competitiva

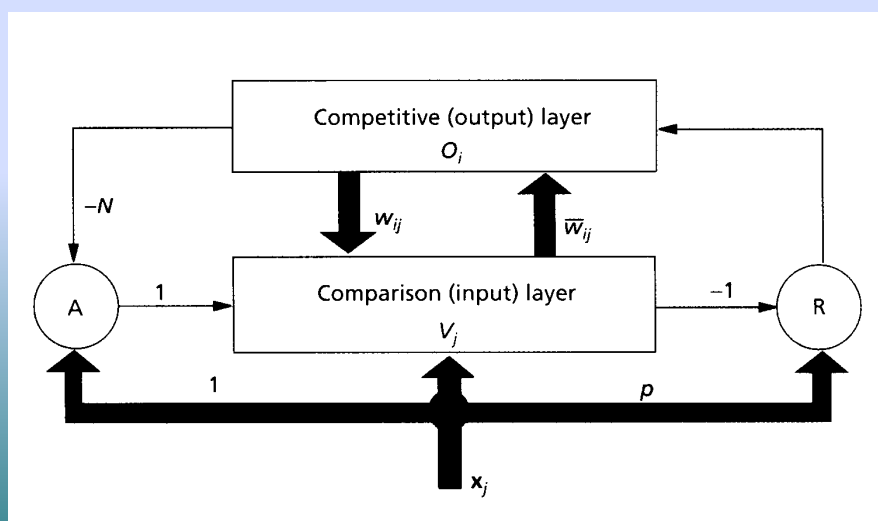




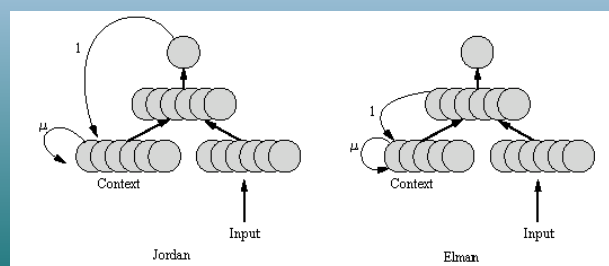
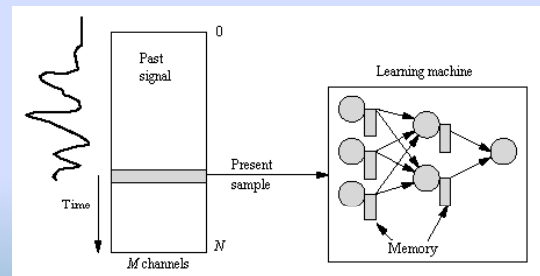
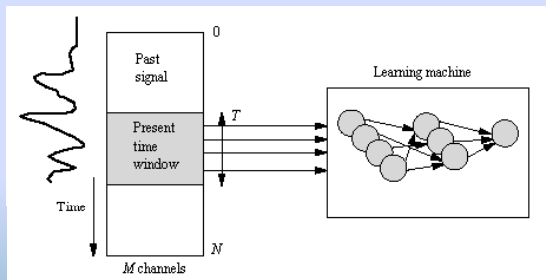
# Rede SOM (ou de Kohonen)



# Rede ART



# Redes Temporais



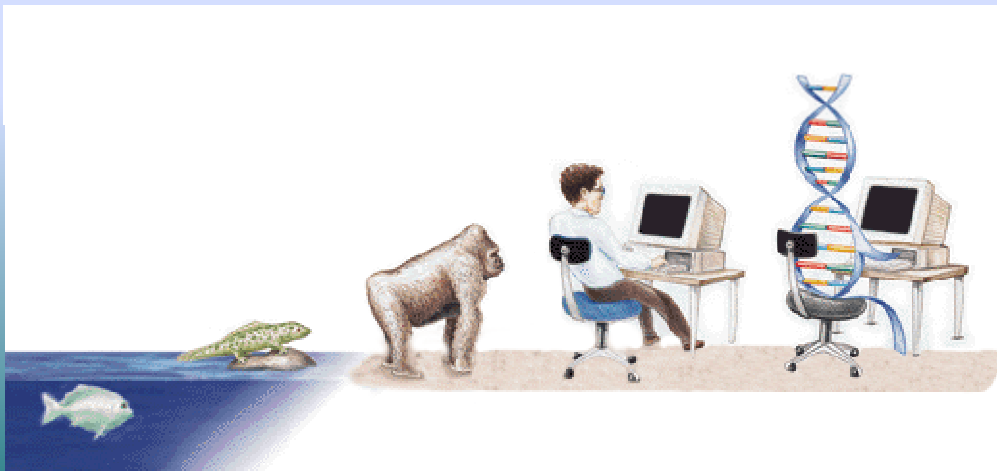
# Implementações de RNAs

- Características-chave:
  - Computacionalmente intensivas
  - Massivamente paralelas
  - Grandes requisitos de memória
- Possibilidades de implementação
  - Computadores convencionais
  - Computadores dedicados
  - Implementação em hardware específico

## Aplicações

- Boas características
  - Regras de resolução do problema desconhecidas ou difíceis de formalizar
  - Dispõe-se de um grande conjunto de exemplos e suas soluções
  - Necessita-se de grande rapidez na resolução do problema (p.ex. tempo real)
  - Não existem soluções tecnológicas atuais
- Exemplos
  - Reconhecimento de formas
  - Tratamento de sinal
  - Visão, fala
  - Previsão e modelagem
  - Auxílio à decisão
  - Robótica

## Computação Evolucionária



## Características

- Considera a Teoria da Evolução Darwiniana como um processo adaptativo de otimização
- Usa populações de estruturas computacionais que evoluem
- Busca uma melhora na “adequabilidade” da população com respeito ao ambiente

## Paradigmas de Computação Evolucionária

- Algoritmos Genéticos
- Programação Evolucionária
- Estratégias Evolucionárias
- Programação Genética
- Sistemas Classificadores

## Problema de Otimização

- Dada uma função  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  e um espaço de busca  $S \subseteq \mathbb{R}^n$ :

$$\text{maximizar } f(x) | x \in S$$

- Em termos de problemas de busca, o problema pode ser escrito como:

$$\text{encontrar } x^* | f(x^*) \geq f(x), \forall x \in S$$

## Métodos de Otimização

- Probabilísticos: usam a idéia da busca probabilística, não sendo baseados totalmente em sorte como métodos aleatórios
- Numéricos
  - Analíticos: quando a função é explicitamente conhecida e derivável
  - Cálculo Numérico: usando Pesquisa Operacional ou técnicas de gradiente ou estatística de alta ordem
- Enumerativos: examinam os pontos do espaço de busca, um a um, em busca dos pontos ótimos



## Definição de Algoritmos Genéticos

Algoritmos Genéticos (AGs) são métodos computacionais de busca baseados nos mecanismos de evolução natural e na genética. Em AGs uma população de possíveis soluções para o problema em questão evolui de acordo com operadores probabilísticos concebidos a partir de metáforas biológicas de modo que há uma tendência de que, na média, os indivíduos representem soluções cada vez melhores à medida que o processo evolutivo continua.

## Características Primárias de AGs

- Operam em uma população de pontos
- Operam em um espaço de soluções codificadas
- Necessitam somente de informação sobre o valor de uma função objetivo para cada membro da população
- Usam transições probabilísticas

## Representação Cromossômica

- Cada possível solução no espaço de busca é representada por uma seqüência de símbolos  $s$  gerados a partir de um alfabeto (e.g. binário)
- Cada seqüência  $s$  corresponde a um *cromossomo*
- Cada elemento de  $s$  é equivalente a um *gene*
- *Indivíduo*  $\equiv$  *cromossomo*

## Fluxo Básico de um Algoritmo Genético

```
begin
  t  $\leftarrow$  0
  inicializar P(t)
  avaliar P(t)
  while (not Condição_Terminal) do
    begin
      t  $\leftarrow$  t + 1
      selecionar P(t) a partir de P(t - 1)
      recombinar e mutar P(t)
      avaliar P(t)
    end
  end
```

## Inicialização

- População inicial de  $n$  indivíduos
- Geração aleatória ou por processo heurístico
- A população inicial deve cobrir a maior área do espaço de busca

## Avaliação e Adequabilidade

- AGs necessitam de uma função objetivo para cada indivíduo da população
- ***Fitness*** (“adequabilidade”) deve ser um valor não-negativo
- O fitness indica o quão bem adaptado ao ambiente um indivíduo está



## Seleção

- Emula os processos de reprodução assexuada e seleção natural
- Gera-se uma população temporária de  $N$  indivíduos extraídos com probabilidade proporcional ao fitness de cada um na população:

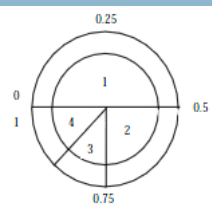
$$p_{sel} = \frac{a(s)}{\sum_{i=1}^N a(s_i)}$$

## Processo de Seleção

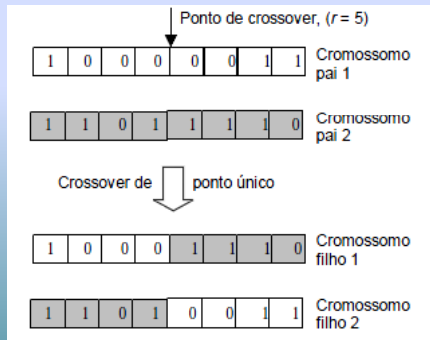
- Em uma população ela define os indivíduos que terão suas características repassadas para os indivíduos da próxima geração.
- Várias técnicas: roleta, torneio etc.

• Exemplo:

N	Cromossomo	Fitness	Graus
1	0001100101010	6.0	180
2	0101001010101	3.0	90
3	1011110100101	1.5	45
4	1010010101001	1.5	45

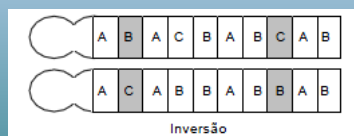
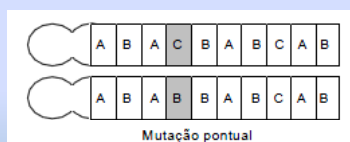


## Cruzamento ou recombinação



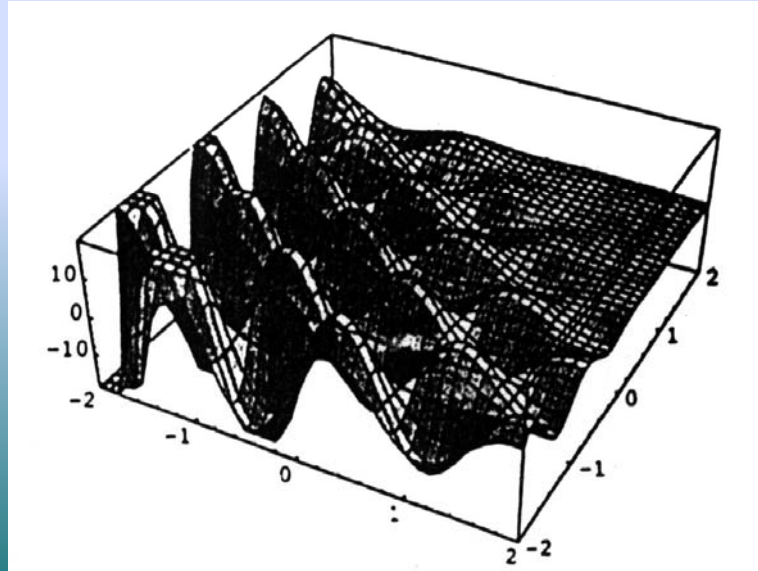
- Envolve a troca de fragmentos entre pares de cromossomos
- A forma mais simples é um processo aleatório com probabilidade fixa  $p_{rec}$

## Mutação



- Equivale à busca aleatória
- Seleciona-se uma posição num cromossomo e muda-se o valor do gene para outro valor possível, com probabilidade  $p_{mut}$

## Função Multimodal



## Condições de Término

- Ideal = ponto ótimo
- Prática = funções multimodais
- Critérios:
  - Número máximo de gerações
  - Tempo limite de processamento
  - Estagnação: quando não se observa melhorias na população após várias gerações consecutivas

## Áreas de Aplicação

- Problema do Caixeiro Viajante
- Planejamento de tarefas (“job-shop scheduling”)
- Síntese de Redes Neurais
- Análise de dados: agrupamento (“clustering”)
- Controle de processos: controle ótimo
- Robótica
- Modelos econômicos, ecológicos e sociais

## Programação Evolucionária

- Inicialmente usava máquinas de estados finitos.
- Meta-programação
  - Processo de auto-adaptação dos parâmetros
- Sem cruzamento, somente mutação

## Estratégias Evolucionárias

- Estratégia geral: algoritmo ( $\mu + \lambda$ ).
- Regra de sucesso 1/5.
- Operadores de cruzamento e mutação.
- Seleção ocorre depois da recombinação e mutação.
- Auto-adaptação dos parâmetros

## Programação Genética

- Cada cromossomo representa uma árvore computacional com tamanho variável
- Cada estrutura cromossômica representa um programa (e.g. em Lisp)
- Cada árvore equivale a um indivíduo e uma população de programas evolui com os Ags
- Realiza a geração automática de programas computacionais para resolução de programas



## Sistemas Classificadores

- São basicamente sistemas de regras de produção adaptativas
- Regras do tipo “if-then” evoluem de acordo com um AG
- Cada regra é um classificador
- Populações de classificadores são manipulados por AGs

## Mineração de Dados



## Histórico

- Década de 60: Coleções de dados, criação de BD
- Década de 70: Modelos de dados relacionais, implementação de DBMS relacionais
- Década de 80: RDBMS, modelos avançados de dados (relacional estendido, OO, dedutivo etc.) e DBMS orientados à aplicação (espaciais, científicos, de engenharia etc.).
- Década de 90: Mineração de dados e data warehousing, bases de dados multimídia, e tecnologia Web

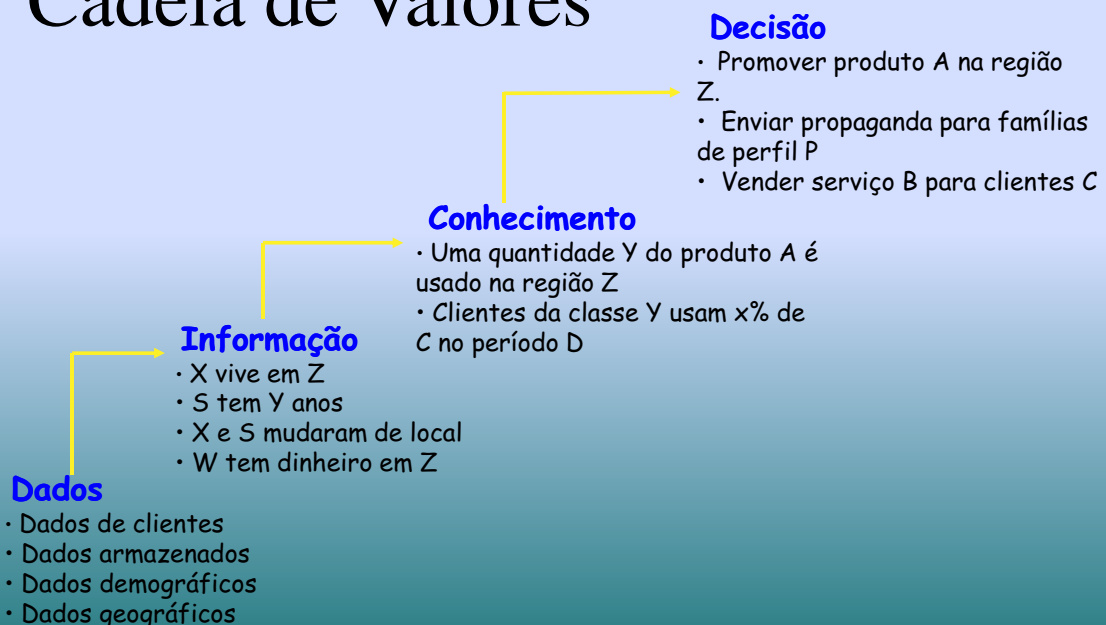
## Aprendizagem de Máquina

- *Aprendizagem de Máquina* – Melhoria no desempenho de alguma tarefa através da experiência
- *Data Mining* – Parte de um processo maior (KDD) interessado em:
  - Melhoria no desempenho
  - Representação inteligível
  - Conhecimento obtido interessante, inovador
- *We are drowning in information, but starving for knowledge!* (John Naisbett)

## Motivações para MD

- Abundância de dados industriais e comerciais
- Foco competitivo – Gerenciamento do conhecimento
- Computadores poderosos e baratos
- Fundamentos avançados em
  - Aprendizagem de máquina & lógica
  - Estatística
  - Sistemas de gerenciamento de BD

## Cadeia de Valores

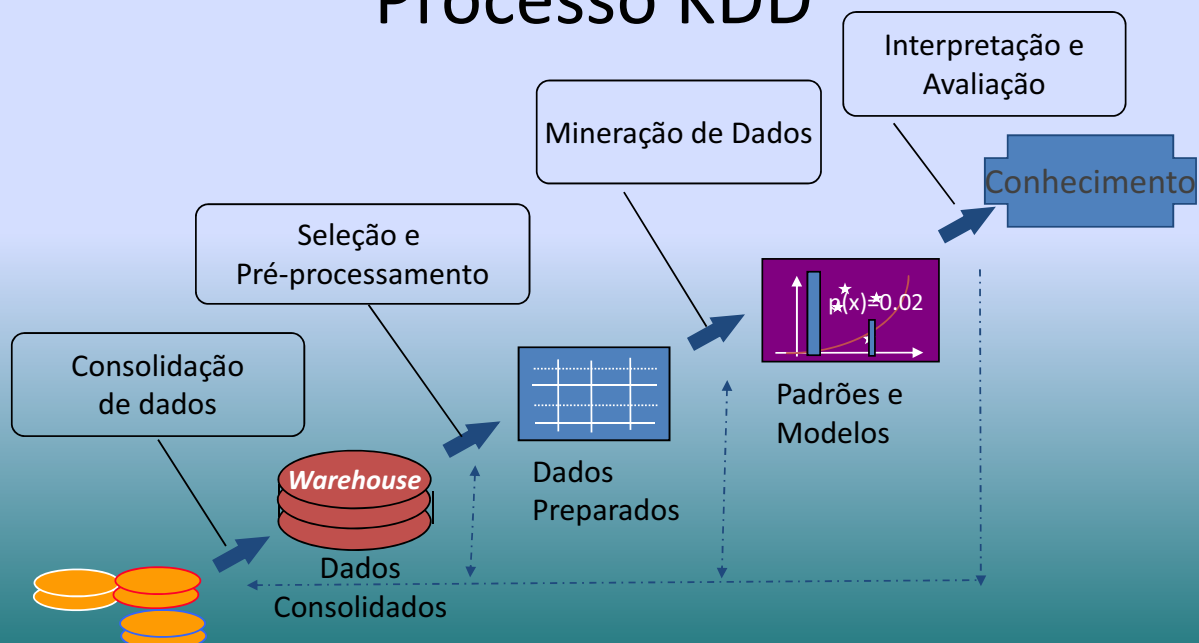




## KDD x MD

- KDD é a seleção e o processamento de dados para:
  - Identificar conhecimento novo, preciso e útil, &
  - Modelar fenômenos do mundo real
- Mineração de Dados é o principal componente do processo KDD – descoberta de conhecimento em BD

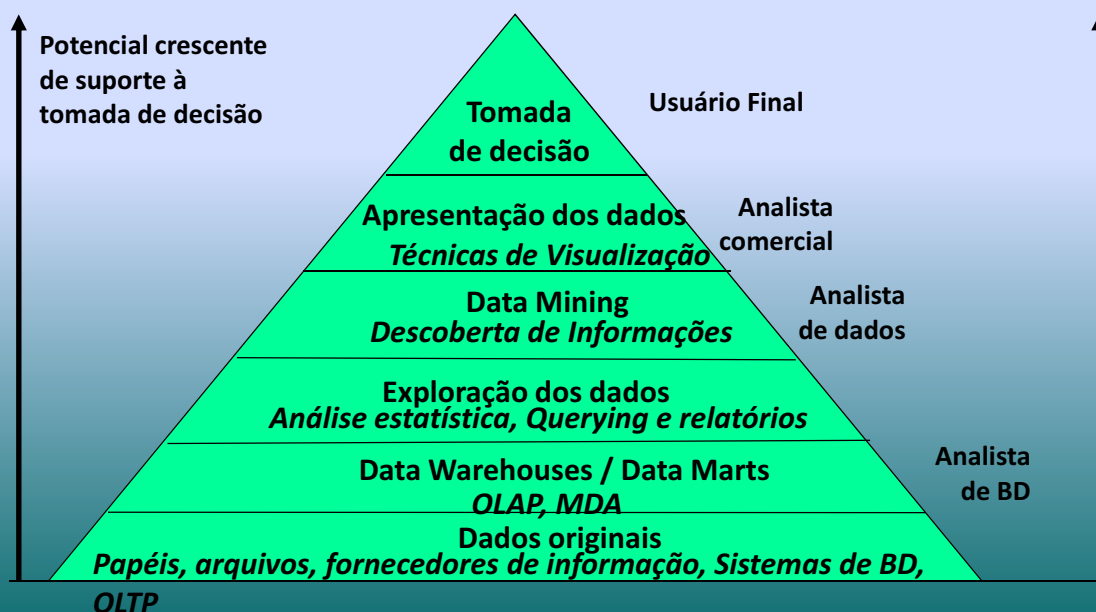
## Processo KDD



## Passos em MD

- Seleção e Pré-Processamento
  - Limpeza dos dados: (pode exigir 60% do tempo total)
  - Redução de dados:
    - Encontrar características úteis, redução de dimensionalidade e ou de variáveis
- Determinar a tarefa de MD
  - Classificação (ou regressão), associação, agrupamento
- Escolha do algoritmo
- Mineração de Dados: busca pelos padrões interessantes
- Interpretação e avaliação: análise dos resultados
  - Visualização, transformação, remoção de padrões redundantes
- Uso do conhecimento descoberto

## Mineração de Dados no Processo Decisório



## Descoberta de Regras de Associação

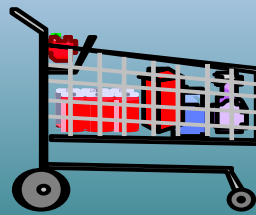
Também chamada de "Market Basket Analysis", surgiu com a idéia de determinar hábitos de compra de clientes, encontrando associações e correlações entre diferentes itens que um cliente coloca em seu "carrinho de compras".

Leite, ovos, açúcar,  
pão



Cliente 1

Leite, ovos, cereal, pão



Cliente 2

Ovos, açúcar



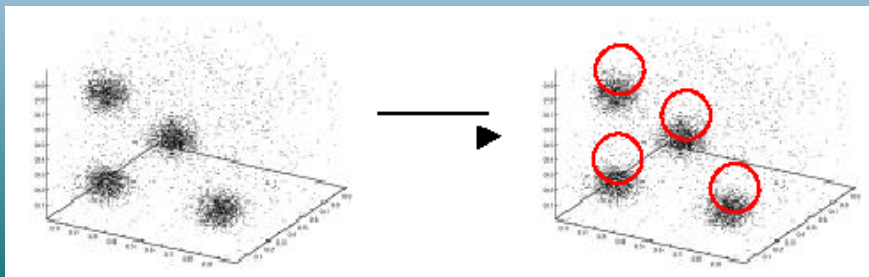
Cliente 3

## Objetivos da Descoberta de Regras de Associação

- Extrair informação sobre comportamento de compra
- Informação obtida pode sugerir
  - Novos leiautes de lojas
  - Novo conjunto de produtos
  - Quais produtos colocar em promoção
- MBA é aplicável onde um cliente compra vários itens em proximidade
  - Cartões de crédito
  - Serviços de companhias de telecomunicações
  - Serviços bancários
  - Tratamentos médicos

# Agrupamento

- Dado:
  - BD grande de dados de clientes, contendo suas propriedades e seu histórico.
- Objetivo:
  - Encontrar grupos com comportamento similar
  - Encontrar dados com comportamento não usual



# Agrupamento

- Dado:
  - Um conjunto de dados com  $N$  dados  $d$ -dimensionais
- Encontrar:
  - Uma partição natural do conjunto de dados em um número de grupos ( $k$ ) e ruído
  - Os grupos devem ser tais que
    - Itens em um mesmo grupo são similares, ou seja, similaridade intra-grupos é maximizada &
    - Itens de grupos diferentes são diferentes, ou seja, similaridade inter-grupos é minimizada

## Uso do agrupamento

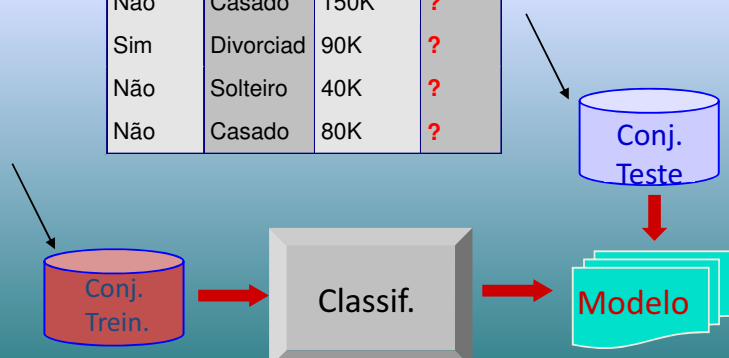
- Sem classes pré-definidas
- Usado como técnica individual para determinar distribuição de dados ou como etapa de pré-processamento para outros algoritmos
- Auxilia a compreender como objetos pertencentes a um conjunto de dados tendem a se agrupar naturalmente

## Classificação

ID	Res-sarce?	Estado civil	Rendi-mentos	Enga-na?
1	Sim	Solteiro	125K	Não
2	Não	Casado	100K	Não
3	Não	Solteiro	70K	Não
4	Sim	Casado	120K	Não
5	Não	Divorciad	95K	Sim
6	Não	Casado	60K	Não
7	Sim	Divorciad	220K	Não
8	Não	Solteiro	85K	Sim
9	Não	Casado	75K	Não
10	Não	Solteiro	90K	Sim

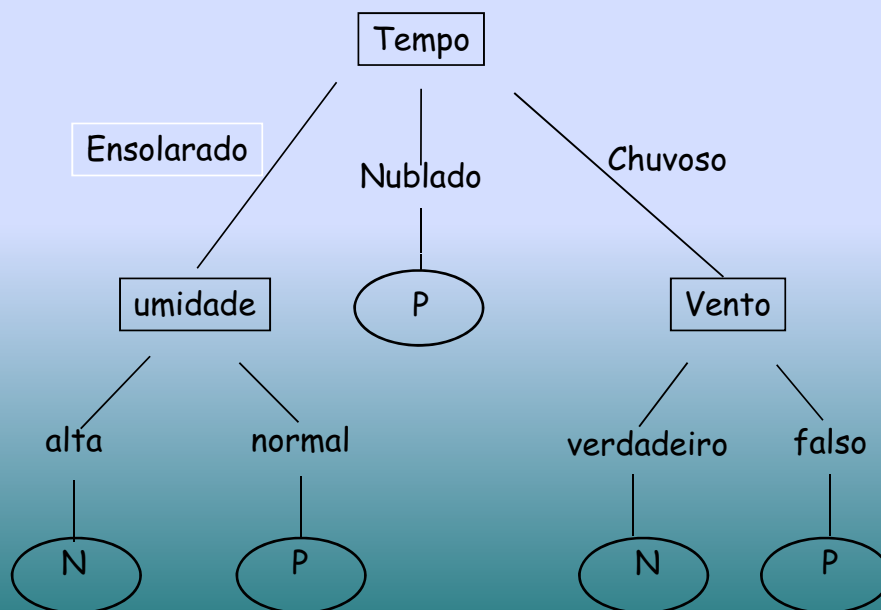
classe

Res-sarce?	Estado civil	Rendi-mentos	Enga-na?
Não	Solteiro	75K	?
Sim	Casado	50K	?
Não	Casado	150K	?
Sim	Divorciad	90K	?
Não	Solteiro	40K	?
Não	Casado	80K	?

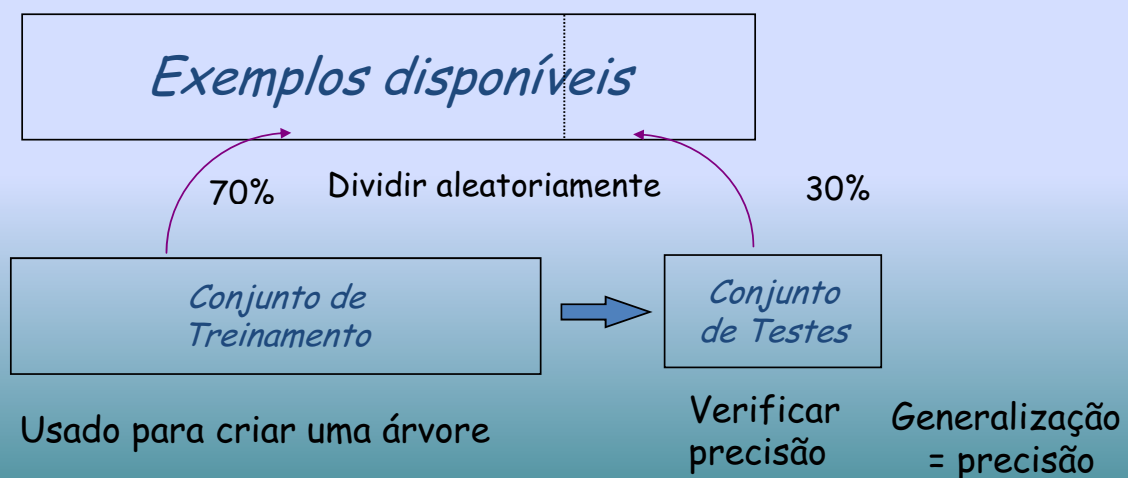




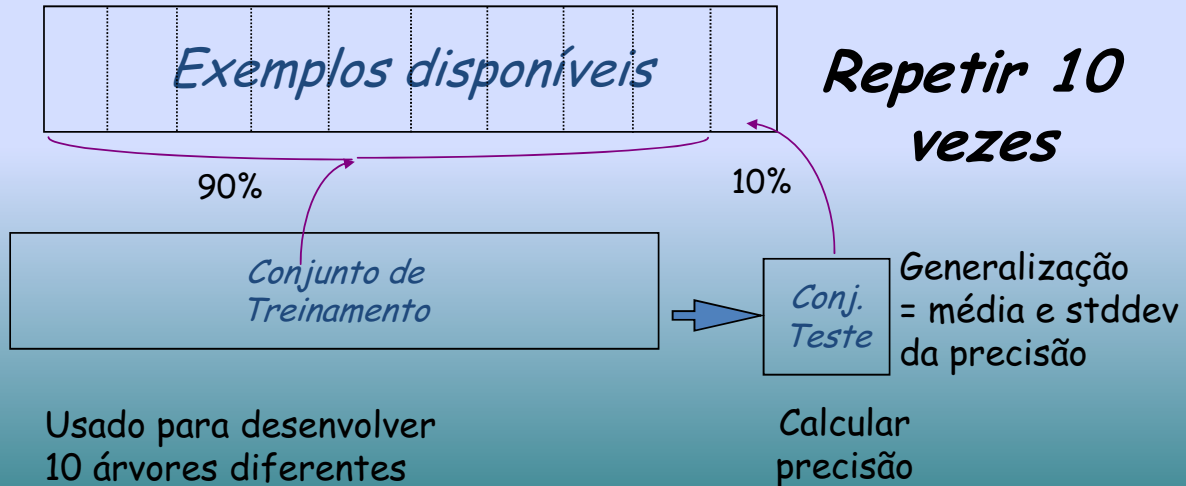
# Árvore de Decisão



## Avaliação - BD grande – Hold-Out



## Avaliação - BD pequena Validação Cruzada



## Referências – RNA

- Haykin, S., "Neural Networks – A Comprehensive Foundation", 2nd. Ed., Prentice Hall, 1999.
- Principe, J.C.; Euliano, N.R.; Lefebvre, W.C., "Neural and Adaptive Systems – Fundamentals through Simulations", John Wiley & Sons, 2000.
- <http://www.heatonresearch.com/encog/>
- <http://nn.cs.utexas.edu/>
- <http://www.inns.org/>
- <http://www.ip-atlas.com/pub/nap/nn-src/>
- <http://www.ra.cs.uni-tuebingen.de/SNNS/>
- <http://www.sbrn.org.br/>
- <http://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=72>

## Referências – CE

- Bäck, T.; Fogel, D.B.; Michalewicz, T., “Evolutionary Computation 1 – Basic Algorithms and Operators”, Institute of Physics Publishing, 2000.
- Bäck, T.; Fogel, D.B.; Michalewicz, T., “Evolutionary Computation 2 – Advanced Algorithms and Operators”, Institute of Physics Publishing, 2000.
- <http://cs.gmu.edu/~eclab/projects/ecj/>
- <http://www.gp-field-guide.org.uk/>
- <http://cs.gmu.edu/~sean/book/metaheuristics/>
- [http://en.wikipedia.org/wiki/Genetic\\_algorithm](http://en.wikipedia.org/wiki/Genetic_algorithm)
- <http://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=4235>
- <http://www.obitko.com/tutorials/genetic-algorithms/index.php>
- <http://www.talkorigins.org/faqs/origin.html>

## Referências – MD

- Tan, P.-N.; Steinbach, M.; Kumar, V., “Introduction to Data Mining”, Addison-Wesley, 2006.
- Witten, I.H.; Frank, E., “Data Mining – Practical Machine Learning Tools and Techniques”, Morgan Kaufmann Publishers, 2005.
- <http://www.crisp-dm.org/>
- <http://www.kdnuggets.com/>
- <http://www.cs.waikato.ac.nz/ml/weka/>
- <http://rapid-i.com/content/blogcategory/38/69/>
- <http://www.keel.es/>
- <http://www-stat.stanford.edu/~tibs/ElemStatLearn/>
- <http://www.sigkdd.org/explorations/issue.php?issue=current>