



Inteligência Artificial

Profa. Patrícia R. Oliveira
EACH / USP

Parte 2 – Introdução ao Aprendizado de Máquina

Este material é parcialmente baseado em slides
dos Profs. Thiago Pardo, Ricardo Campello (ICMC/USP) e José Augusto
Baranauskas (FFCLRP/USP)



O que é Aprendizado?

- Aprendizado é qualquer processo pelo qual um sistema melhora o seu desempenho em uma determinada tarefa a partir da experiência (Herbert Simon).
- Quais seriam as tarefas?
 - Classificação
 - Solução de problemas / Planejamento / Controle



Classificação

- Associa um objeto ou evento a uma das categorias pertencentes a um conjunto finito de possibilidades.
- Exemplos:
 - Diagnóstico médico:
 - classifica registros de sintomas de pacientes em um tipo de doença.
 - Filtragem de SPAM em emails;
 - Reconhecimento de objetos em imagens;



Solução de Problemas / Planejamento / Controle

- Desempenha ações em um determinado ambiente com o objetivo de alcançar uma determinada meta.
- Exemplos:
 - Solução de problemas de cálculo;
 - Jogos de dama, xadrez, etc.;
 - Direção de veículos;
 - Pilotagem de aeronaves, helicópteros ou foguetes;
 - Controle de elevadores;
 - Controle de robôs móveis.

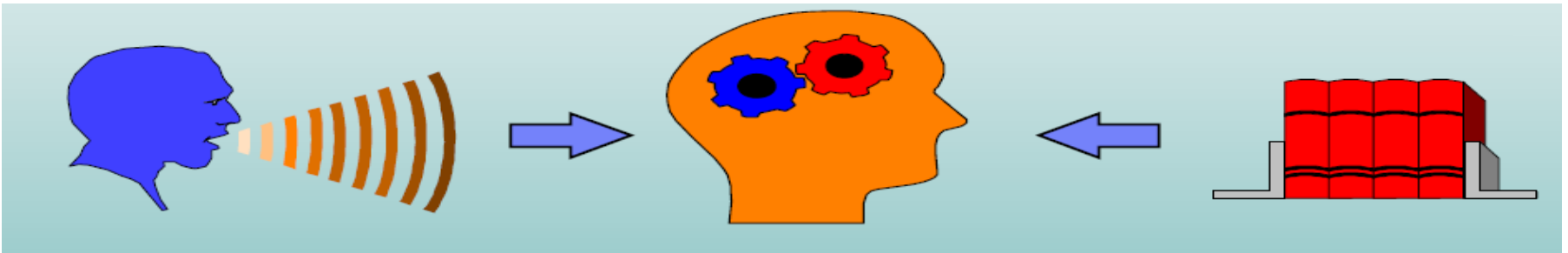
Aprendizado de Máquina (AM)

- Definição: é uma área da Inteligência Artificial que pesquisa métodos relacionados à aquisição de novos conhecimentos, novas habilidades e novas formas de organizar o conhecimento já existente.

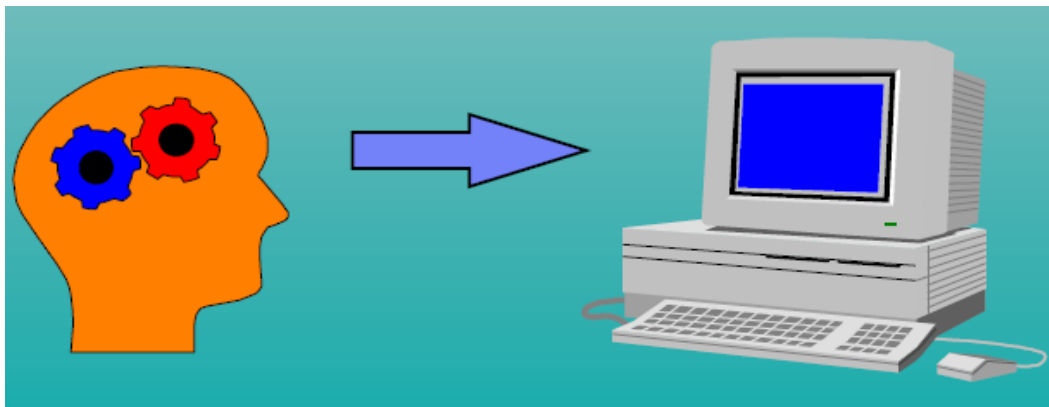


Objetivos de AM

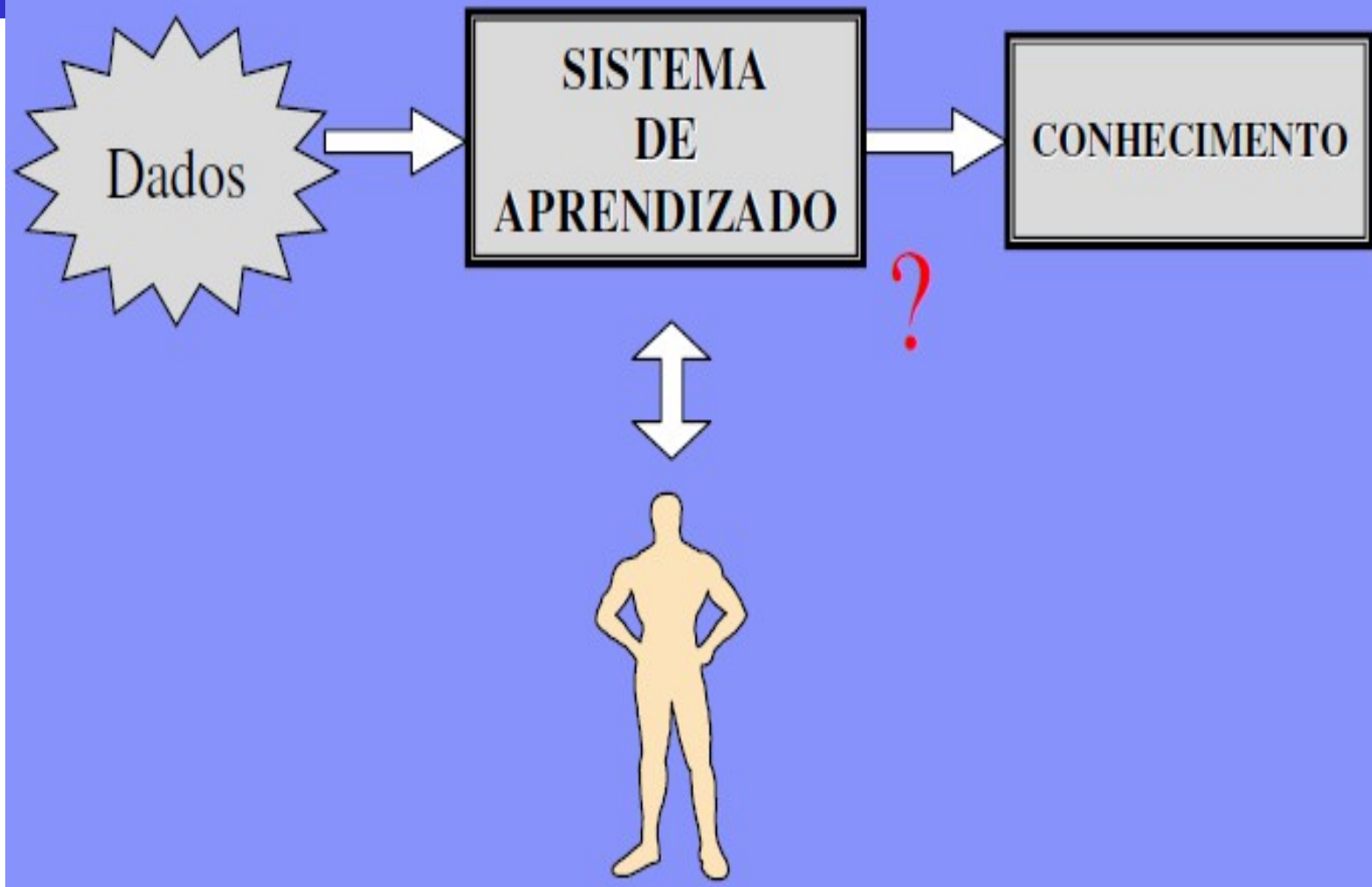
- Um melhor entendimento dos mecanismos de aprendizado humano.



- Automação da aquisição do conhecimento.



Aprendizado

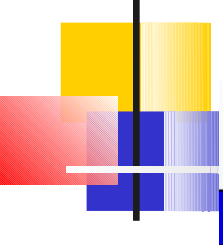


Paradigmas de AM

1) Simbólico

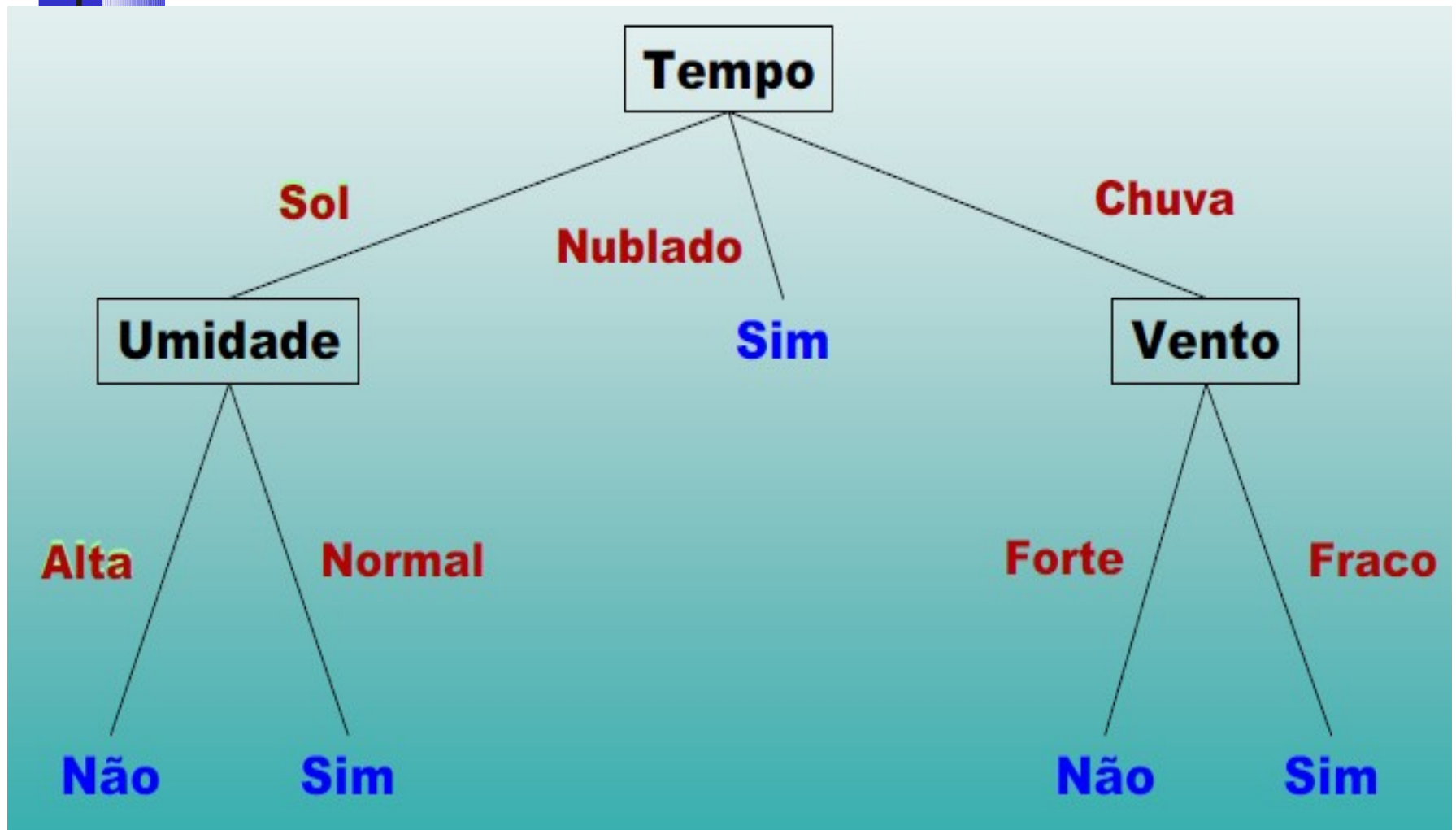
- Busca aprender construindo representações simbólicas de um conceito.
 - análise de exemplos e contra-exemplos desse conceito.
- Representações simbólicas:
 - expressões lógicas;
 - árvores de decisão;
 - regras;
 - redes semânticas.

Conjunto de Exemplos / Contra-exemplos



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

Esboço de uma árvore de decisão



Paradigmas de AM

2) Estatístico

- Utiliza modelos estatísticos para encontrar uma boa aproximação do conceito a ser aprendido.
- Destaque para o Aprendizado Bayesiano:
 - usa um modelo probabilístico baseado no conhecimento prévio do problema.

Paradigmas de AM

3) Baseado em Exemplos

- Classifica exemplos nunca vistos por meio de exemplos similares conhecidos.
- Mantém os exemplos disponíveis na memória
 - ao contrário de outros paradigmas que utilizam os exemplos apenas para induzir o modelo, descartando-os logo após a indução.
 - Exemplos:
 - Raciocínio baseado em casos (CBR)
 - k-Nearest Neighbours (kNN)

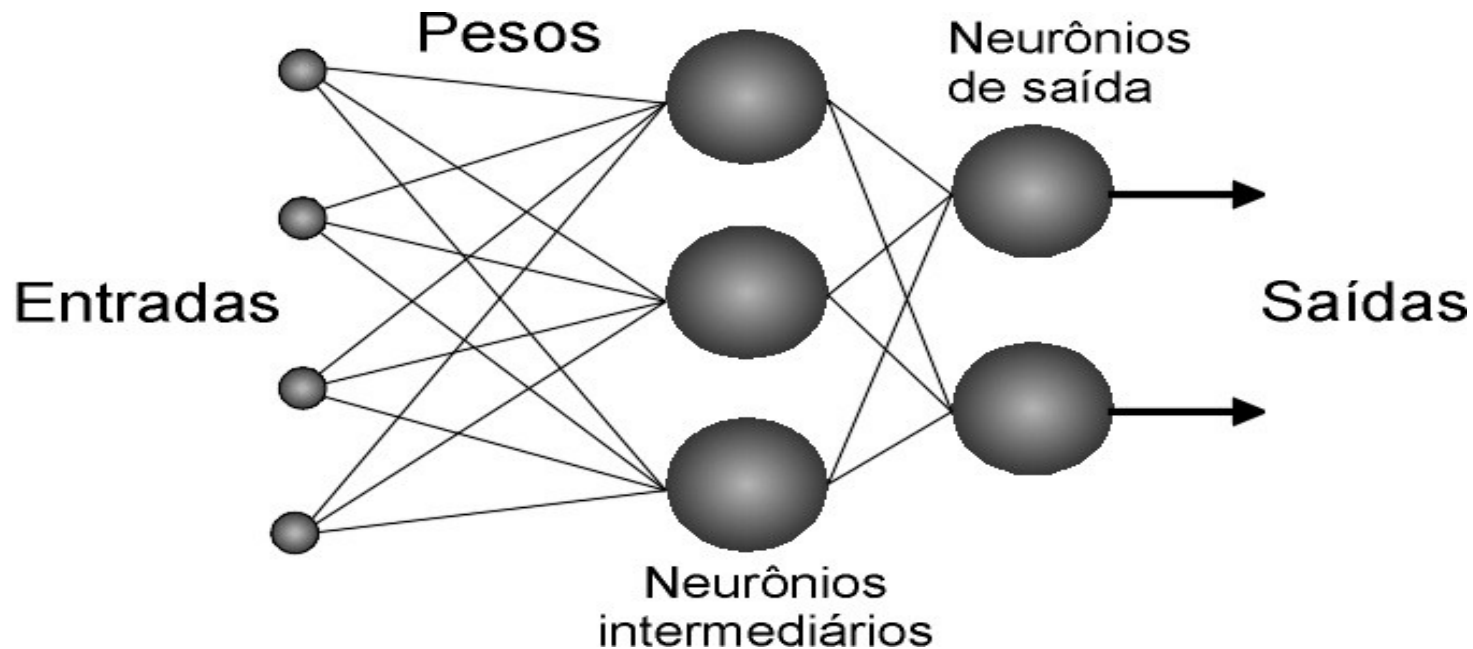
Paradigmas de AM

4) Conexionista

- Redes Neurais são construções matemáticas simplificadas inspiradas no modelo biológico do sistema nervoso humano.
- O termo conexionismo advém do fato que a representação de uma rede neural envolve unidades altamente interconectadas.

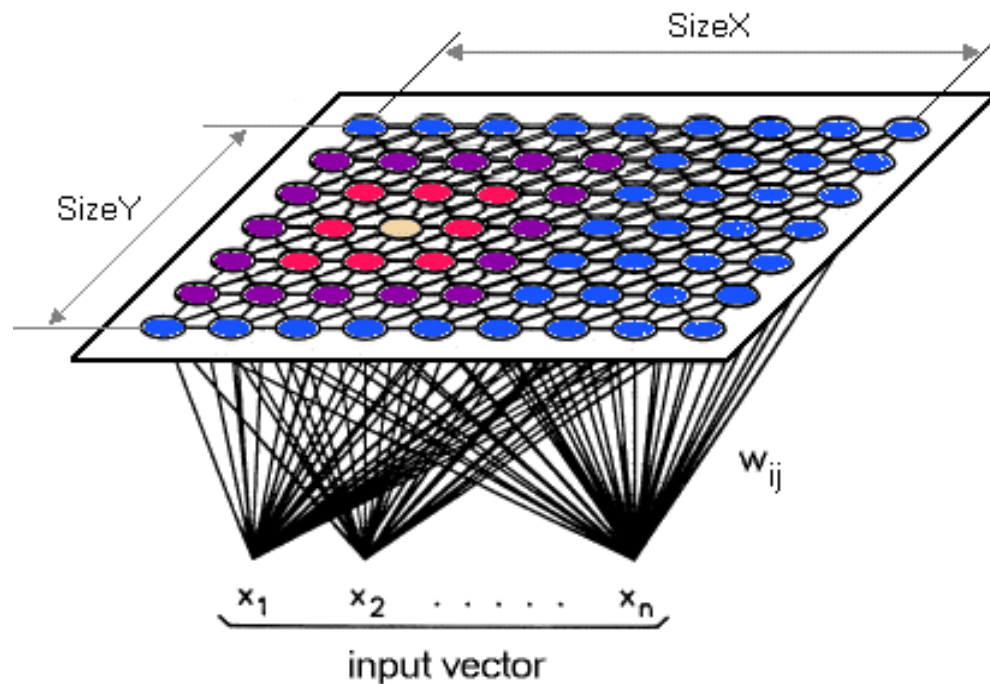
Redes Neurais

- Exemplo de uma Rede Neural Perceptron Multicamadas (Multilayer Perceptron).



Redes Neurais

- Exemplo de um Mapa Auto-Organizável (SOM – Self-Organizing Map).

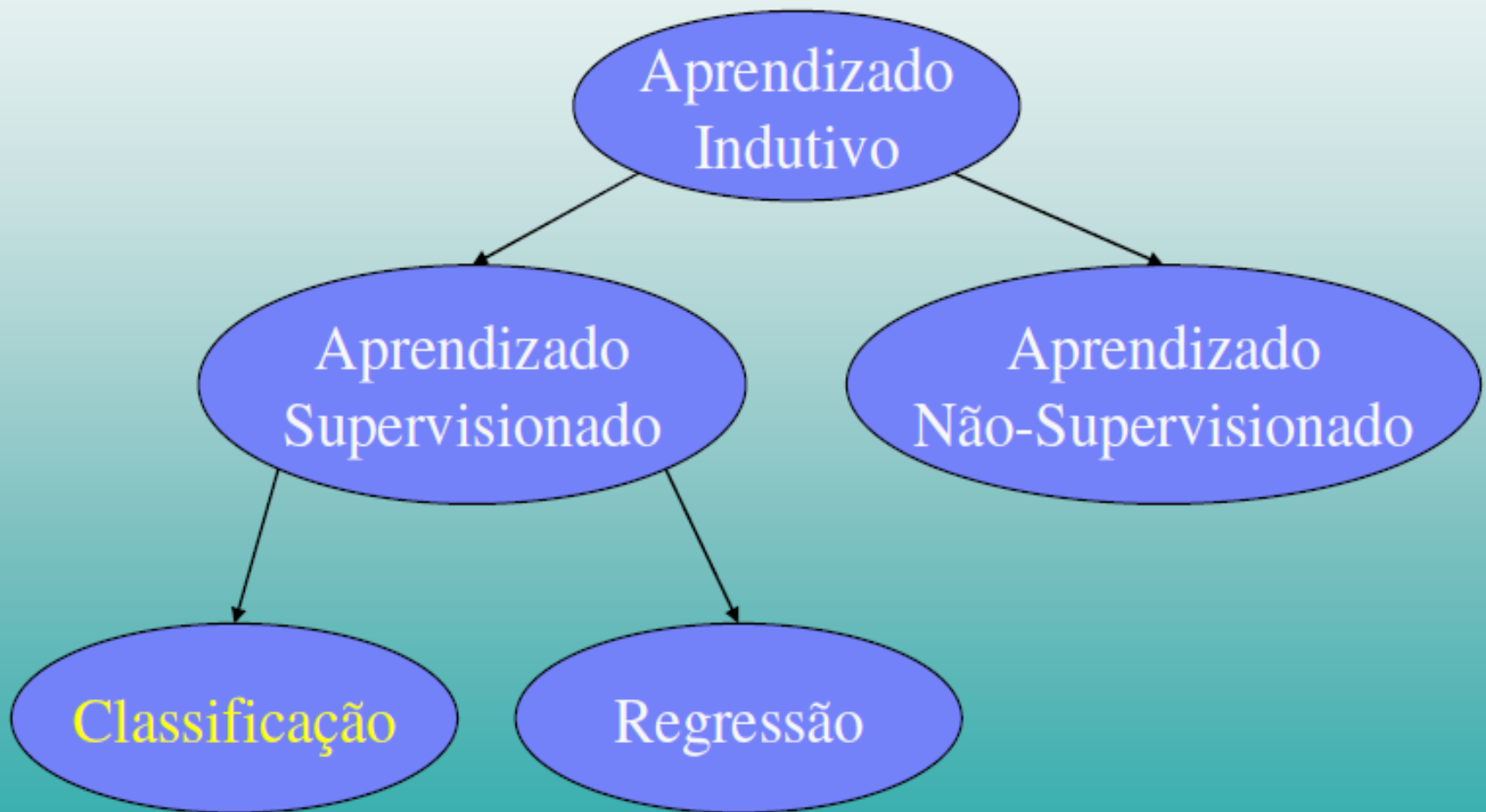


Paradigmas de AM

5) Evolutivo

- Possui analogia direta com a Teoria de Darwin, em que sobrevivem os mais bem adaptados ao ambiente.
- População: conjunto de soluções para um determinado problema.
- A população evolui por meio da aplicação de operadores, como *crossover* e mutação.

Hierarquia do Aprendizado de Máquina





Aprendizado Indutivo

- A inferência indutiva permite obter conclusões genéricas sobre um conjunto particular de exemplos.
- Caracterizado pelo raciocínio originado em um conceito específico, que é ao longo do processo, generalizado.
- Utilizado para derivar conhecimento novo e prever eventos futuros.
- Efetuado a partir de raciocínio sobre exemplos fornecidos por um processo externo ao sistema.



Aprendizado Supervisionado

- O algoritmo de aprendizado (indutor) recebe um conjunto de exemplos de treinamento para os quais os rótulos da classe associada são conhecidos.
- Cada exemplo (instância ou padrão) é descrito por um vetor de valores (atributos) e pelo rótulo da classe associada.
- O objetivo é construir um classificador que possa determinar corretamente a classe de novos exemplos ainda não rotulados.
- Para rótulos de classe discretos, esse problema é chamado de classificação e para valores contínuos é conhecido como regressão.



Aprendizado Supervisionado

- Alguns métodos:
 - Classificação e regressão
 - Redes Neurais Perceptron Multicamadas
 - Somente classificação
 - Árvores de decisão
 - Naive Bayes
 - Redes Neurais Perceptron



Aprendizado Não Supervisionado

- O indutor analisa os exemplos (não rotulados) fornecidos e tenta determinar se alguns deles podem ser agrupados de alguma maneira, formando agrupamentos ou clusters.
- Após a determinação dos agrupamentos, em geral, é necessário uma análise para determinar o que cada agrupamento significa no contexto problema sendo analisado.



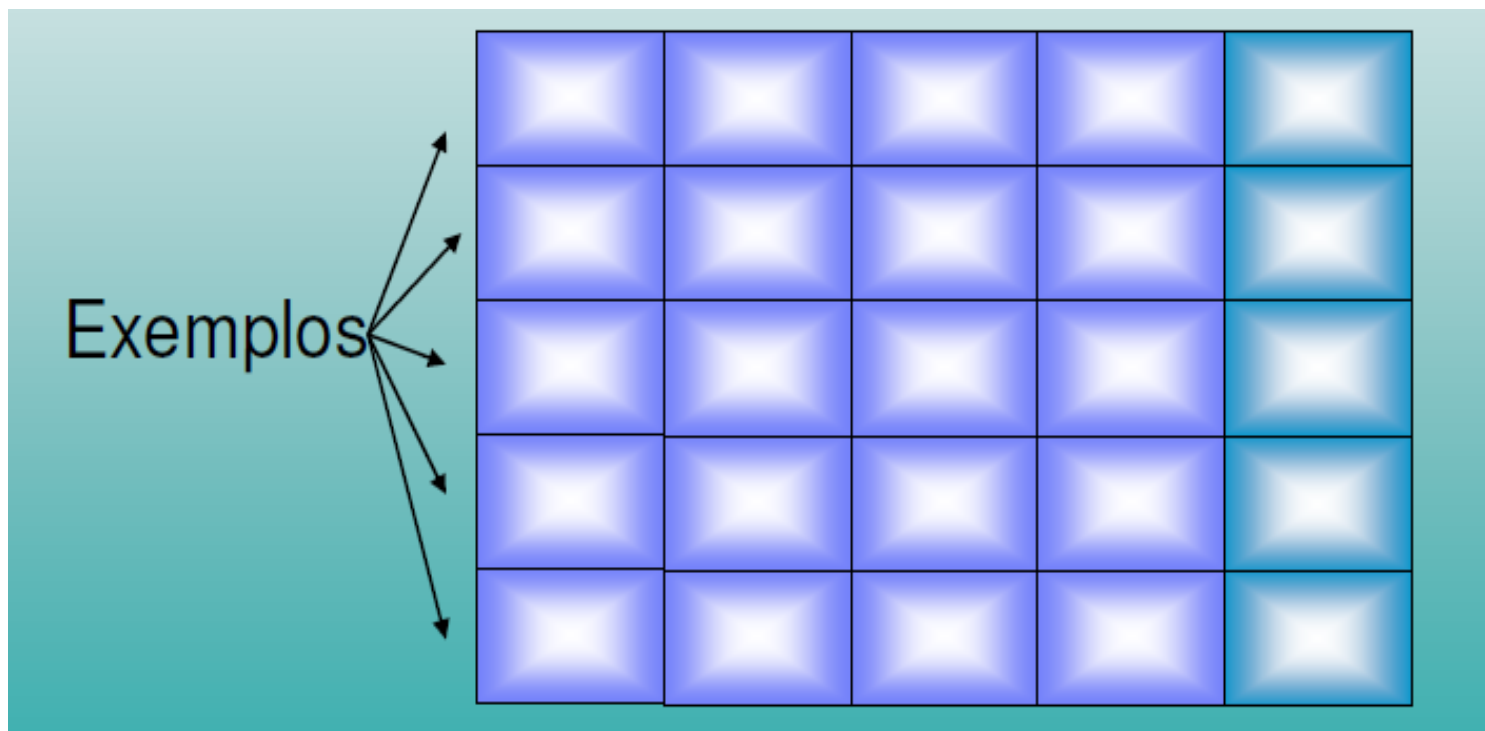
Aprendizado Não Supervisionado

- Alguns métodos:
 - k-médias
 - Técnicas de Agrupamento Hierárquico
 - Mapas Auto-organizáveis (Redes Neurais SOM)



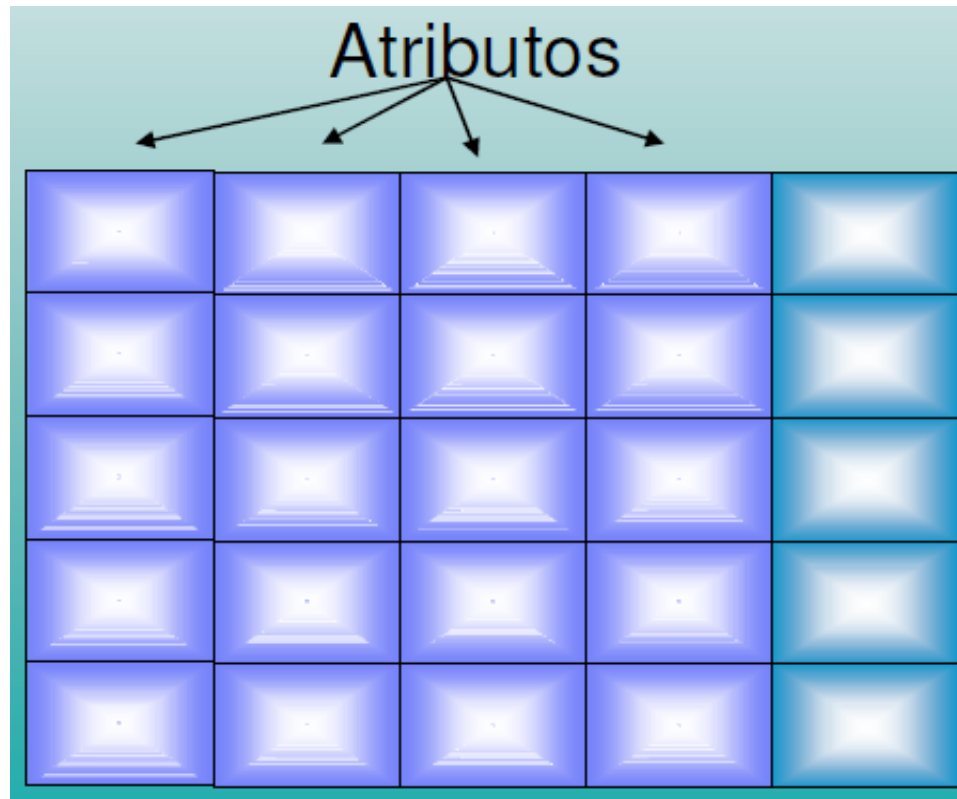
Exemplo, instância ou padrão

- Cada exemplo fornecido ao indutor é caracterizado por um conjunto de valores fixos de atributos.



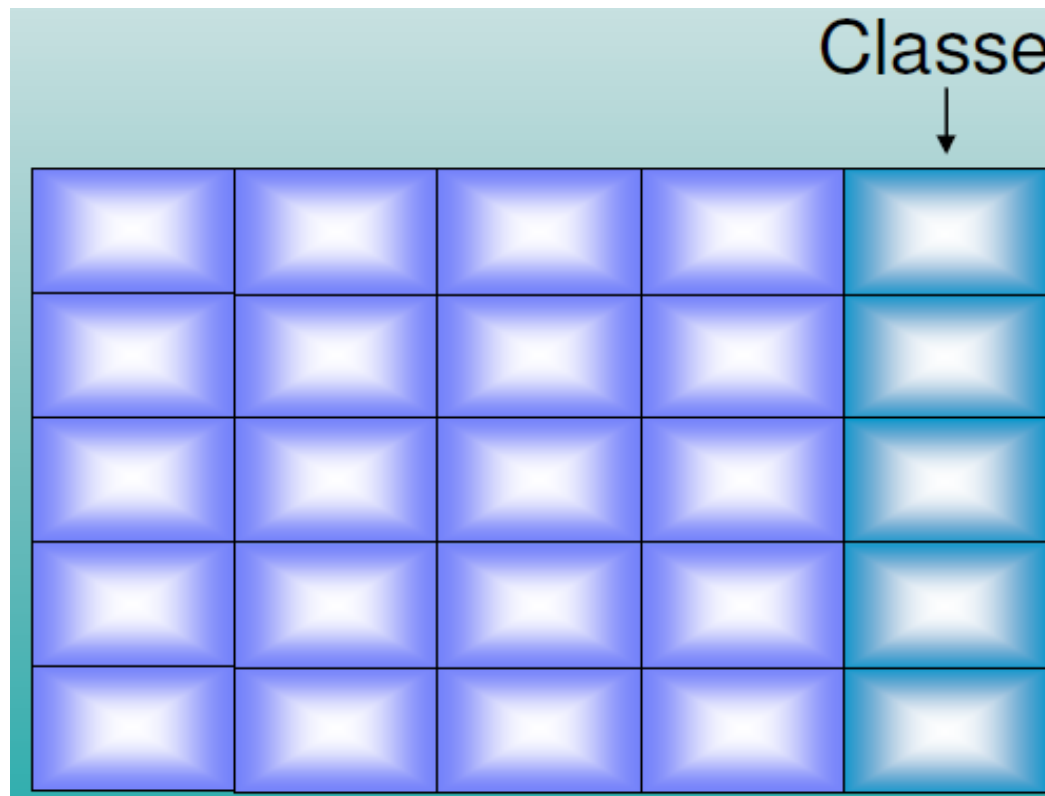
Atributo ou campo

- É uma determinada característica de um exemplo.

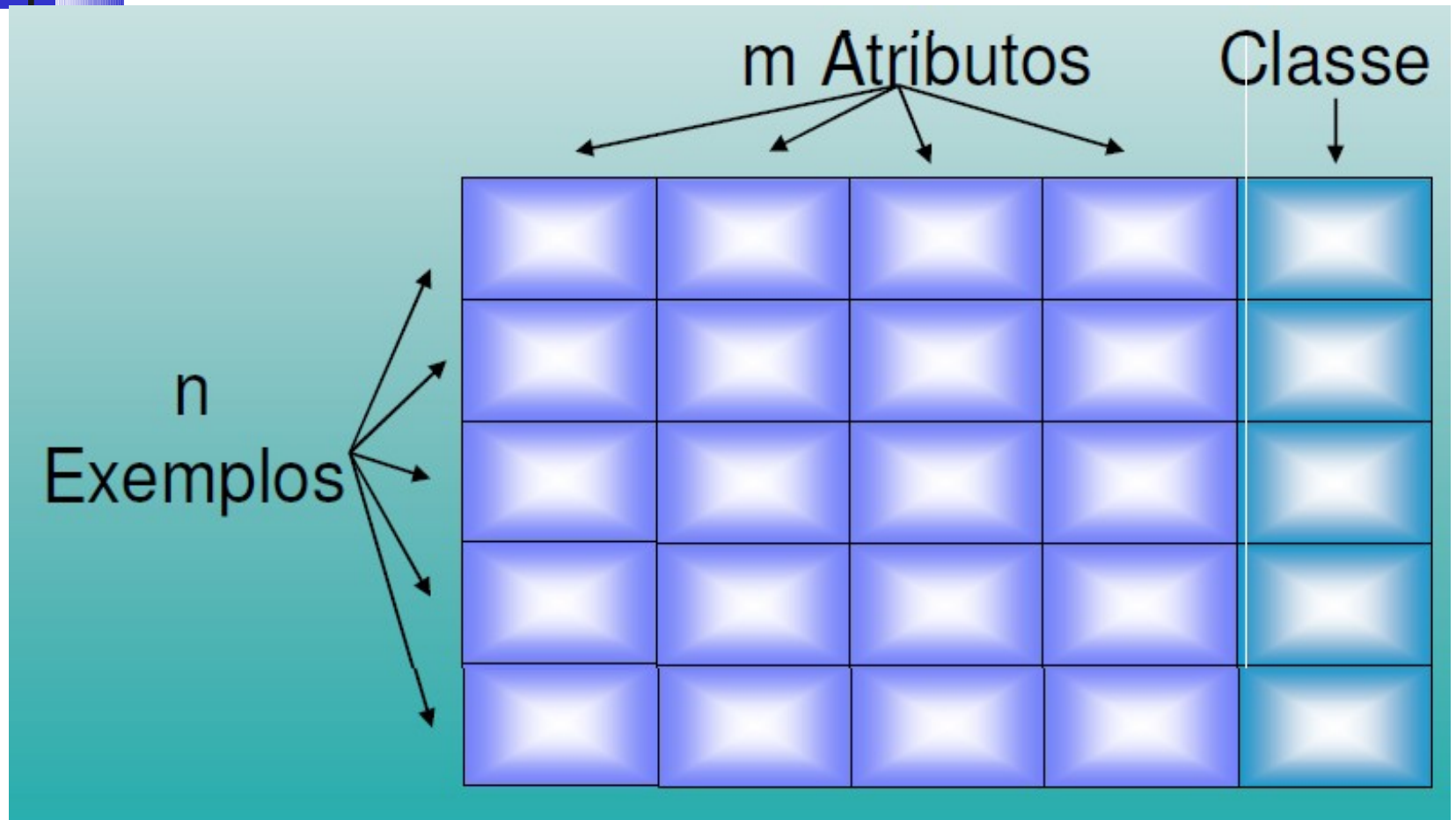


Classe

- É um atributo especial que descreve o fenômeno de interesse.



Conjunto de Dados

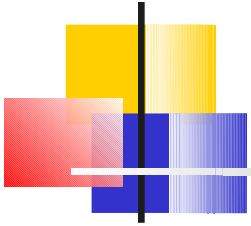


Esquema de um Conjunto de Dados

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

Esquema de um 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

Esquema de um 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

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



Conjuntos de Treinamento e Teste

- De modo geral, o conjunto de exemplos disponíveis deve ser dividido em dois conjuntos disjuntos:
 - conjunto de treinamento
 - usado para o aprendizado do conceito (construção do modelo).
 - conjunto de teste
 - usado para medir o desempenho e grau de generalização do modelo construído.
 - a disjunção dos conjuntos é exigida para tornar essa medida estatisticamente válida.



Classificador

- Dado um conjunto de exemplos de treinamento, um indutor gera como saída um classificador.
- Para um novo exemplo, dado como entrada para o classificador, espera-se que este possa predizer, com a maior precisão possível, a sua classe.



Classificador

- Formalmente, um exemplo pode ser representado pelo par:

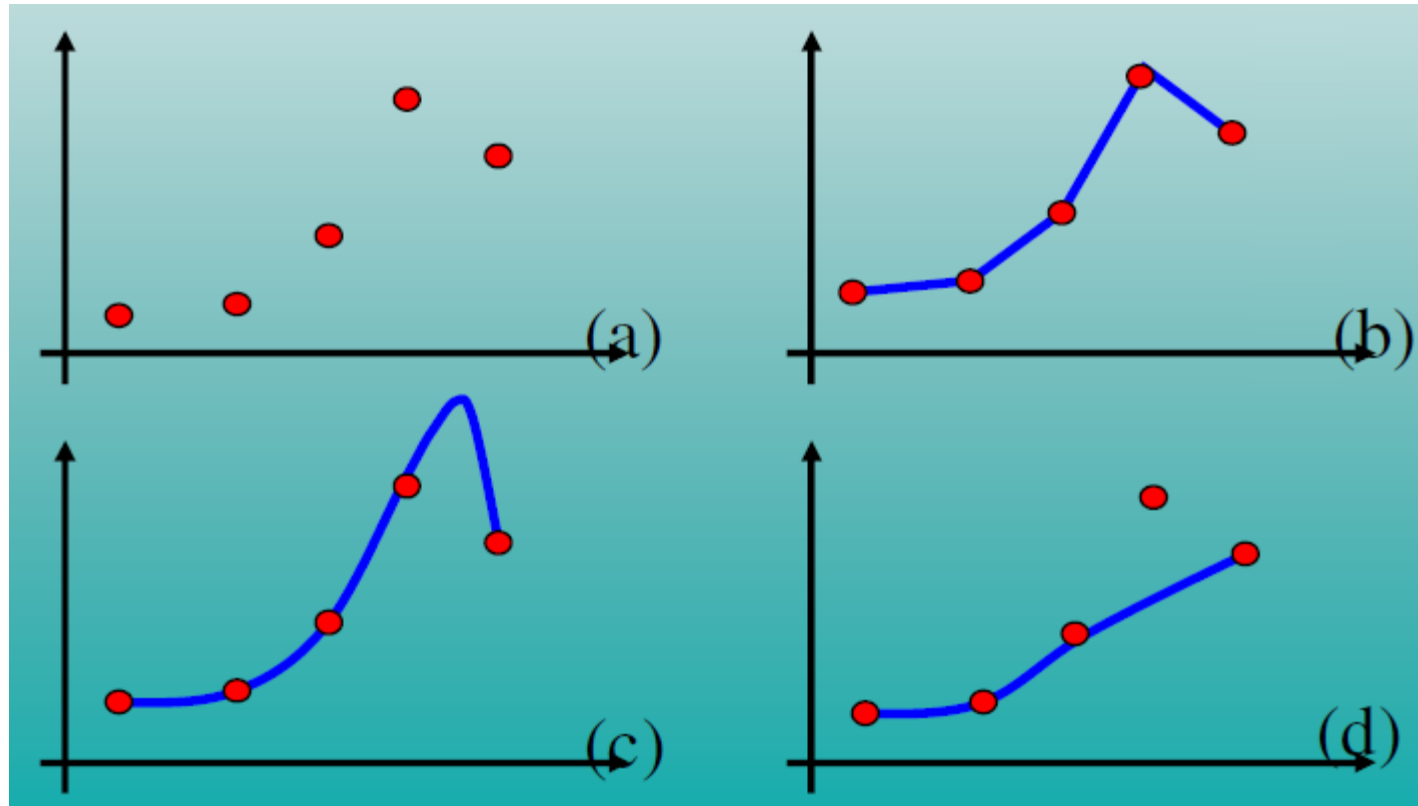
$$(x, y) = (x, f(x)),$$

- em que:

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

Exemplos de Hipóteses

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

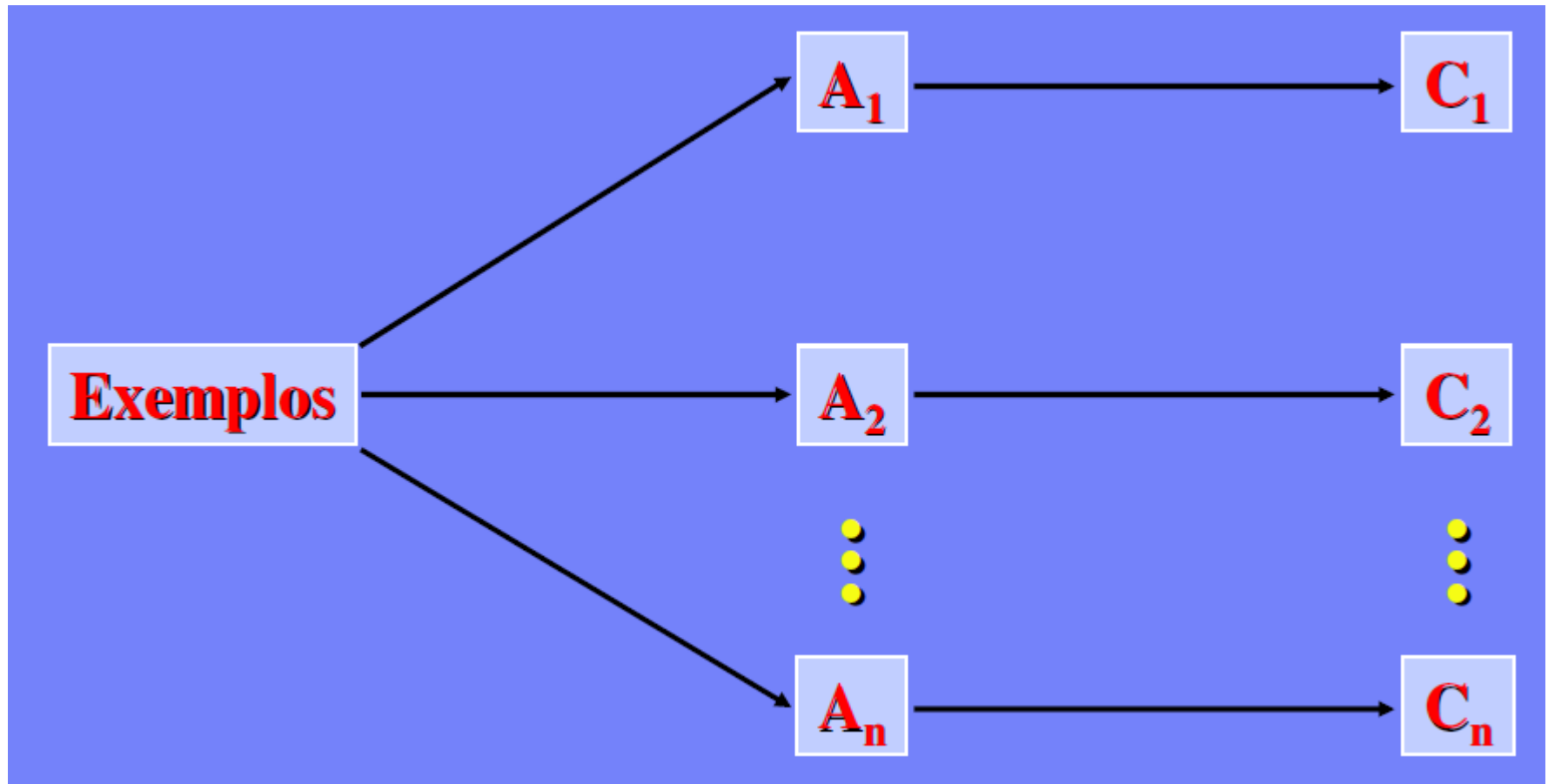




Classificação e Regressão

- Qual é a diferença entre Classificação e Regressão?
 - Em problemas de regressão a variável de saída y assume valores contínuos.
 - Em problemas de classificação a variável de saída y é estritamente categórica.

Qual Algoritmo (A_i) gera o melhor Classificador (C_i)?





Qual Algoritmo (A_i) gera 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 terá um bom desempenho em um conjunto de exemplos.



Erro e Acurácia

- Lembrando a notação adotada
 - Exemplo $(x, y) = (x, f(x))$
 - Atributos: x
 - Classe (rotulada): $y = f(x)$
 - Classe (classificada pelo modelo): $h(x)$
 - n : número de exemplos



Erro e Acurácia

- Classificação

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

(Erro)

$$acc(h) = 1 - err(h)$$

(Acurácia)

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



Erro e Acurácia

- Regressão: distância entre o valor real e o predito

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

- mse: erro quadrático médio (*mean squared error*)
- mad: distância absoluta média (*mean absolute distance*)



Distribuição de Classes

- Dado um conjunto de exemplos T , é possível calcular sua distribuição de classes.
- Para cada classe C_j , sua distribuição $Distr(C_j)$ é calculada como:

$$Distr(C_j) = \frac{1}{n} \sum_{i=1}^n \|y_i = C_j\|$$



Distribuição de Classes

- Se um conjunto com 100 exemplos, possui 60 exemplos da classe C_1 , 15 exemplos da classe C_2 e 25 exemplos da classe C_3 , então sua distribuição de classes é:

$$\text{Distr}(C_1, C_2, C_3) = (0.60, 0.15, 0.25) = (60\%, 15\%, 25\%)$$

- Nesse exemplo, a classe C_1 é a classe majoritária ou prevalente.




Erro Majoritário

- Dada a distribuição de k classes em um conjunto de exemplos T , pode-se calcular o erro majoritário desse conjunto como:

$$maj-err(T) = 1 - \max_{i=1,2,\dots,k} Distr(C_i)$$

- No exemplo anterior, o erro majoritário é $maj-err(T) = 1 - 0.6 = 40\%$.

Qual o erro majoritário?



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



Erro Majoritário

- O erro majoritário é independente do algoritmo de aprendizado.
- Fornece um limiar máximo abaixo do qual o erro de um classificador deve ficar.
- Prevalência de Classe: quando há um desbalanceamento de classes no conjunto de exemplos.



Prevalência de Classe

- Suponha um conjunto de exemplos T com a seguinte distribuição de classes:

$$\text{Distr}(C_1, C_2, C_3) = (99\%, 0.25\%, 0.75\%)$$

- Prevalência da classe C_1 .

- Um modelo bem simples, que classifique todos os exemplos como sendo da classe C_1 teria uma acurácia de 99%!!



Prevalência de Classe

- Pode ser indesejável quando as classes minoritárias possuem uma informação muito importante.
- Para o exemplo anterior, considere:
 - C_1 : paciente normal;
 - C_2 : paciente com doença A;
 - C_3 : paciente com doença B.

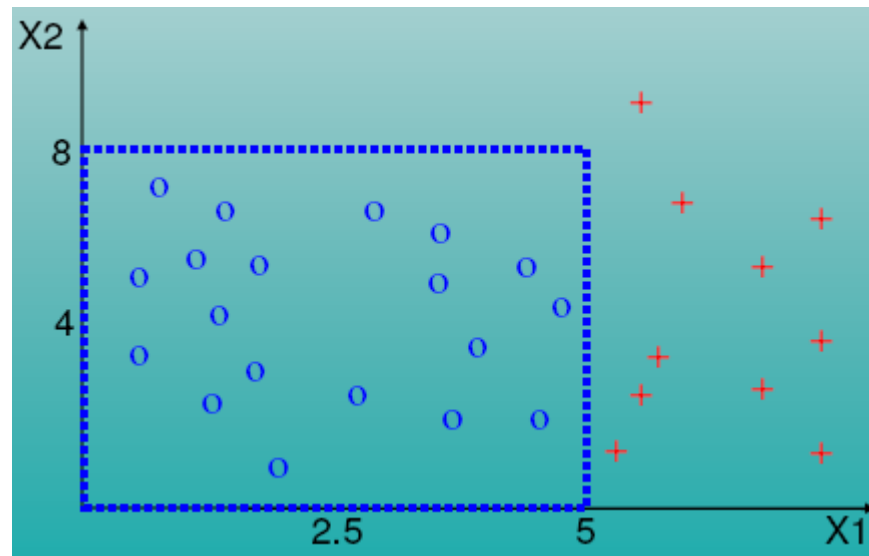


Espaço de Descrição

- m atributos podem ser vistos como um vetor.
- Cada atributo corresponde a uma coordenada num espaço m-dimensional denominado espaço de descrição.
- Cada ponto no espaço de descrição pode ser rotulado com a classe associada aos atributos.

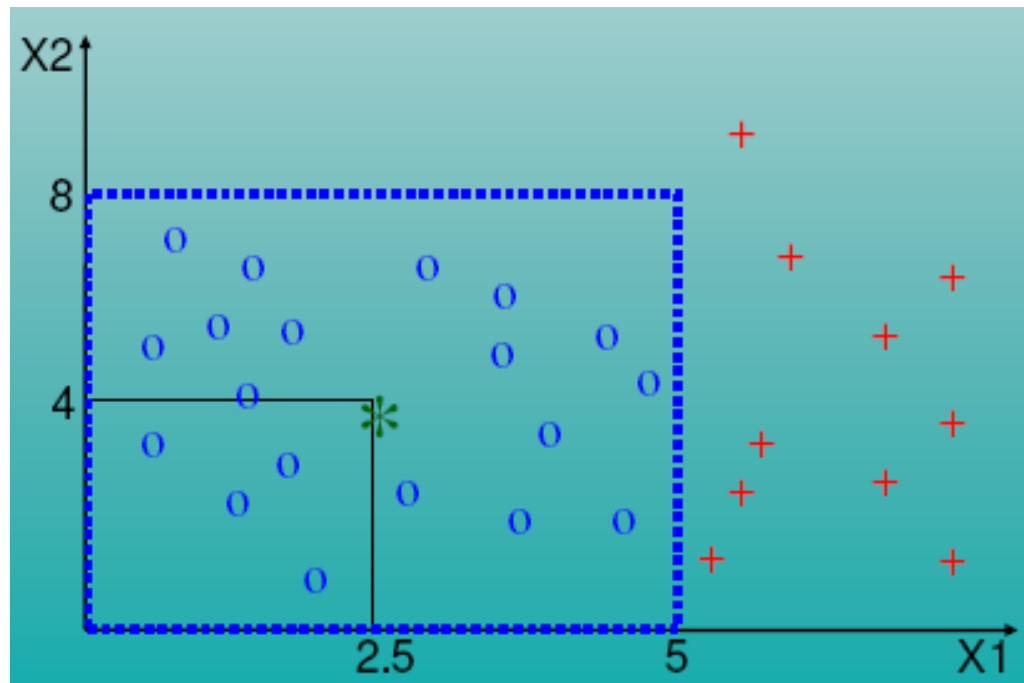
Espaço de Descrição

- Um indutor divide o espaço de descrição em regiões.
 - Cada região é rotulada com uma classe.
- Exemplo: para 2 atributos X_1 e X_2 ,
 - if $X_1 < 5$ and $X_2 < 8$ then classe "o" else classe "+",
divide o espaço em duas regiões.



Espaço de Descrição

- Para classificar um novo exemplo com $(X1, X2) = (2.5, 4)$, basta verificar em qual região ela se localiza e atribuir a classe associada àquela região (neste caso, classe "o").

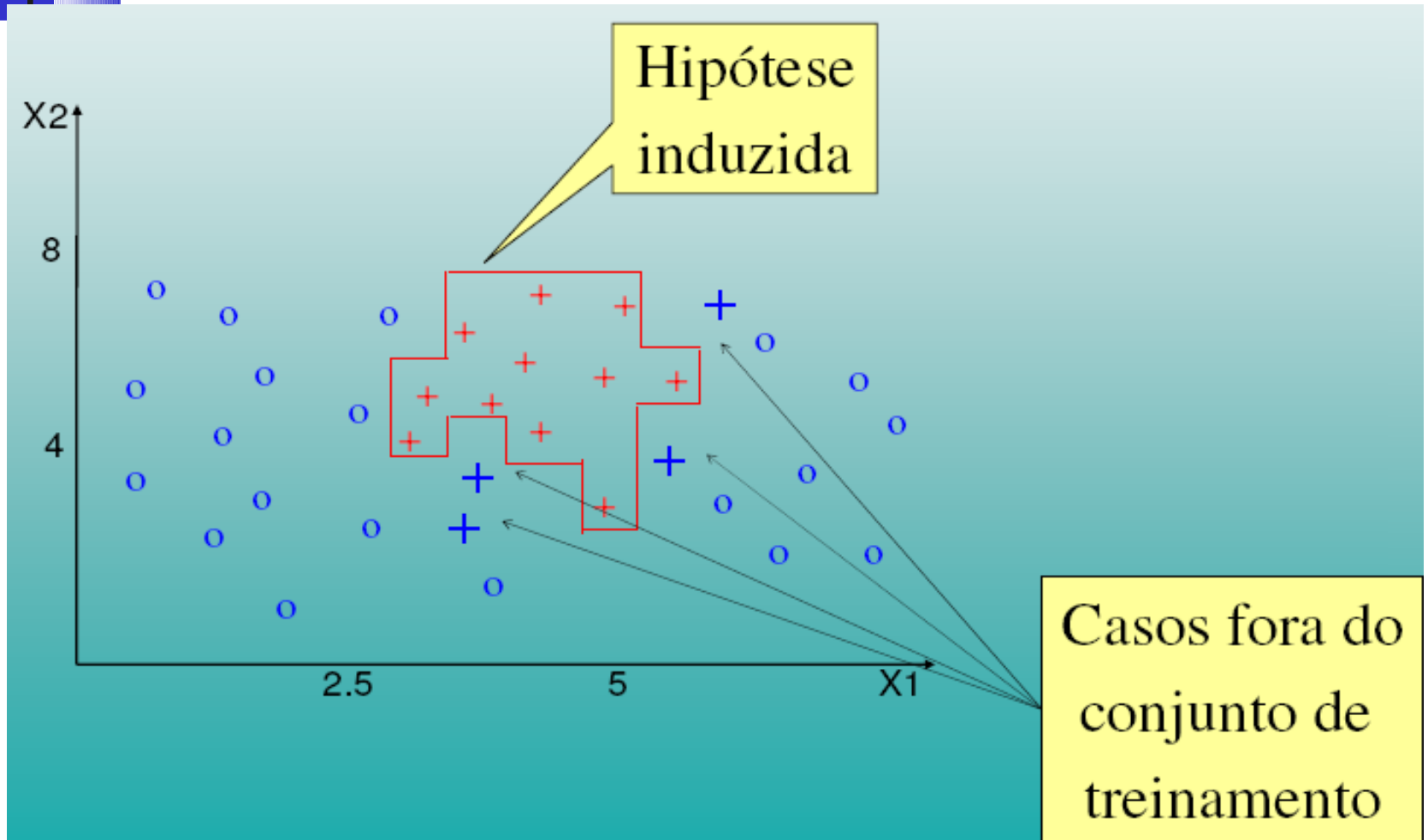




Overfitting

- Ocorre quando a hipótese extraída a partir dos dados é muito específica para o conjunto de treinamento.
- A hipótese apresenta um bom desempenho para o conjunto de treinamento, mas um desempenho ruim para os casos fora desse conjunto.

Overfitting - Exemplo





Underfitting

- A hipótese induzida apresenta um desempenho ruim tanto no conjunto de treinamento como de teste.
 - poucos exemplos representativos foram dados ao sistema de aprendizado.
 - o usuário predefiniu um tamanho muito pequeno para o classificador.
 - alto fator de poda para uma árvore de decisão.
 - número insuficiente de neurônios e conexões para uma rede neural.



Métodos para Avaliação de Classificadores

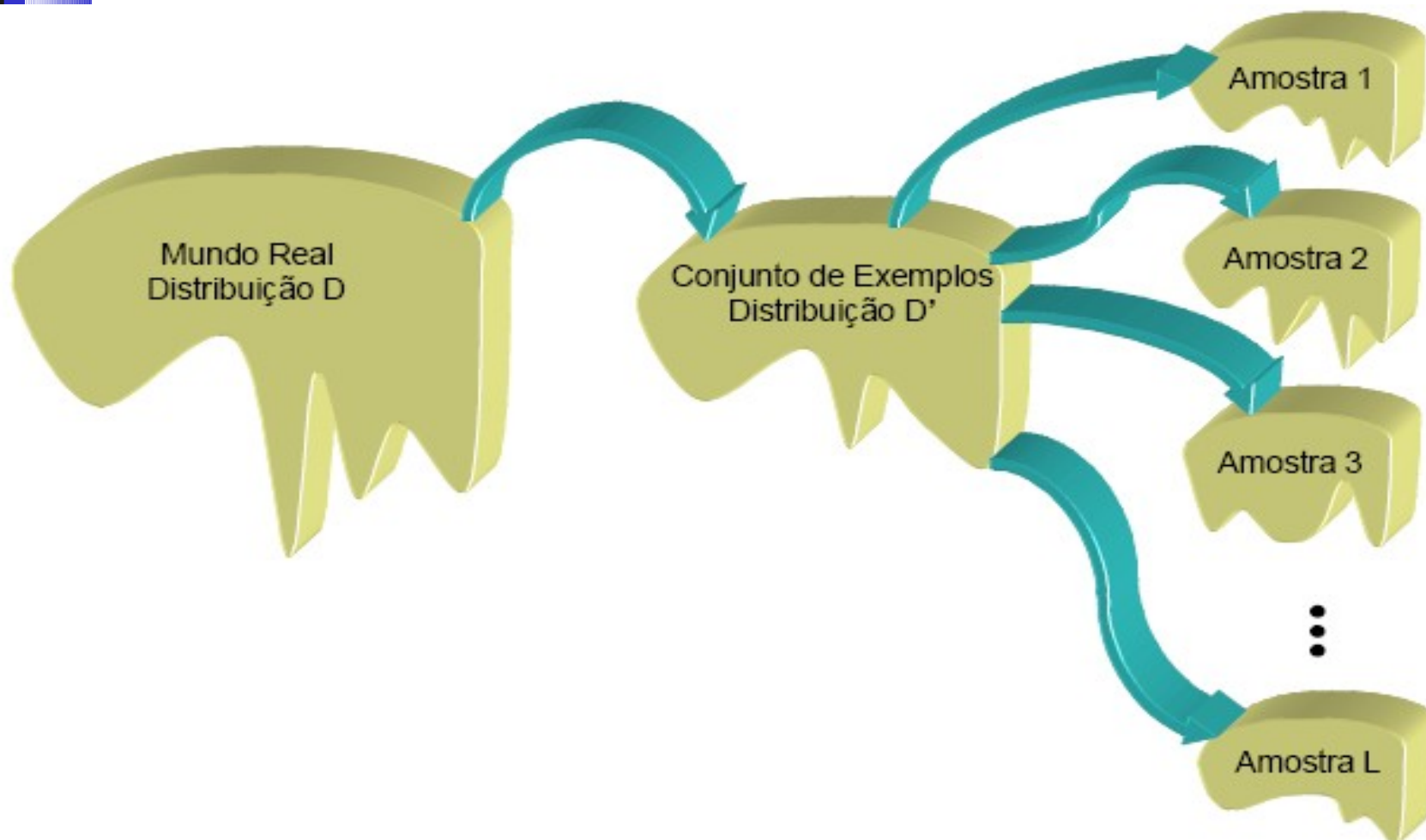
- Métodos de Amostragem:
 - Holdout
 - Amostragem aleatória
 - r-fold cross-validation
 - r-fold stratified cross-validation
 - Leave-one-out
 - Bootstrap



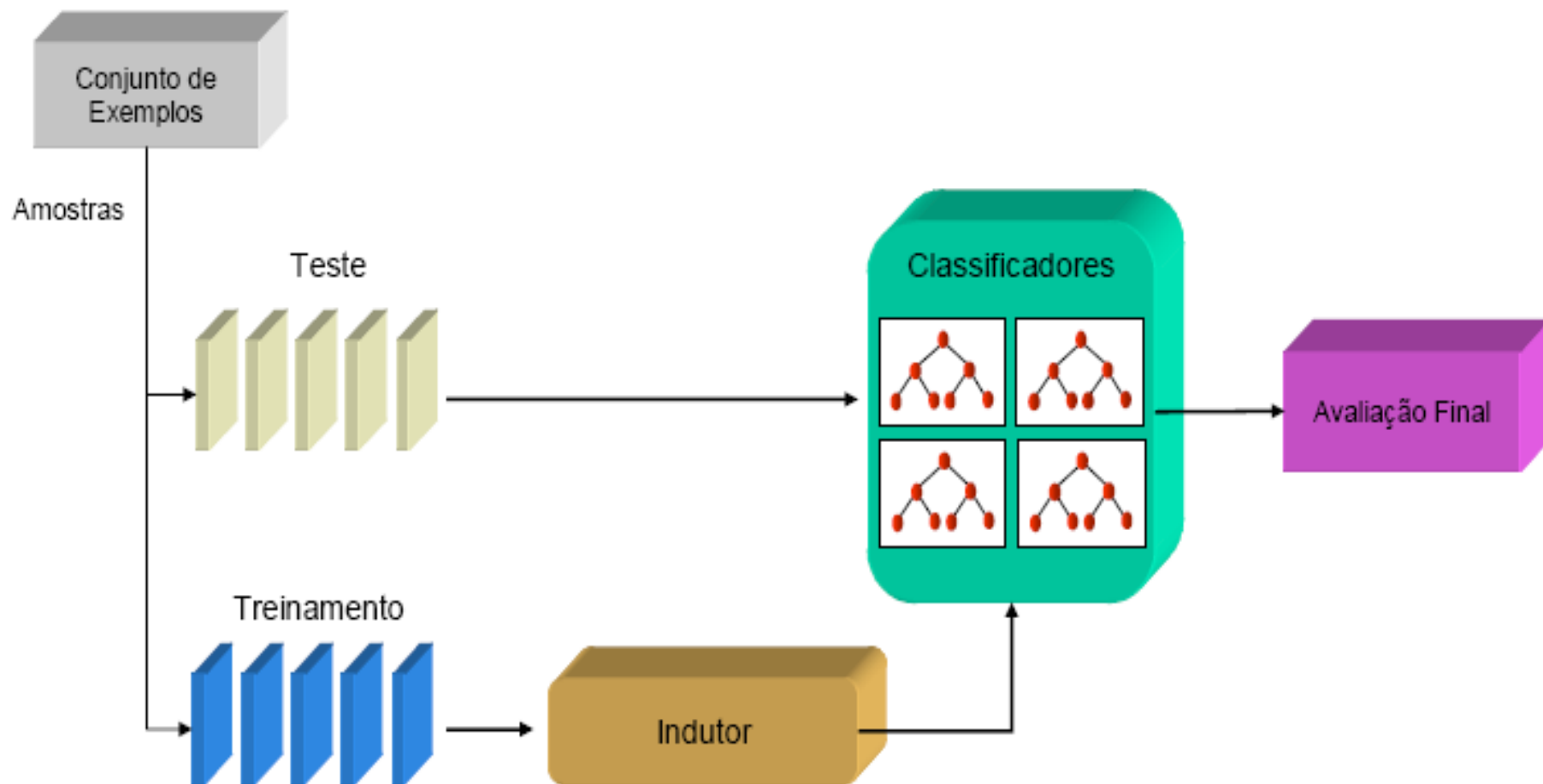
Métodos de Amostragem e Avaliação de Algoritmos

- Amostragem (resampling): metodologia de avaliação, utilizada para comparar dois algoritmos de classificação.
- Para se estimar o erro verdadeiro de um classificador, a amostra para teste deve ser aleatoriamente escolhida.
- Amostras não devem ser pré-selecionadas de nenhuma maneira.
- Para problemas reais, tem-se uma amostra de uma única população, de tamanho n , e a tarefa é estimar o erro verdadeiro para essa população.

Métodos de Amostragem



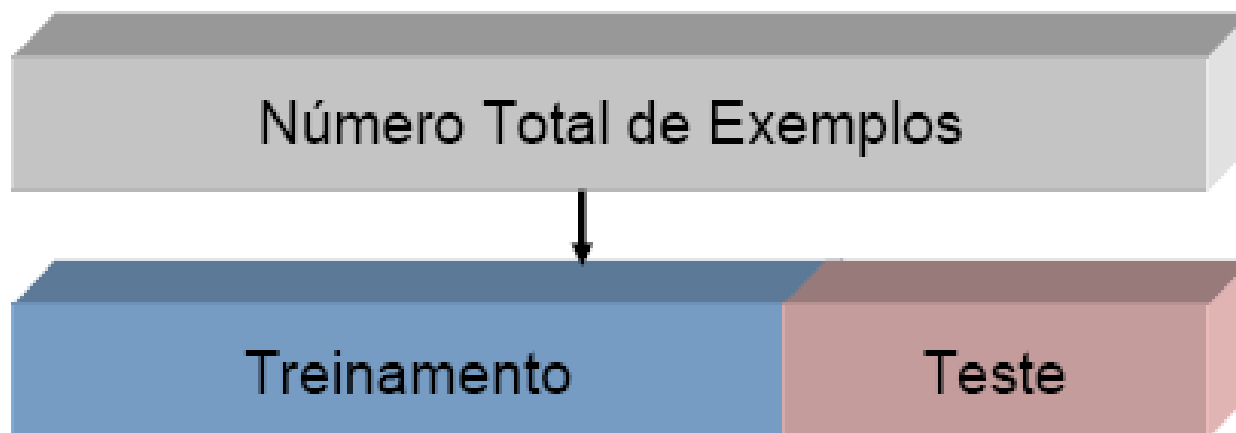
Métodos de Amostragem





Holdout

- Este método divide os exemplos em uma porcentagem de exemplos p para treinamento e $(1-p)$ para teste, considerando normalmente $p > 1/2$.
- Valores típicos são $p = 2/3$ e $(1-p) = 1/3$, embora não existam fundamentos teóricos sobre estes valores.

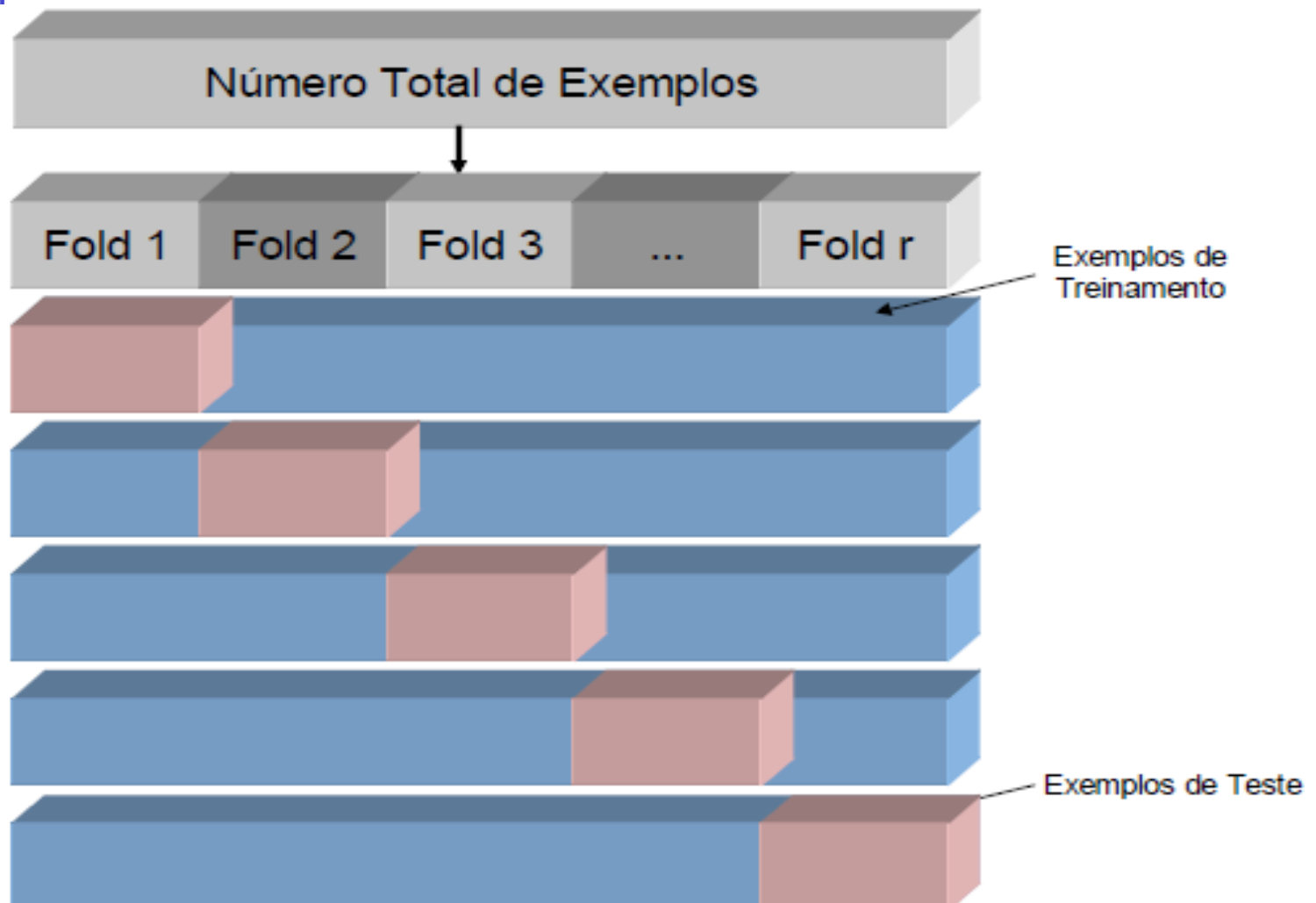




Amostragem Aleatória

- Na amostragem aleatória, L hipóteses, $L \ll n$, são induzidas a partir de cada um dos L conjuntos de treinamento.
- O erro final é calculado como sendo a média dos erros de todas as hipóteses induzidas e calculados em conjuntos de teste independentes e extraídos aleatoriamente.
- Amostragem aleatória pode produzir melhores estimativas de erro que o método holdout.

r-fold cross-validation





r-fold cross-validation

- Os exemplos são aleatoriamente divididos em r partições (folds) de tamanho aproximadamente igual a (n/r) .
- Os exemplos de $(r-1)$ folds são usados no treinamento e os classificadores obtidos são testados com o fold remanescente.
- O processo é repetido r vezes, e a cada repetição um fold diferente é usado para teste. O erro do cross-validation é a média dos erros dos r folds.



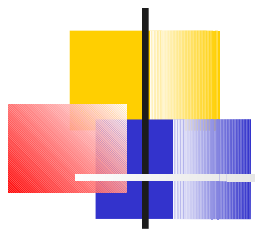
r-fold stratified cross-validation

- É similar ao cross-validation, mas no processo de geração dos folds a distribuição das classes no conjunto de exemplos é levada em consideração durante a amostragem.
- Por exemplo, se o conjunto de exemplos tiver duas classes com uma distribuição de 80% para uma classe e 20% para a outra, cada fold também terá essa proporção.



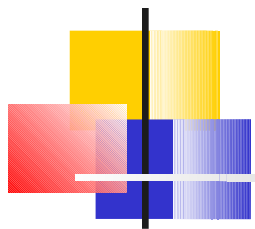
Leave-one-out

- Para um conjunto de exemplos de tamanho n , um classificador é gerado usando $n-1$ exemplos, e testado no exemplo remanescente.
- O processo é repetido n vezes, utilizando cada um dos n exemplos para teste.
 - O erro é a soma dos erros dos testes para cada exemplo dividido por n .
 - Caso especial de cross-validation.
- Computacionalmente caro e usado apenas quando o conjunto de exemplos é pequeno.



Bootstrap

- Funciona melhor que o cross-validation para conjuntos muito pequenos.
- Consiste em repetir diversas vezes o processo inteiro de classificação.
 - Cada experimento é baseado em um conjunto de treinamento novo, obtido por amostragem com reposição do conjunto de dados original.



Bootstrap

- Caso mais simples do bootstrap:
 - O conjunto de treinamento é formado por n exemplos (mesmo tamanho do conjunto de exemplos original) extraídos aleatoriamente com reposição.
 - Os exemplos que não aparecem no conjunto de treinamento são colocados no conjunto de teste.

Bootstrap

Conjunto Completo
de Exemplos

E_1 E_2 E_3 E_4 E_5

E_3 E_1 E_3 E_3 E_5



E_2 E_4

E_5 E_5 E_3 E_1 E_2



E_4

E_5 E_1 E_5 E_2 E_1



E_3 E_4

\vdots

E_4 E_4 E_4 E_1 E_4



E_2 E_3 E_5

Conjuntos de Treinamento

Conjuntos de Teste



Matriz de Confusão

- Oferece uma medida da eficácia do classificador, mostrando o número de classificações corretas versus as classificações previstas para cada classe.

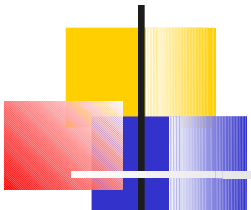
Classe	prevista C_1	prevista C_2	...	prevista C_k
real C_1	$M(C_1, C_1)$	$M(C_1, C_2)$...	$M(C_1, C_k)$
real C_2	$M(C_2, C_1)$	$M(C_2, C_2)$...	$M(C_2, C_k)$
\vdots	\vdots	\vdots	\ddots	\vdots
real C_k	$M(C_k, C_1)$	$M(C_k, C_2)$...	$M(C_k, C_k)$

Matriz de Confusão

Classe	prevista C_1	prevista C_2	...	prevista C_k
real C_1	$M(C_1, C_1)$	$M(C_1, C_2)$...	$M(C_1, C_k)$
real C_2	$M(C_2, C_1)$	$M(C_2, C_2)$...	$M(C_2, C_k)$
\vdots	\vdots	\vdots	\ddots	\vdots
real C_k	$M(C_k, C_1)$	$M(C_k, C_2)$...	$M(C_k, C_k)$

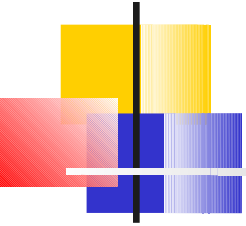
$$M(C_i, C_j) = \sum_{\{ \forall (x, y) \in T : y = C_i \}} \|h(x) = C_j\|$$

Matriz de Confusão - Exemplo



Classe Prevista	Classe Verdadeira		
	1	2	3
1	25	10	0
2	0	40	0
3	5	0	20

Matriz de Confusão para 2 Classes

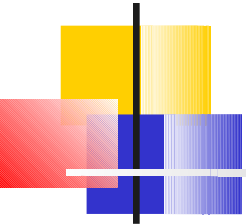


Classe Prevista	Classe Verdadeira	
	P	N
P	70	40
N	30	60



		Classe Verdadeira	
		P	N
Classe Prevista	P	VP	FP
	N	FN	VN

Matriz de Confusão para 2 Classes



- Medidas de erro:

$$\text{Taxa de FP} = \frac{FP}{FP + VN}$$

(alarmes falsos)

Erro do tipo I

		Classe Verdadeira	
		P	N
Classe Prevista	P	VP	FP
	N	FN	VN

$$\text{Taxa de FN} = \frac{FN}{VP + FN}$$

Erro do tipo II

		Classe Verdadeira	
		P	N
Classe Prevista	P	VP	FP
	N	FN	VN



Exemplo

- Matrizes de confusão para 3 classificadores:

		Classe Verdadeira	
		P	N
Classe Prevista	P	20	15
	N	30	35

Classificador 1
TFN = 0.6
TFP = 0.3

		Classe Verdadeira	
		P	N
Classe Prevista	P	70	50
	N	30	50

Classificador 2
TFN = 0.3
TFP = 0.5

		Classe Verdadeira	
		P	N
Classe Prevista	P	60	20
	N	40	80

Classificador 3
TFN = 0.4
TFP = 0.2

Exercício

		Classe Verdadeira	
		P	N
Classe Prevista	P	25	10
	N	45	60

Classificador 1
TFN =
TFP =

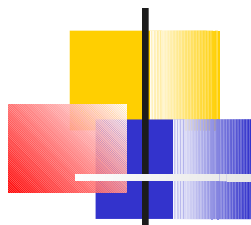
		Classe Verdadeira	
		P	N
Classe Prevista	P	70	20
	N	15	30

Classificador2
TFN =
TFP =

		Classe Verdadeira	
		P	N
Classe Prevista	P	70	95
	N	30	5

Classificador 3
TFN =
TFP =

Estimando a Taxa de Erro de um Classificador



- Etapa 1) Compute a taxa de erro (média) do classificador usando uma técnica de amostragem.
- Suponha que:
 - A tarefa é classificar o conjunto de dados Iris (com 4 atributos e 150 exemplos).
 - Em “Machine Learning Repository”:
<http://archive.ics.uci.edu/ml/datasets/Iris>
 - Foi utilizada a técnica 10-fold-cross-validation.
 - Nesse caso $n = 150$.
 - A média de erros calculada é designada por $\hat{\epsilon}$.





Estimando a Taxa de Erro de um Classificador

- Etapa 2) Compute um intervalo de confiança para essa estimativa de erro.
 - O erro padrão para essa estimativa é:

$$SE = \sqrt{\frac{\hat{\epsilon} \cdot (1 - \hat{\epsilon})}{n}}$$

- Um intervalo de confiança $(1 - \alpha)$ para o erro verdadeiro é:

$$\hat{\epsilon} - z_{\alpha/2}SE \leq \epsilon \leq \hat{\epsilon} + z_{\alpha/2}SE$$

- Para um intervalo de confiança de 95%, $Z_{0,025} = 1.96$, então:

$$\hat{\epsilon} - 1.96SE \leq \epsilon \leq \hat{\epsilon} + 1.96SE.$$



Estimando a Taxa de Erro de um Classificador

- Exemplo: suponha que o classificador apresentou erro médio = 10%, quando avaliado pela técnica 10-fold cross-validation.
- Portanto, para $n = 150$ e $\hat{e} = 0.10$

$$SE = \sqrt{\frac{\hat{e}(1 - \hat{e})}{n}} = \sqrt{\frac{0.10(1 - 0.10)}{150}} = 0.0245$$

- Assim, com 95% de confiança, o erro verdadeiro vai estar entre $0.10 - 1.96 \times 0.0245 = 0.052$ e $0.10 + 1.96 \times 0.0245 = 0.148$.



Leituras

- **Obrigatória:**

- REZENDE, S. (Ed.) Sistemas Inteligentes - Fundamentos e Aplicações. Manole, 2003.
 - Capítulo 4: Conceitos sobre Aprendizado de Máquina.

- **Complementar:**

- DIETTERICH, T.G. 1998. Approximate statistical tests for comparing supervised classification learning algorithms. Neural Computation, 10(7), 1895-1924, 1998. PDF Disponível no Col.]