



Aprendizado de Máquina

Introdução

Thiago A. S. Pardo
Daniel Honorato
Solange O. Rezende
Ronaldo C. Prati

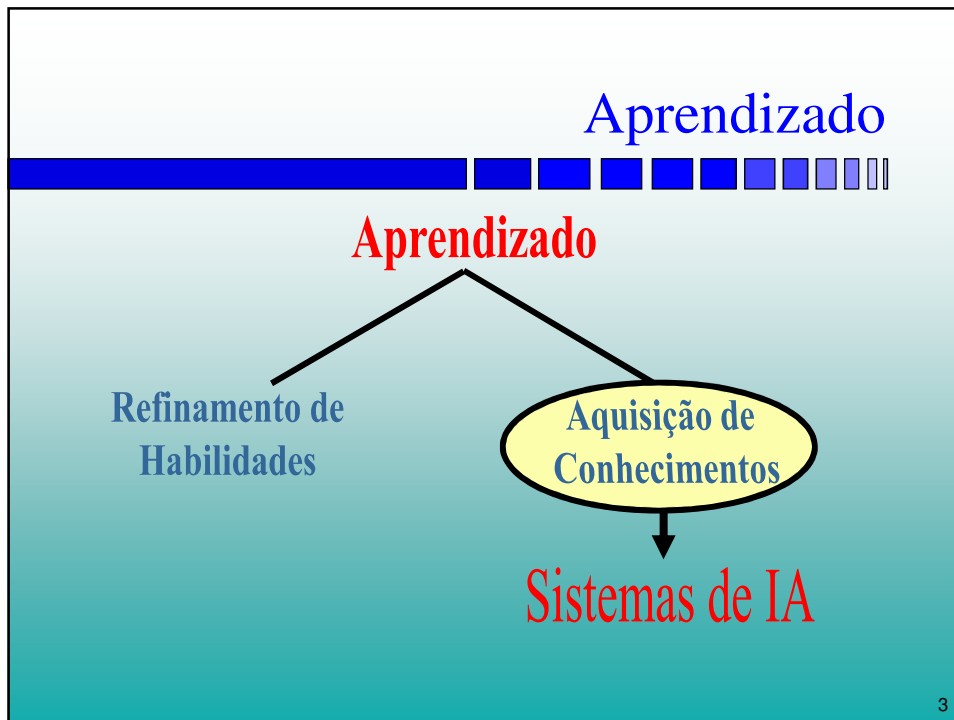
1



Inteligência vs. Aprendizado

- Aprendizado é a chave da superioridade da Inteligência Humana
- Para que uma máquina tenha **Comportamento Inteligente**, deve-se aumentar a **Capacidade de Aprendizado**

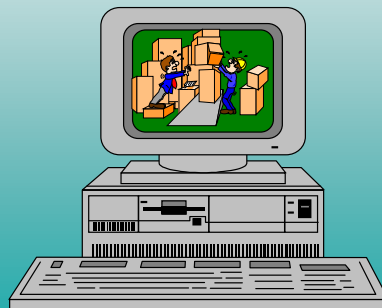
2



- ## Inteligência vs. Aprendizado
-
- O ser humano está **pré-programado para o aprendizado**; aprende ampliando o alcance do conhecimento que já possui, através de reordenações sucessivas
 - O computador não possui o programa inicial para procurar por informações e realizar aprendizado em geral
 - **Paradigmas e técnicas de AM** possuem um alvo bem mais limitado do que o aprendizado humano
- 4

Aprendizado de Máquina (AM)

Definição: é uma subárea de IA que pesquisa métodos computacionais relacionados à aquisição de novos conhecimentos, novas habilidades e novas formas de organizar o conhecimento já existente



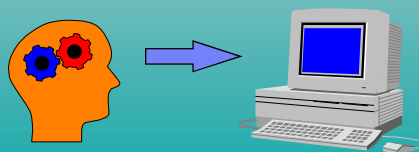
5

Objetivos de AM

- Um melhor entendimento dos mecanismos de aprendizado humano

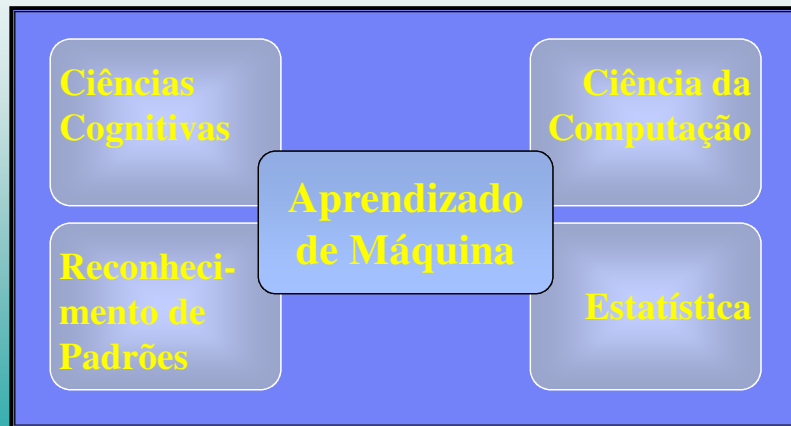


- Automação da aquisição do conhecimento



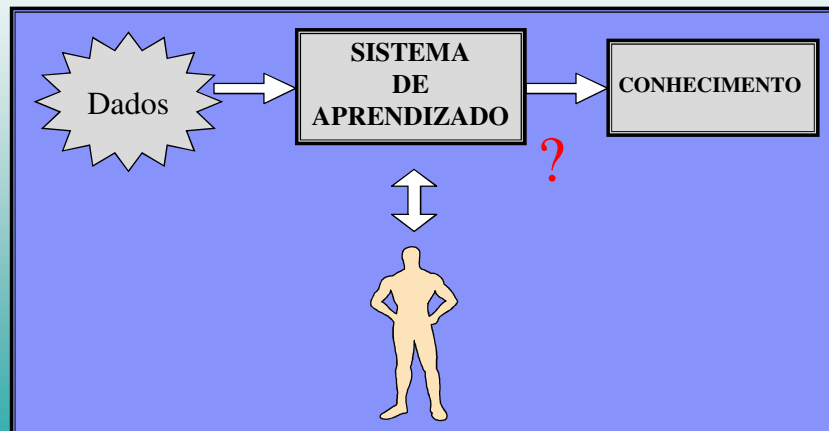
6

AM: multidisciplinar



7

Aprendizado



8

Paradigmas de AM

- Paradigmas de AM
 - Simbólico
 - Baseado em Exemplos
 - Estatístico
 - Conexionista
 - Evolutivo

9

Paradigmas de AM - Simbólico

- Explora representações de estruturas gráficas ou lógicas, no lugar de métodos estatísticos ou numéricos
- Descrições simbólicas representam um **conhecimento de alto nível**
- As representações simbólicas estão tipicamente na forma de alguma expressão lógica, árvore de decisão, regras de produção ou rede semântica

10

Paradigmas de AM – Baseado em Exemplos

- Exemplos de treinamento são armazenados
- Casos nunca vistos são classificados através de **casos similares** conhecidos
- Classificação de um caso é lembrar de um caso similar cuja classe é conhecida e assumir que o novo caso terá a mesma classe

11

Paradigmas de AM - Estatístico

- Decisões tomadas através de raciocínio sobre **probabilidades dos dados**
- Como regra geral, técnicas estatísticas tendem a focar tarefas em que todos os atributos têm valores contínuos ou ordinais
- Muitas técnicas são paramétricas, assumindo alguma forma de modelo, e então encontrando valores apropriados para os parâmetros do modelo a partir de dados

12

Paradigmas de AM - Conexionista

- Estudo de Redes Neurais Artificiais foi inspirado em parte na observação de que sistemas de aprendizado biológico são compostos de redes muito complexas de neurônios interconectados
- Redes Neurais Artificiais são redes construídas a partir de **conjuntos de unidades simples altamente interconectadas**, daí o nome *conexionismo*

13

Paradigmas de AM - Evolutivo

- Este paradigma possui uma analogia direta com a teoria de Darwin, onde **sobrevivem os mais bem adaptados** ao ambiente
- Um classificador evolutivo consiste em uma população de elementos de classificação que competem para fazer a predição; elementos que possuem uma performance fraca são descartados, enquanto os elementos mais fortes proliferam, produzindo variações de si mesmos

14

Qual o paradigma?

Dia	Tempo	Temperatura	Umidade	Vento	Jogou tênis?
1	Sol	Quente	Alta	Fraco	Não
2	Sol	Quente	Alta	Forte	Não
3	Nublado	Quente	Alta	Fraco	Sim
4	Chuva	Mediana	Alta	Fraco	Sim
5	Chuva	Frio	Normal	Fraco	Sim
6	Chuva	Frio	Normal	Forte	Não
7	Nublado	Frio	Normal	Forte	Sim
8	Sol	Mediana	Alta	Fraco	Não
9	Sol	Frio	Normal	Fraco	Sim
10	Chuva	Mediana	Normal	Fraco	Sim
11	Sol	Mediana	Normal	Forte	Sim
12	Nublado	Mediana	Alta	Forte	Sim
13	Nublado	Quente	Normal	Fraco	Sim
14	Chuva	Mediana	Alta	Forte	Não

15

Estratégias de AM

Aprendizado por Hábito

Aprendizado por Instrução

Aprendizado por Dedução

Aprendizado por Analogia

Aprendizado por Indução

16

Aprendizado por Indução

- **Inferência Indutiva** é um dos principais meios para a aquisição de novos conhecimentos e previsão de eventos futuros
- **Observações permitem descobrir regras e procedimentos**
- Deve-se ter cuidado com o número de observações e a relevância dos dados

17

Argumentos Dedutivos vs. Indutivos

- **Argumento dedutivo**
Nenhum dos alunos gosta de Inteligência Artificial
Francisco é um aluno
⇒ *Francisco não gosta de Inteligência Artificial*
- **Argumento indutivo**
Nenhum dos alunos que foram entrevistados gosta de Inteligência Artificial
⇒ *Nenhum aluno gosta de Inteligência Artificial*

18

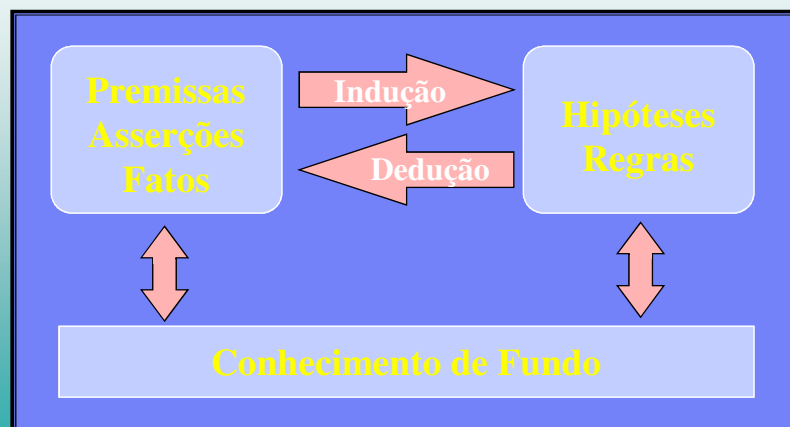
Argumentos Dedutivos vs Indutivos

Argumentos Dedutivos	Argumentos Indutivos
Se todas as premissas são verdadeiras, a conclusão é verdadeira	Se todas as premissas são verdadeiras, a conclusão é <u>provavelmente</u> verdadeira, mas não <u>necessariamente</u> verdadeira (à exceção dos argumentos matemáticos indutivos).
Toda a informação do conteúdo fatural da conclusão já está, pelo menos implicitamente, nas premissas.	A conclusão contém informação que não está nem implicitamente nas premissas.

Argumentos indutivos preservam falsidade

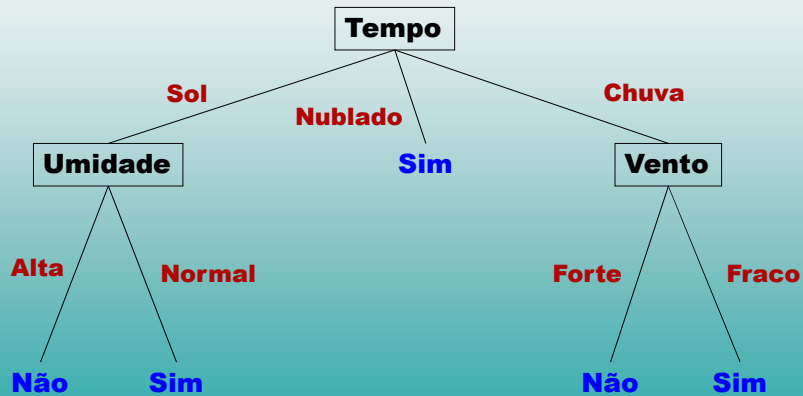
19

Relação entre Dedução e Indução



20

Dedução ou Indução?



21

Aprendizado por Indução

■ Dependendo dos fatos necessários para realizar o aprendizado fornecidos por uma fonte externa ou observados pelo sistema aprendiz, pode-se distinguir dois tipos diferentes de estratégias de aprendizado

- Aprendizado por **observação e descoberta**
- Aprendizado por **exemplos**

22

Aprendizado por Indução (Cont)

1. Aprendizado por observação e descoberta



2. Aprendizado por exemplos

23

Aprendizado por Indução (Cont)

1. Aprendizado por observação e descoberta



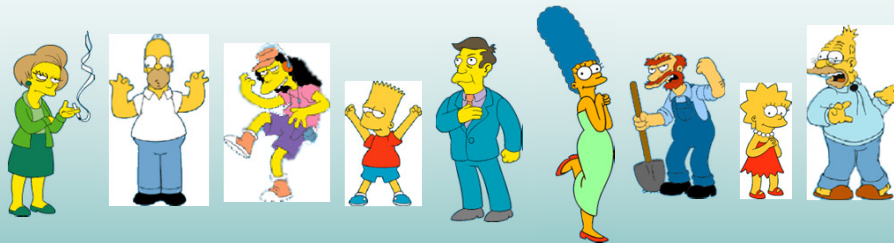
- Não existe o professor
- O aprendiz analisa entidades fornecidas ou observadas e tenta determinar se alguns subconjuntos dessas entidades podem ser agrupados em certas classes de maneira útil

→ **Aprendizado Não-Supervisionado**

24

Agrupamento

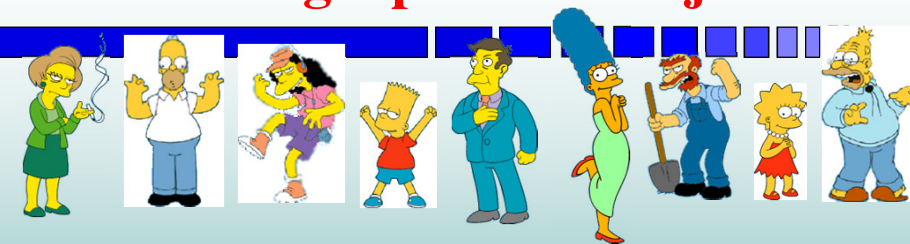
Como agrupar estes objetos?



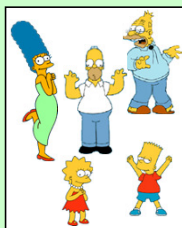
25

Agrupamento

Como agrupar estes objetos?



o agrupamento é subjetivo



Os Simpsons



Empregados da Escola



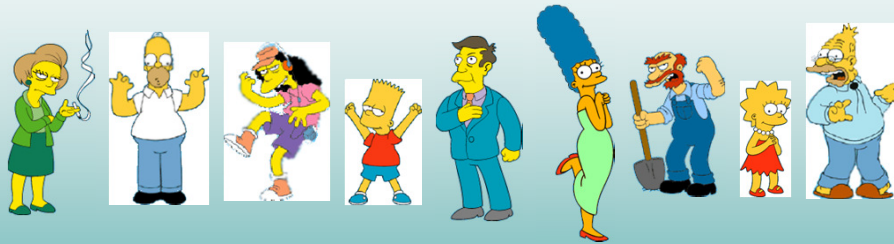
Mulheres



Homens

Agrupamento

Como agrupar estes objetos?



Qual seu critério de agrupamento?

- Similaridade em algum nível

27

O que é similaridade?

A qualidade, caráter ou condição das coisas similares.

(Dicionário Houaiss)

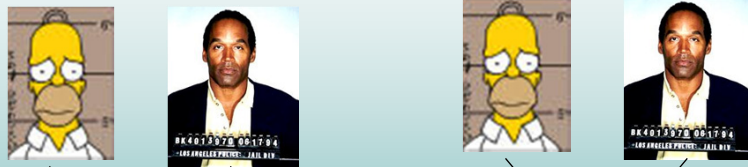


Similaridade é difícil de definir, mas...
Reconhecemos quando a vemos!

28

Como medir similaridade?

■ Função de distância ou similaridade



$$x^2 + y^2 + 2dx + 2ey + f = 0$$

$$a = \pi r^2$$

características = cor da
pele, altura
distância = 0.2

$$x^2 + y^2 + 2dx + 2ey + f = 0$$

$$a = \pi r^2$$

características = sobrenome,
idade
distância = 0.4

29

Aprendizado por Indução (Cont)

- Existe um professor que já tem conhecimento do conceito
 - O professor ajuda na escolha dos exemplos
 - O aprendiz induz a descrição de um conceito, formulando uma regra geral a partir de exemplos e contra-exemplos
- TAREFA: determinar a descrição geral de um conceito

→ **Aprendizado Supervisionado**



2. Aprendizado por exemplos

30

Aprendizado por Exemplos

Nesta estratégia de aprendizado, o sistema aprendiz induz a descrição do conceito ou classe através de processos de generalização e especialização realizados sobre exemplos e, opcionalmente, contra-exemplos do conceito.



31

Jogo de tênis

- Observação e descoberta OU exemplos?
 - ◆ Por quê?

32

Características do Aprendizado Indutivo



- Dois tipos de aprendizado
 - 1.1 Aprendizado Incremental
 - 1.2 Aprendizado Não Incremental

33

Características do Aprendizado Indutivo



- Dois tipos de aprendizado
 - 1.1 Aprendizado Incremental

Em geral, no aprendizado incremental, o processo de aprendizado procede através de uma sequência de hipóteses, H_1, H_2, \dots etc., sobre o conceito que está sendo aprendido. Quando um exemplo é processado, a hipótese corrente é atualizada, se necessário, resultando na próxima hipótese.

34

Características do Aprendizado Indutivo

- Dois tipos de aprendizado

- 1.1 Aprendizado Incremental

- 1.2 Aprendizado Não Incremental

Necessita de que todos os exemplos de treinamento, simultaneamente, estejam disponíveis para que seja induzido um conceito.

É vantajoso usar esses algoritmos para problemas de aprendizado onde todos os exemplos estão disponíveis e, provavelmente, não irão ocorrer mudanças.

35

Características do Aprendizado Indutivo -LD

- Linguagens de Descrição

- ◆ LD de instâncias L_E (Exemplos/Objetos)

- ◆ LD de conceitos L_H (Hipóteses)

- ◆ LD da teoria do domínio L_K (conhecimento de fundo)

36

Características do Aprendizado Indutivo -LD

- Linguagens de Descrição
 - ◆ LD de instâncias $L_{\mathcal{E}}$ (Exemplos/Objetos)
 - ◆ LD de conceitos $L_{\mathcal{H}}$ (Hipóteses)
 - ◆ LD da teoria do domínio $L_{\mathcal{K}}$ (conhecimento de fundo)

37

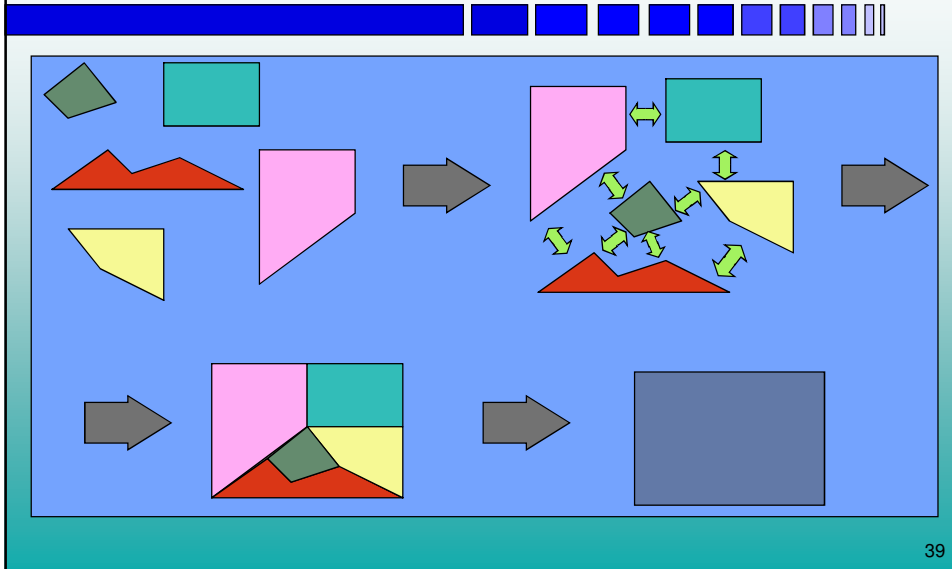
Características do Aprendizado Indutivo - LD

Qualquer que seja o tipo de aprendizado, é necessário uma linguagem para descrever objetos (ou possíveis eventos) e uma linguagem para descrever conceitos

Em geral, é possível distinguir dois tipos de descrições para objetos: **estrutural** e **atributos**

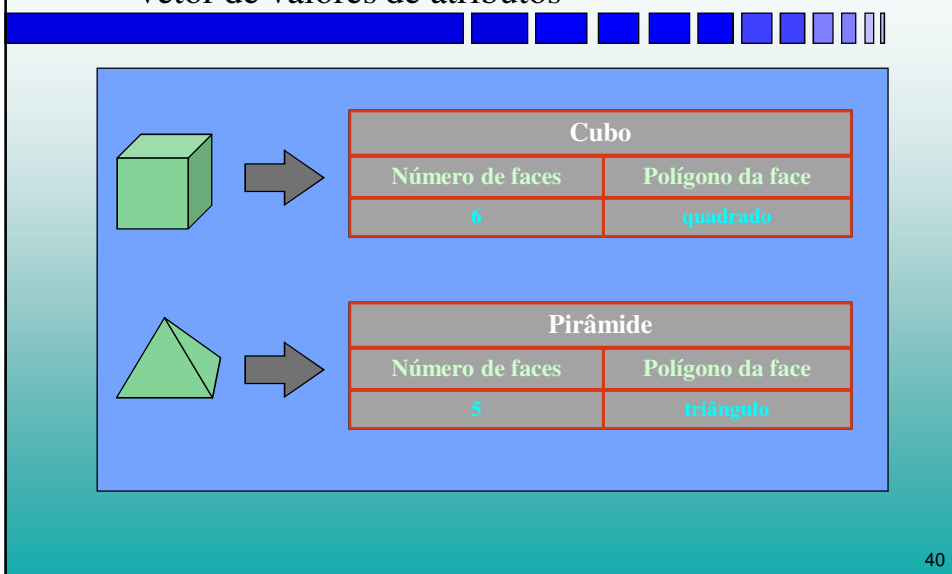
38

1. Descrições estruturais: um objeto é descrito em termos de seus componentes e a relação entre eles



39

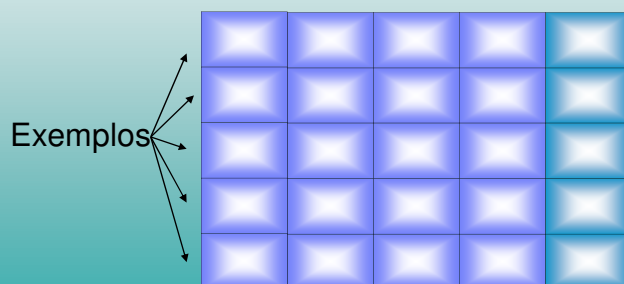
2. Descrições de atributos: um objeto é descrito em termos de suas características globais como um vetor de valores de atributos



40

Exemplo

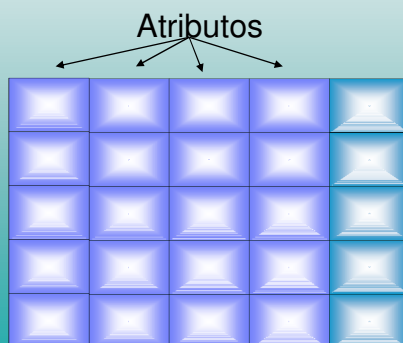
- Exemplo, caso ou registro
- É um conjunto fixo de atributos



41

Atributo

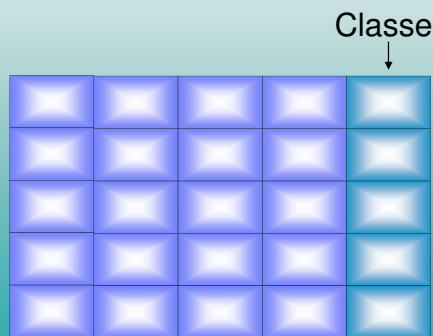
- Atributo, ou campo ou *feature*
- Uma única característica de um exemplo



42

Classe

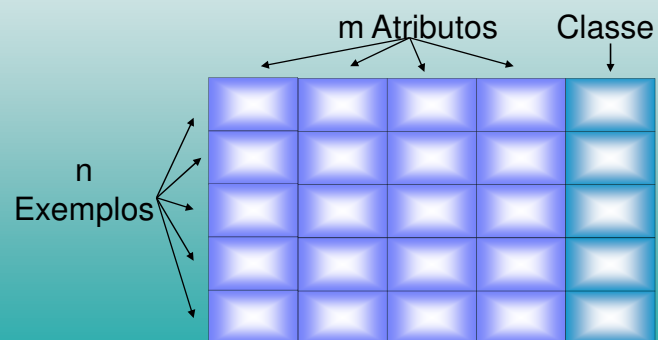
- Atributo especial que descreve o fenômeno de interesse



43

Conjunto de Dados

- Conjunto de exemplos rotulados, segundo sua classe



44

Exemplo de Conjunto de Dados

	X_1	X_2	...	X_m	Y
T1	X_{11}	X_{12}	...	X_{1m}	y_1
T2	X_{21}	X_{22}	...	X_{2m}	y_2
...
Tn	X_{n1}	X_{n2}	...	X_{nm}	y_n

45

Exemplo de Conjunto de Dados

Exemplo

	X_1	X_2	...	X_m	Y
T1	X_{11}	X_{12}	...	X_{1m}	y_1
T2	X_{21}	X_{22}	...	X_{2m}	y_2
...
Tn	X_{n1}	X_{n2}	...	X_{nm}	y_n

46

Exemplo de Conjunto de Dados

Atributo

	X_1	X_2	...	X_m	Y
T1	X_{11}	X_{12}	...	X_{1m}	y_1
T2	X_{21}	X_{22}	...	X_{2m}	y_2
...
Tn	X_{n1}	X_{n2}	...	X_{nm}	y_n

47

Exemplo de Conjunto de Dados

Classe

	X_1	X_2	...	X_m	Y
T1	X_{11}	X_{12}	...	X_{1m}	y_1
T2	X_{21}	X_{22}	...	X_{2m}	y_2
...
Tn	X_{n1}	X_{n2}	...	X_{nm}	y_n

48

Exemplo de Conjunto de Dados

Dia	Tempo	Temperatura	Umidade	Vento	Jogou tênis?
1	Sol	Quente	Alta	Fraco	Não
2	Sol	Quente	Alta	Forte	Não
3	Nublado	Quente	Alta	Fraco	Sim
4	Chuva	Mediana	Alta	Fraco	Sim
5	Chuva	Frio	Normal	Fraco	Sim
6	Chuva	Frio	Normal	Forte	Não
7	Nublado	Frio	Normal	Forte	Sim
8	Sol	Mediana	Alta	Fraco	Não
9	Sol	Frio	Normal	Fraco	Sim
10	Chuva	Mediana	Normal	Fraco	Sim
11	Sol	Mediana	Normal	Forte	Sim
12	Nublado	Mediana	Alta	Forte	Sim
13	Nublado	Quente	Normal	Fraco	Sim
14	Chuva	Mediana	Alta	Forte	Não

49

Características do Aprendizado Indutivo -LD

■ Linguagens de Descrição

- ◆ LD de instâncias L_E (Exemplos/Objetos)
- ◆ LD de conceitos L_H (Hipóteses)
- ◆ LD da teoria do domínio L_K (conhecimento de fundo)

50

Formalismos freqüentemente usados em AM para descrever conceitos são:

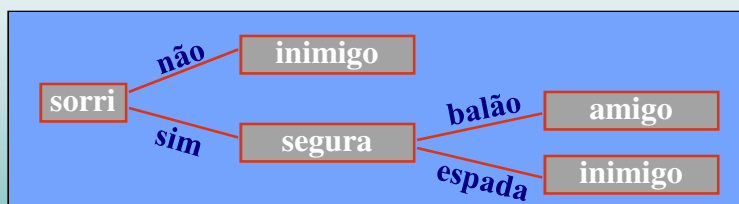
- regras *se-então* (*if-then*) para representar conceitos

*Se Nublado ou Chovendo
então Levar_Guarda-Chuva*

51

Formalismos freqüentemente usados em AM para descrever conceitos são: (Cont)

- árvores de decisão para representar conceitos



- lógica de predicados

$filha(X, Y) \leftarrow mulher(X), pais(X, Y).$

52

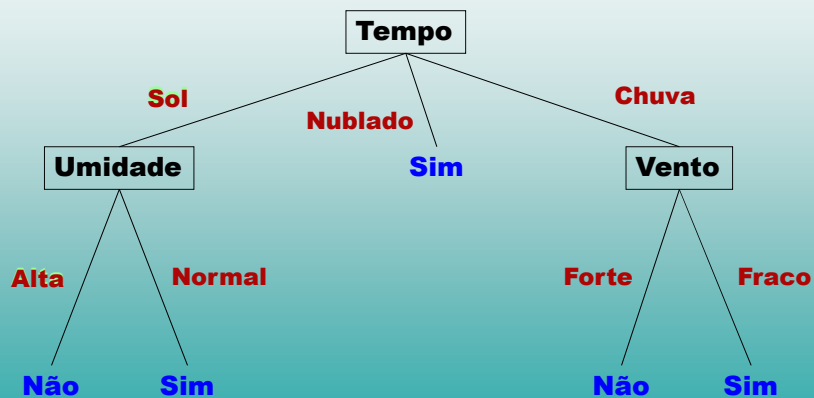
Formalismos freqüentemente usados em AM para descrever conceitos são: (Cont)

■ redes semânticas



53

Qual o formalismo?



54

Características do Aprendizado Indutivo -LD

■ Linguagens de Descrição

- ◆ LD de instâncias $L_{\mathcal{E}}$ (Exemplos/Objetos)
- ◆ LD de conceitos $L_{\mathcal{H}}$ (Hipóteses)
- ◆ LD da teoria do domínio $L_{\mathcal{K}}$ (conhecimento de fundo)

55

Conhecimento de fundo: constituído por algum conhecimento relevante do domínio do problema

■ Exemplo do viajante na Itália:

- ◆ A generalização de que todos os italianos falam italiano é sustentada pela regularidade mais geral de que em um dado país a maioria da população fala a mesma língua; por outro lado, não é assumido que todos os italianos são chamados de Giuseppe devido à regularidade mais geral de que a maioria dos grupos sociais utilizam nomes diversos para diferentes indivíduos

56

Aprendizado Indutivo de Conceitos - AIC

Dados

$\mathcal{E} = \mathcal{E}^+ \cup \mathcal{E}^-$: conjunto de exemplos de treinamento de um conceito C

→ encontrar uma hipótese H , expressa em uma linguagem de descrição L tal que:

- cada exemplo $e \in \mathcal{E}^+$ é coberto por H
- nenhum exemplo negativo $e \in \mathcal{E}^-$ é coberto por H

57

Aprendizado Indutivo de Conceitos - AIC (Cont)

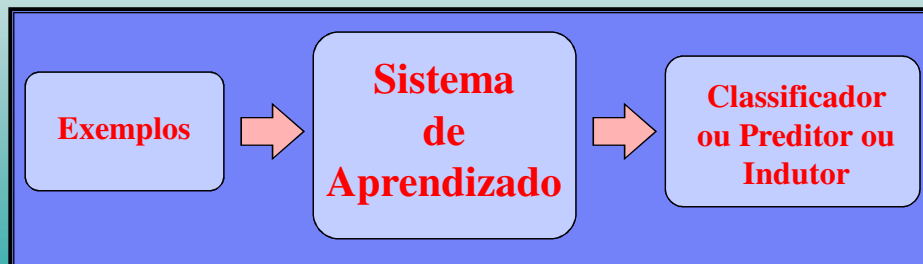
$cobre(H, \mathcal{E}) = \{e \in \mathcal{E}^+ \mid cobre(H, e) = true\}$
(instância positiva)

$cobre(H, \mathcal{E}) = \{e \in \mathcal{E}^- \mid cobre(H, e) = false\}$
(instância negativa)

58

Sistemas de Aprendizado de Máquina Indutivo

O paradigma de aprendizado indutivo busca **aprender conceitos através de instâncias** destes conceitos



59

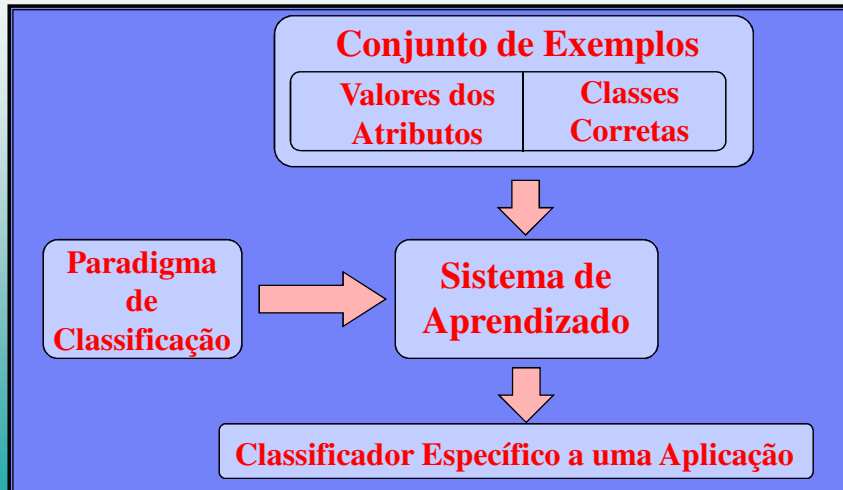
Sistemas de Aprendizado de Máquina Indutivo (Cont)

O classificador utiliza os conceitos aprendidos para **classificar novos exemplos**



60

Representação da Classificação



61

Exercício de Classificação

Grupos de 2 alunos

Atributos

Comprimento do Cabelo











Peso

Idade

Atributo de Classe: Sexo

62

Induza a hipótese em 10 minutos!

	Pessoa	Comprimento do Cabelo	Peso	Idade	Classe: Sexo
Dados	 Homer	0	250	36	M
	 Marge	10	150	34	F
	 Bart	2	90	10	M
	 Lisa	6	78	8	F
	 Maggie	4	20	1	F
	 Abe	1	170	70	M
	 Selma	8	160	41	F
	 Otto	10	180	38	M
	 Krusty	6	200	45	M
	 Comic	8	290	38	?

Preparação de Dados

- Fase que antecede o processo de aprendizagem, para **facilitar ou melhorar o processo**
- Exemplos
 - ◆ Remover **exemplos incorretos**
 - ◆ **Transformar o formato** dos dados para que possam ser usados com um determinado indutor
 - ◆ Selecionar **atributos relevantes** (Seleção de Atributos)
 - Diversas técnicas

Ruído

- **Dados imperfeitos** que podem ser derivados do processo de aquisição, transformação ou rotulação das classes

- ◆ Exemplos com os mesmos atributos mas com classes diferentes

	X ₁	X ₂	X ₃	X ₄	Y
overcast	19	65	yes		dont_go
rain	19	70	yes		dont_go
rain	23	80	yes		dont_go
sunny	23	95	no		dont_go
sunny	28	91	yes		dont_go
sunny	30	85	no		dont_go
overcast	19	65	yes		go
rain	21	80	no		go
rain	22	95	no		go
sunny	22	70	no		go
overcast	23	90	yes		go
rain	25	81	no		go
sunny	25	72	yes		go
overcast	26	75	no		go
overcast	29	78	no		go

65

Classificador

- Um exemplo pode ser representado pelo par:

$$(x, y) = (x, f(x))$$

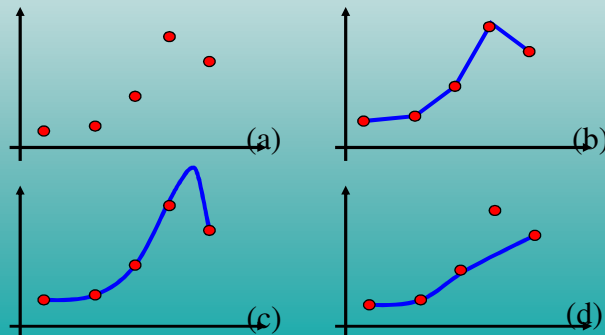
onde

- x é a entrada;
- $f(x)$ é a saída (f desconhecida!)
- Indução ou inferência indutiva: dada uma coleção de exemplos de f , retornar uma função h que aproxima f
- h é denominada uma **hipótese** sobre f

66

Exemplos de Hipóteses

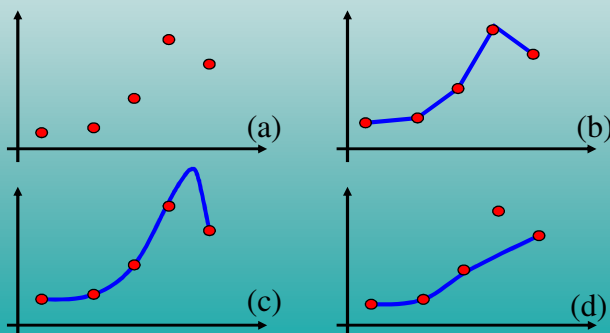
- (a) dados originais
- (b), (c), (d) possíveis hipóteses



67

Bias

- Qualquer critério de **preferência** de uma hipótese sobre outra (além da consistência com os dados)



68

Bias

- Quando o bias pode ocorrer?

69

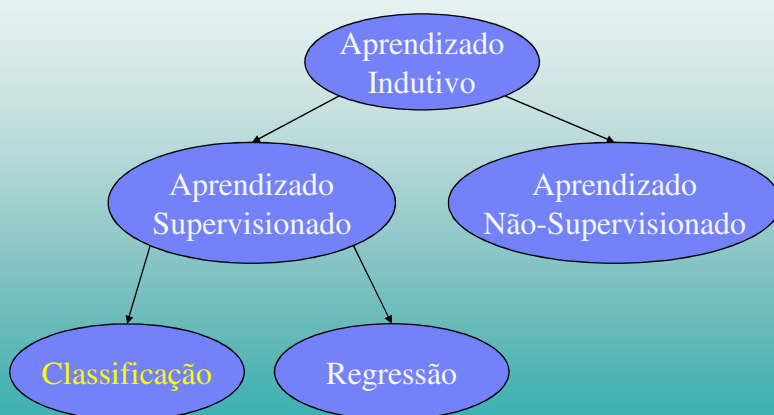
Classificação e Regressão

Qual é a diferença entre Classificação e Regressão ?

Em problemas de Regressão a variável de saída assume valores contínuos, enquanto que em problemas de classificação é estritamente categórica.

70

Hierarquia do Aprendizado



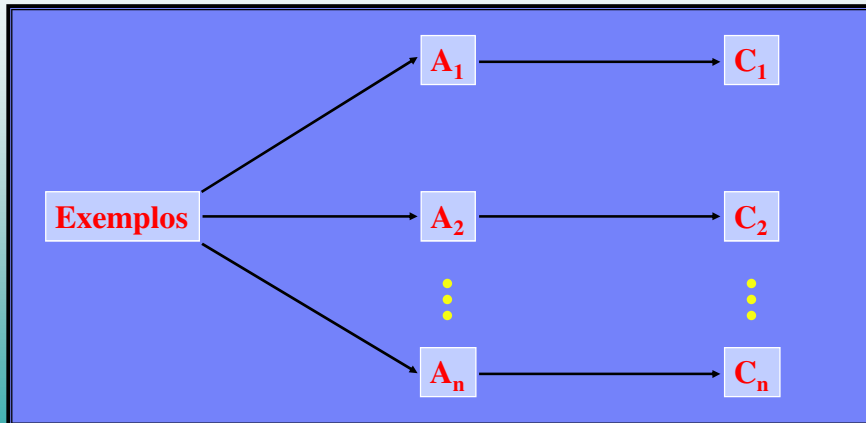
71

Sistemas de Aprendizado de Máquina

<i>Modo de Aprendizado</i>	<i>Paradigmas de Aprendizado</i>	<i>Linguagens de Descrição</i>	<i>Formas de Aprendizado</i>
- Supervisionado - Não Supervisionado	- Simbólico - Estatístico - Baseado em exemplos - Conexionista - Evolutivo	- Instâncias ou Exemplos - Conceitos Aprendidos ou Hipóteses - Teoria de Domínio ou Conhecimento de Fundo	- Incremental - Não Incremental

72

Algoritmo (A_i) Para Construir o Melhor Classificador (C_i)?



73

Algoritmo (A_i) Para Construir o Melhor Classificador (C_i)?

Estudos experimentais são necessários, uma vez que não existe uma análise matemática que possa determinar se um algoritmo de aprendizado irá desempenhar bem em um conjunto de exemplos.

74

Erro e Precisão

Recordando a notação adotada

- Exemplo $(x, y) = (x, f(x))$
- Atributos: x
- Classe (rotulada): $y = f(x)$
- Classe (classificada): $h(x)$
- n é o número de exemplos

75

Erro e Precisão

- Classificação

$$err(h) = \frac{1}{n} \sum_{i=1}^n \|y_i \neq h(x_i)\| \quad (\text{erro})$$

$$acc(h) = 1 - err(h) \quad (\text{precisão})$$

- O operador $\|E\|$ retorna:
 - ◆ 1 se E é verdadeiro
 - ◆ 0 se E é falso

76

Pergunta

- Qual o erro de sua(s) hipótese(s) sobre os Simpsons?

77

Erro e Precisão

- Regressão: distância entre valor real e predito
 - ◆ Duas medidas usualmente utilizadas
 - mse: *mean squared error*
 - mad: *mean absolute distance*

$$\text{mse - err}(h) = \frac{1}{n} \sum_{i=1}^n (y_i - h(x_i))^2$$

$$\text{mad - err}(h) = \frac{1}{n} \sum_{i=1}^n |y_i - h(x_i)|$$

78