

Aprendizado de Máquinas

Objetivo

- A área de aprendizado de máquina preocupa-se em construir programas que automaticamente melhorem seu desempenho com a experiência.

Conceito

- AM estuda métodos computacionais para adquirir novos conhecimentos, novas habilidades e novos meios de organizar o conhecimento já existentes.

O que é aprendizado

- Ganhar conhecimento, habilidades aprendendo, por instrução ou experiência
- Modificação do comportamento pela experiência
- Uma máquina aprende quando muda sua estrutura de programa ou dados de tal maneira que espera-se melhorarias de sua performance no futuro.

Definição

- Um programa de computador aprende da experiência E com referência a algum tipo de tarefa T e medida de performance P . Se sua performance na tarefa T , medida por P , melhora com a experiência E .

Motivação (1)

- Dado um conjunto de objetos, colocar os objetos em grupos baseados na similaridade entre eles



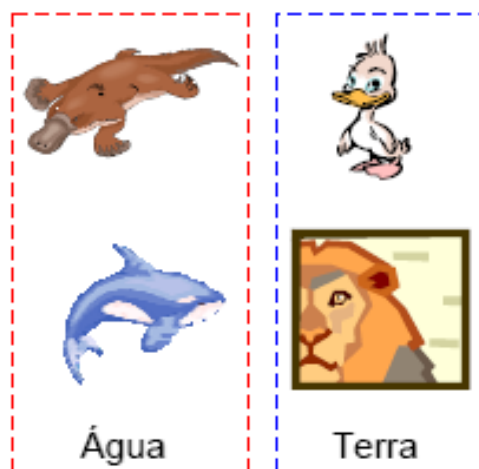
Motivação (1)

- Dado um conjunto de objetos, colocar os objetos em grupos baseados na similaridade entre eles



Motivação (1)

- Dado um conjunto de objetos, colocar os objetos em grupos baseados na similaridade entre eles



Motivação (1)

- Dado um conjunto de objetos, colocar os objetos em grupos baseados na similaridade entre eles



Motivação (2)

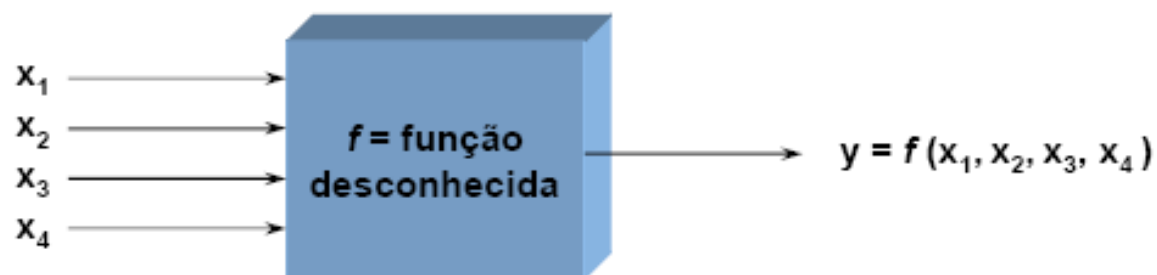
□ Dados pares $(x, f(x))$, inferir f

x	f(x)
1	1
2	4
3	9
4	16
5	?

Dada uma amostra finita, é frequentemente impossível determinar a verdadeira função f .

Abordagem: Encontre um padrão (*hipótese*) nos exemplos de treinamento e assuma que o padrão se repita para casos futuros também

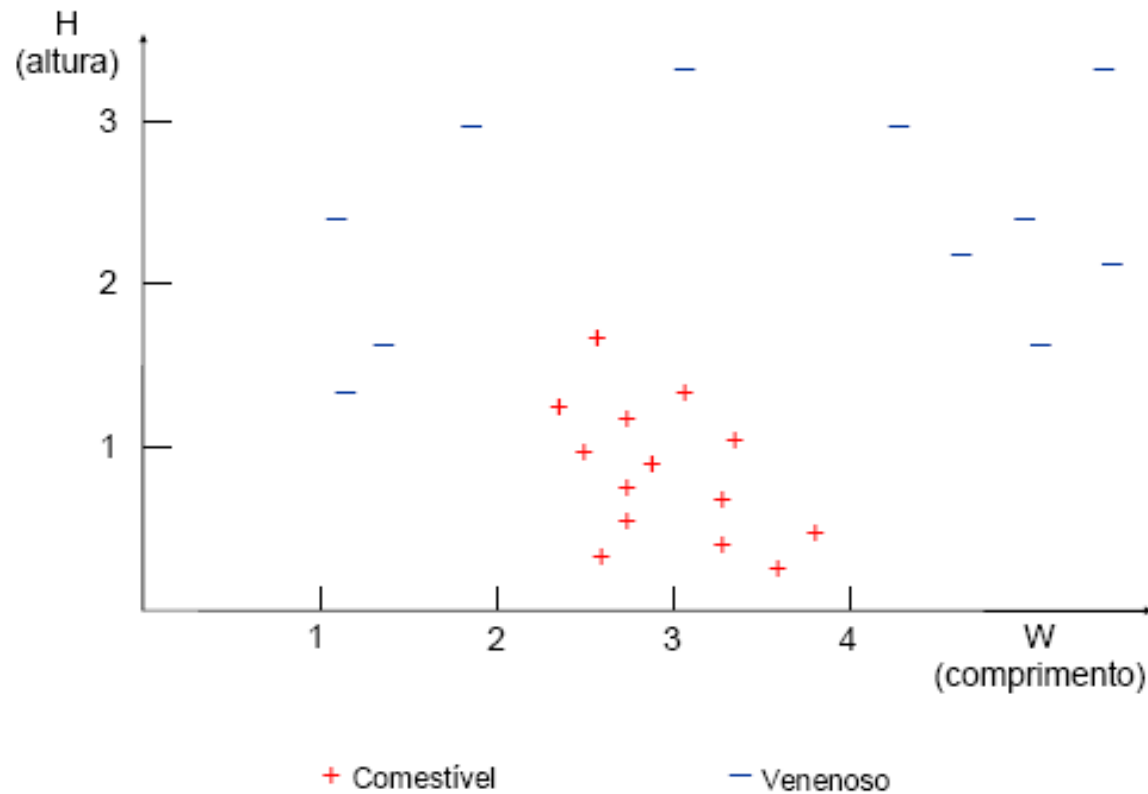
Motivação (2)



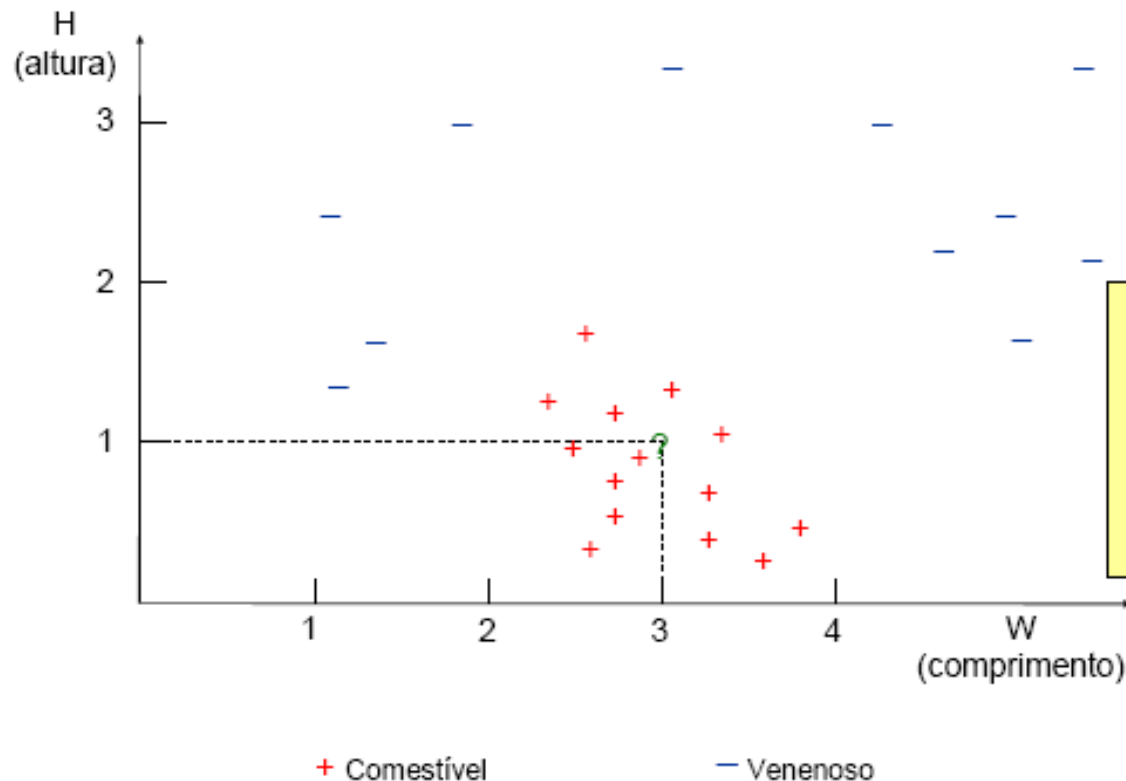
Exemplo	x_1	x_2	x_3	x_4	Y
E_1	0	1	1	0	0
E_2	0	0	0	0	0
E_3	0	0	1	1	1
E_4	1	0	0	1	1
E_5	0	1	1	0	0
E_6	1	1	0	0	0
E_7	0	1	0	1	0

□ $f: (X_1 \times X_2 \times X_3 \times X_4) \rightarrow Y$

Exemplo: Cogumelos Comestíveis x Venenosos

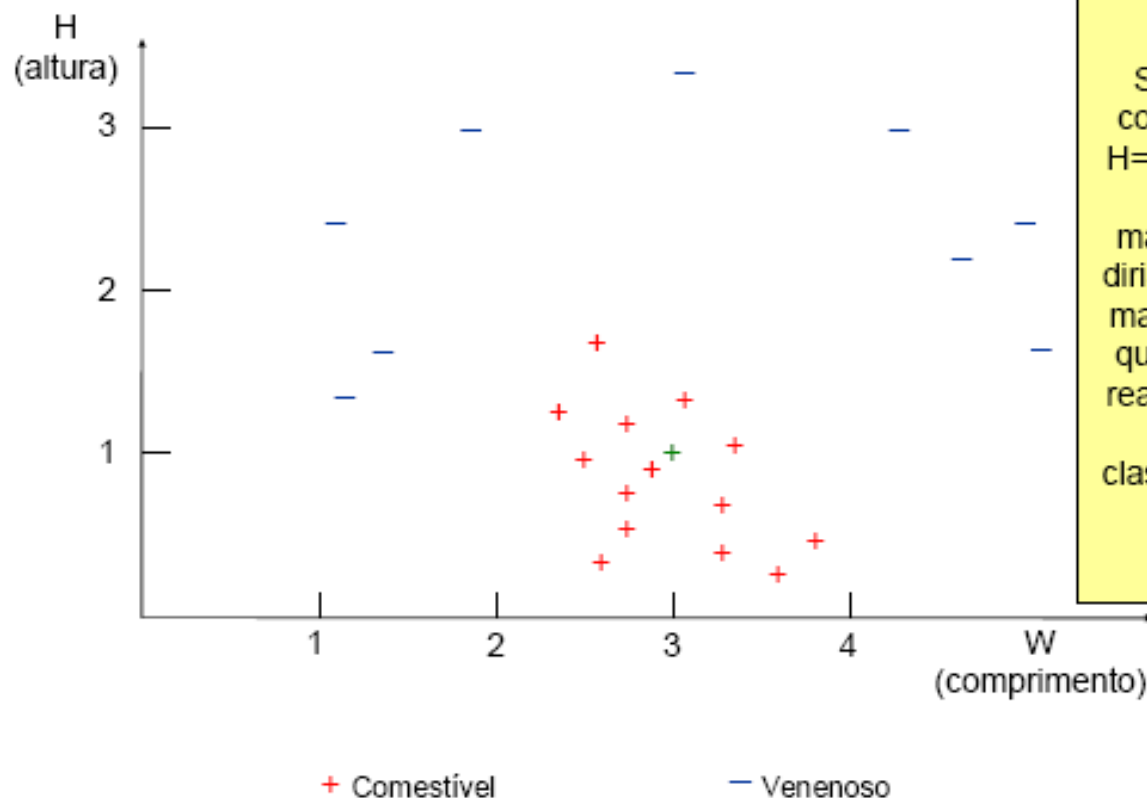


Exemplo: Cogumelos Comestíveis x Venenosos



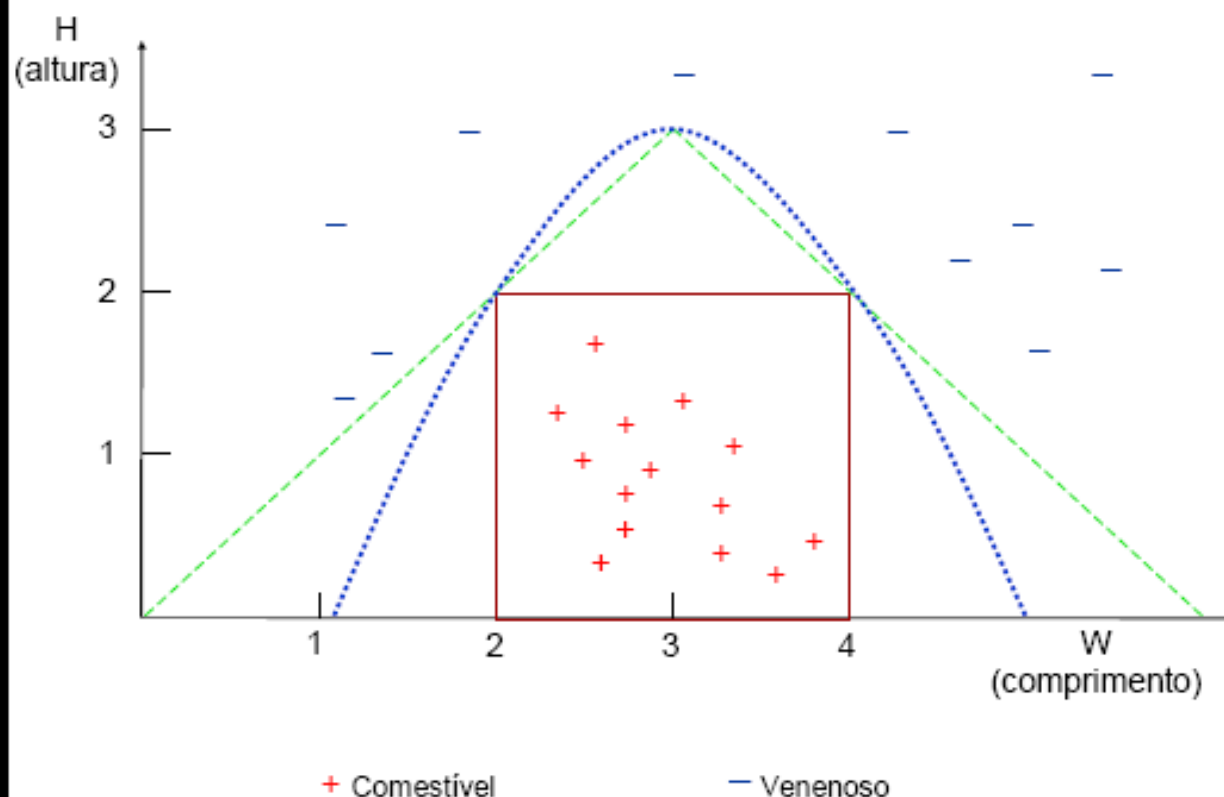
Suponha um novo cogumelo com $W=3$, $H=1$. Ele é comestível ou venenoso?

Exemplo: Cogumelos Comestíveis x Venenosos



Suponha um novo cogumelo com $W=3$, $H=1$. Ele é comestível ou venenoso? A maioria das pessoas diria que é comestível, mas não há garantias que o cogumelo seja realmente comestível. Assim esta classificação é apenas uma *hipótese*

Exemplo: Cogumelos Comestíveis x Venenosos



Hipótese 1:
if $2 < W$ and $W < 4$ and $H < 2$
then comestível
else venenoso

Hipótese 2:
if $H > W$
then venenoso
else if $H > 6 - W$
then venenoso
else comestível

Hipótese 3:
if $H < 3 - (W - 3)^2$
then comestível
else venenoso

Hierarquia do Aprendizado



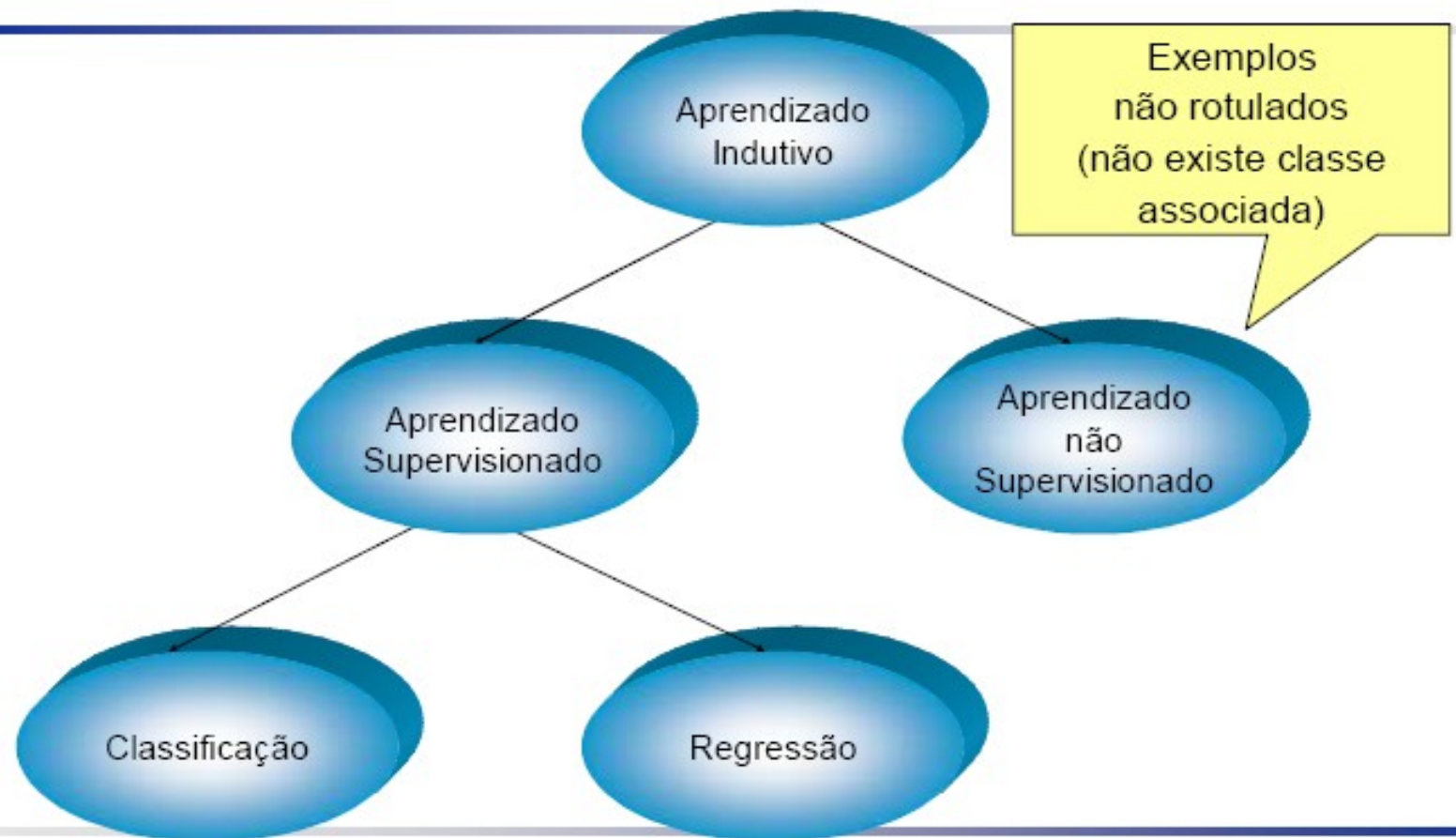
Hierarquia do Aprendizado



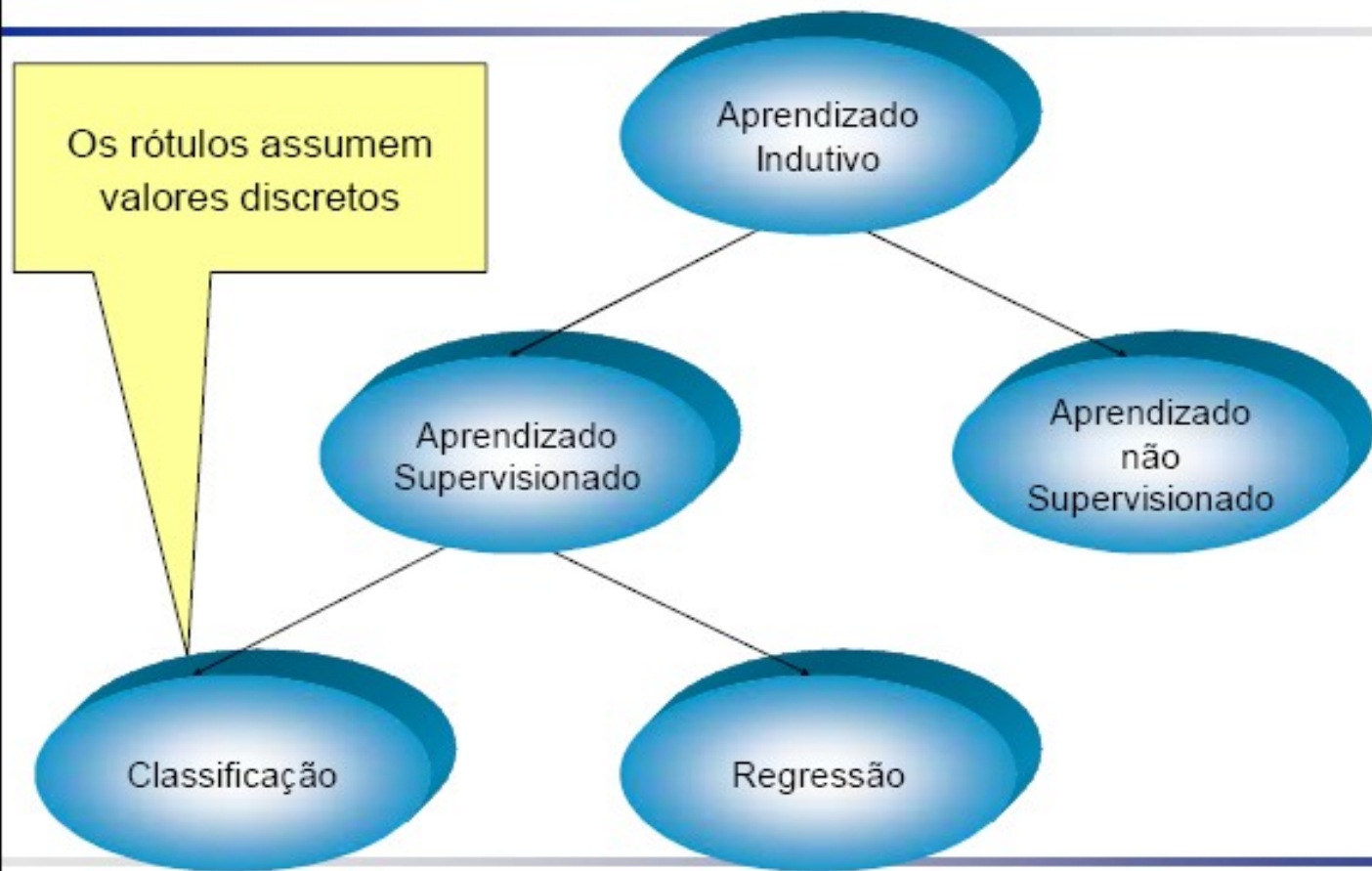
Hierarquia do Aprendizado



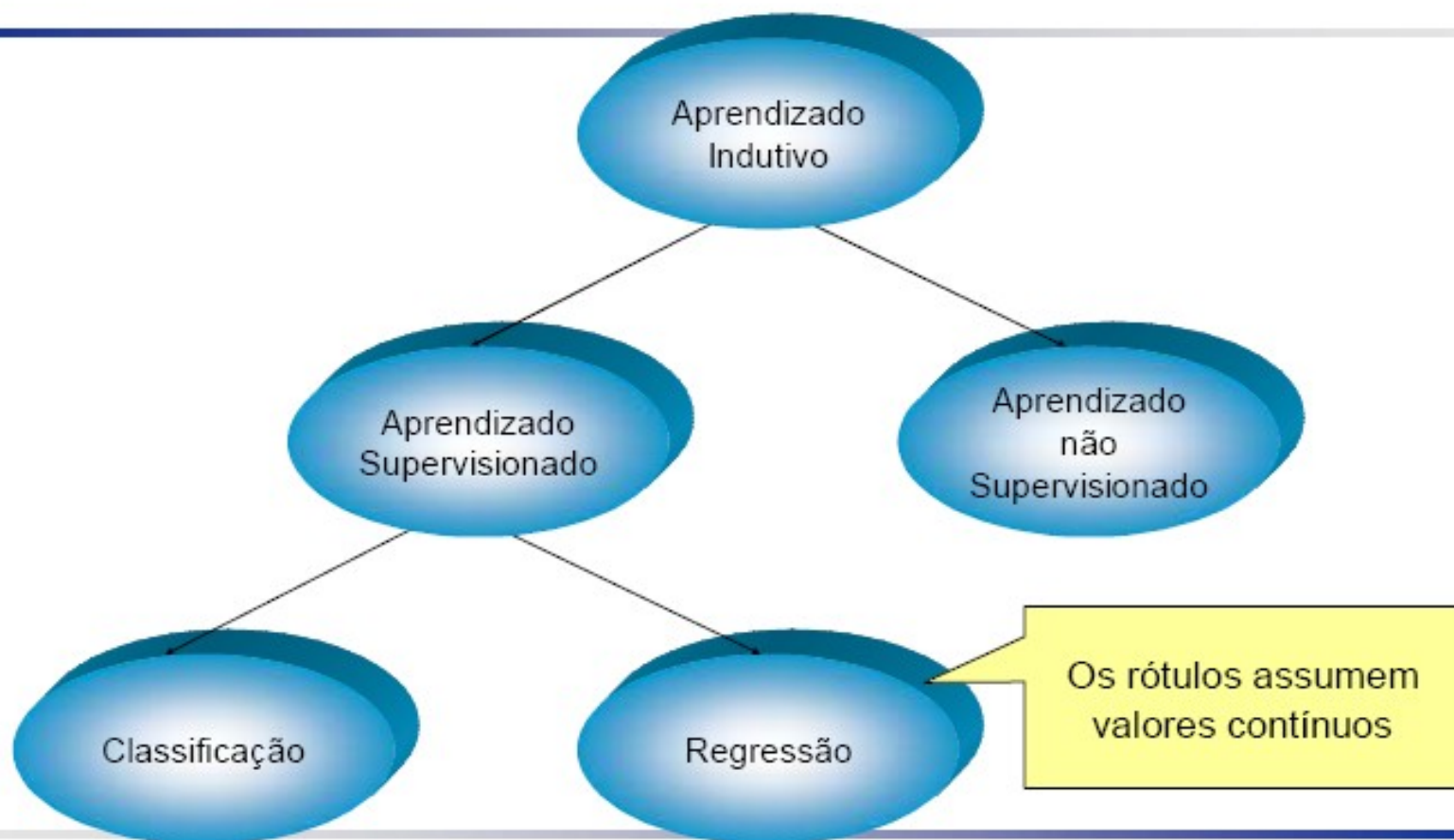
Hierarquia do Aprendizado



Hierarquia do Aprendizado



Hierarquia do Aprendizado



AS x AnS

□ Aprendizado Supervisionado

- Compreender o relacionamento entre os atributos e a classe
- Predizer a classe de novos exemplos o melhor possível

□ Aprendizado Não Supervisionado

- Encontrar representações úteis dos exemplos, tais como:
 - ❖ Encontrar agrupamentos (clusters)
 - ❖ Redução da dimensão
 - ❖ Encontrar as causas ou as fontes ocultas dos exemplos
 - ❖ Modelar a densidade dos exemplos

Tarefa de Classificação

- Cada exemplo pertence a uma classe pré-definida
- Cada exemplo consiste de:
 - Um atributo classe
 - Um conjunto de atributos preditores
- O objetivo é prever a classe do exemplo dado seus valores de atributos preditores.

Exemplo:

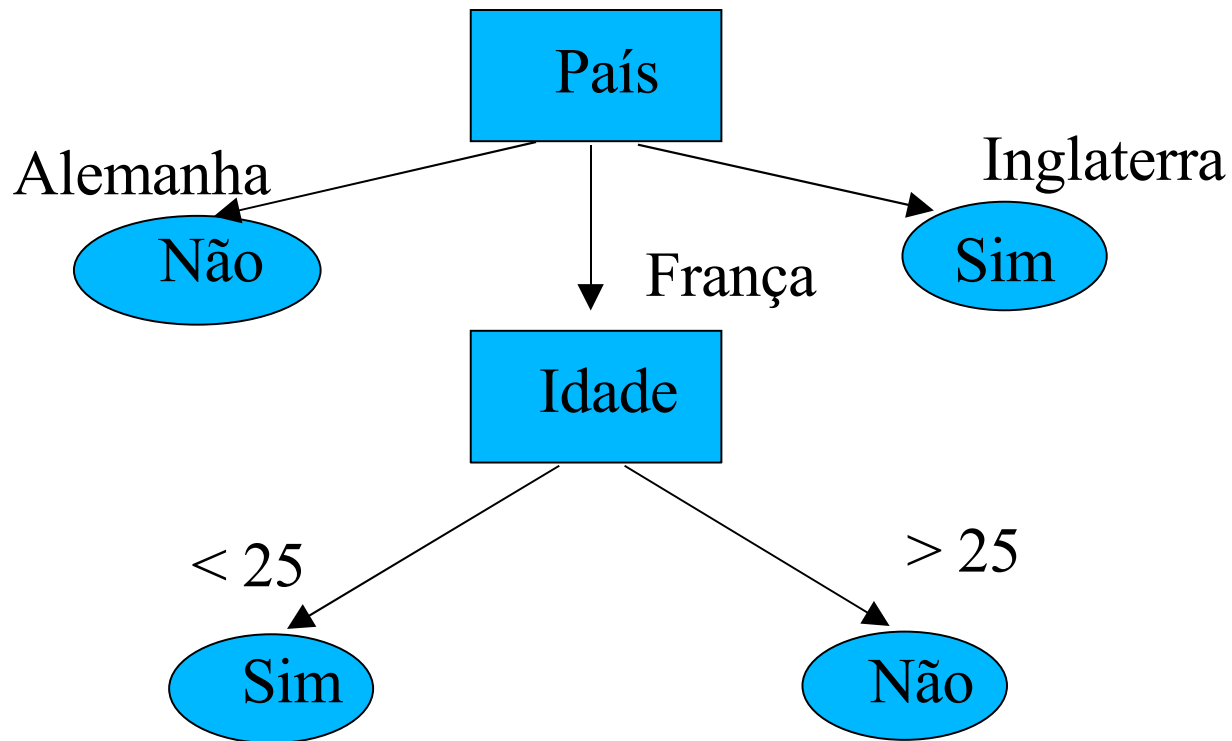
Extraído de Freitas & Lavington 98

- Uma editora internacional publica o livro “Guia de Restaurantes Franceses na Inglaterra” em 3 países: Inglaterra, França e Alemanha.
- A editora tem um banco de dados sobre clientes nesses 3 países, e deseja saber quais clientes são mais prováveis compradores do livro (para fins de mala direta direcionada).
 - Atributo meta: comprar (sim/não)
- Para coletar mais dados: enviar material de propaganda para uma amostra de clientes, registrando se cada cliente que recebeu a propaganda comprou ou não o livro.

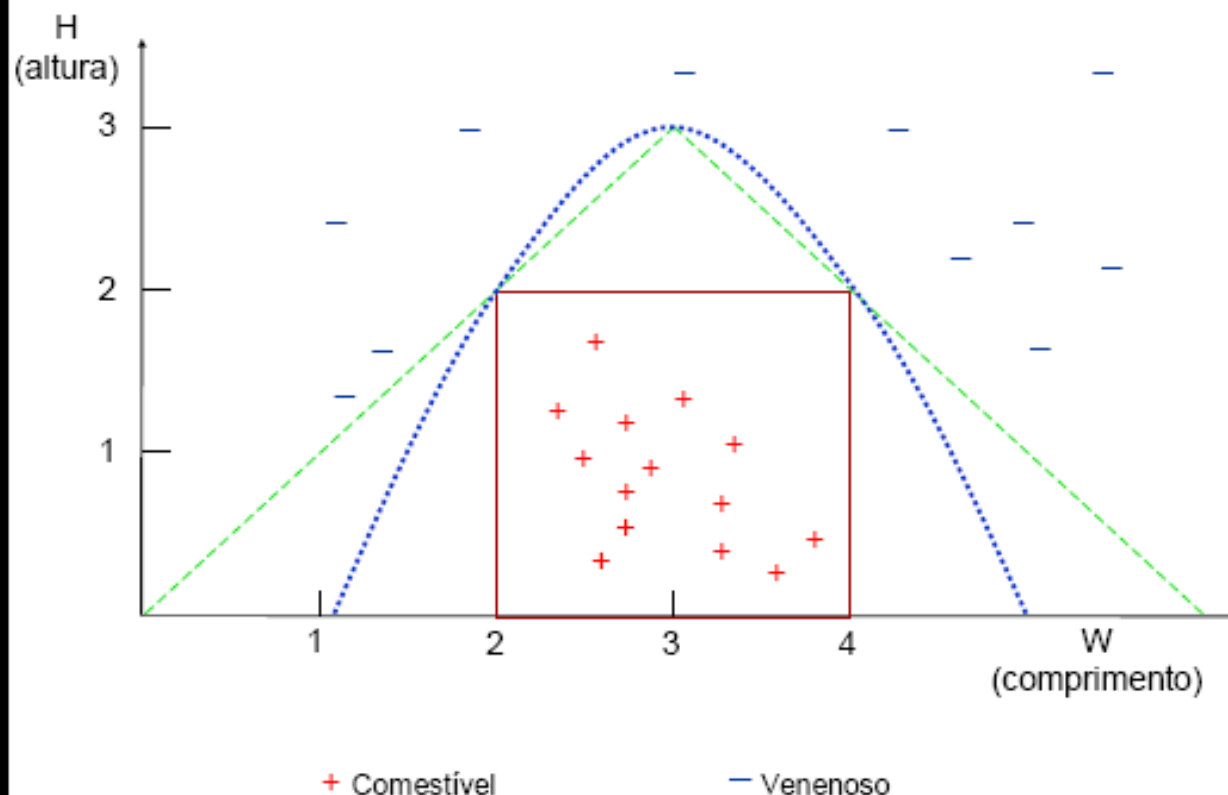
Exemplo de Classificação

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

Árvores de Decisão



Exemplo: Cogumelos Comestíveis x Venenosos



Hipótese 1:
if $2 < W$ and $W < 4$ and $H < 2$
then comestível
else venenoso

Hipótese 2:
if $H > W$
then venenoso
else if $H > 8 - W$
then venenoso
else comestível

Hipótese 3:
if $H < 3 - (W - 3)^2$
then comestível
else venenoso

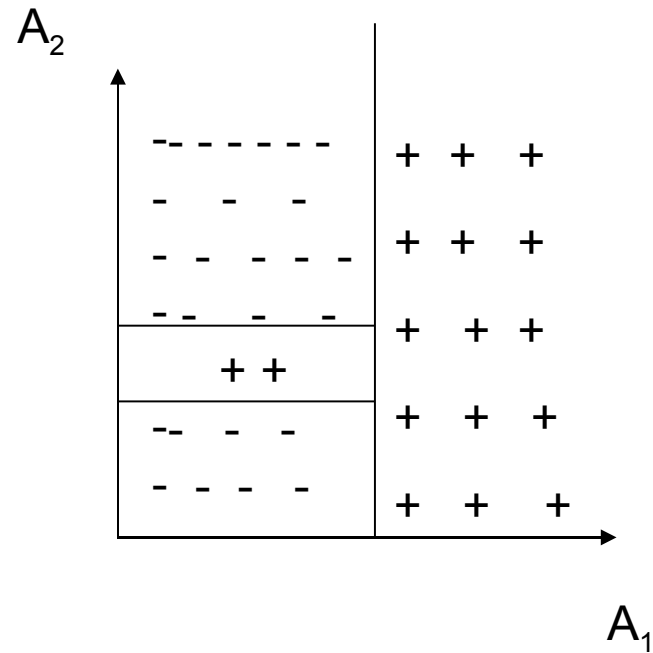
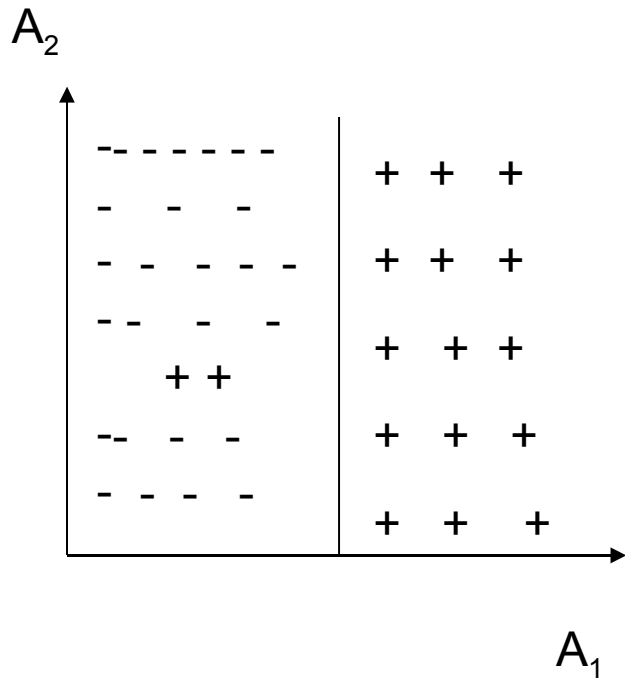
Terminologia

- Exemplo, um objeto, um caso, um registro, um tupla
- Atributo, variável, feature, característica
- Conjunto de treinamento, conjunto de teste
 - Aprendizado
 - Avaliação

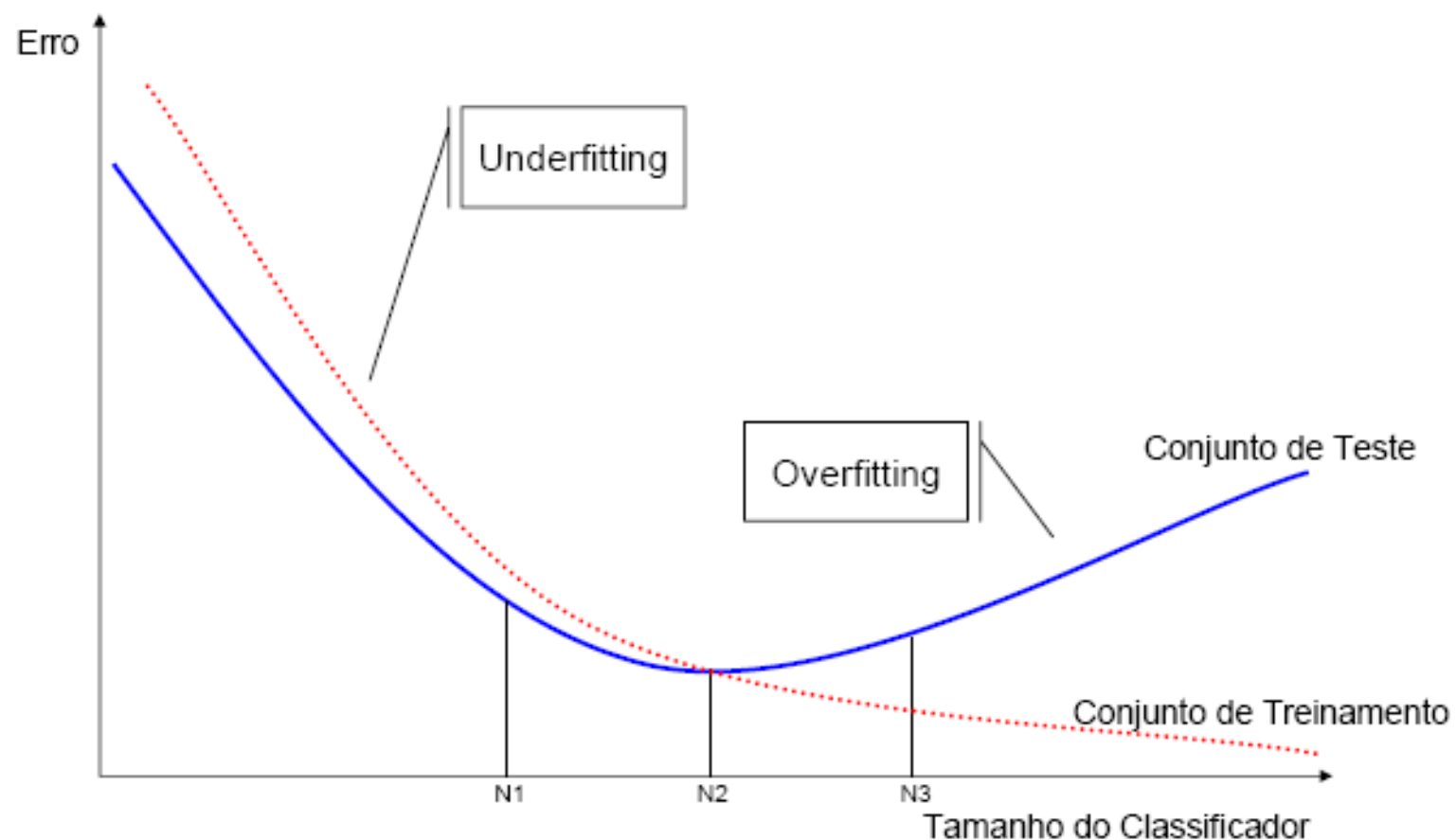
Hipótese do Aprendizado Indutivo

- A tarefa é não determinística
- Qualquer hipótese que aproxime bem o conceito alvo num conjunto de treinamento, suficientemente grande, aproximara o conceito alvo para exemplos não observados.

Overfitting e Underfitting (sobre-especialização)



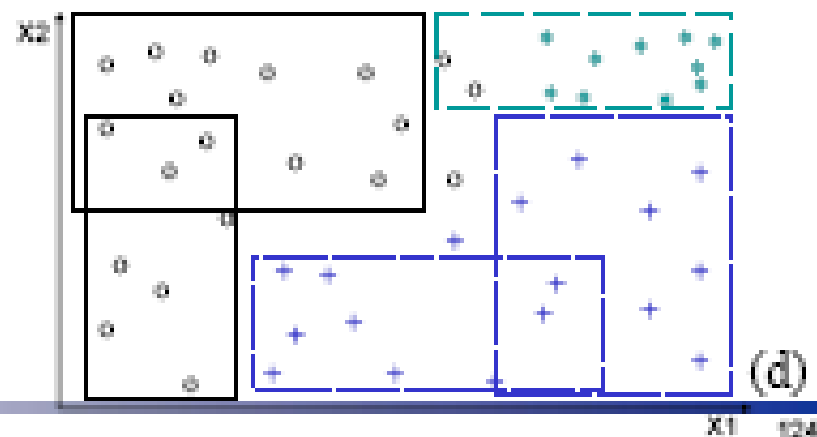
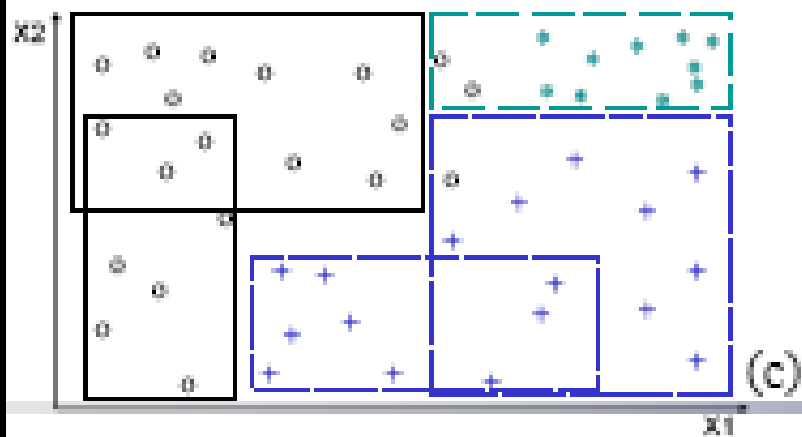
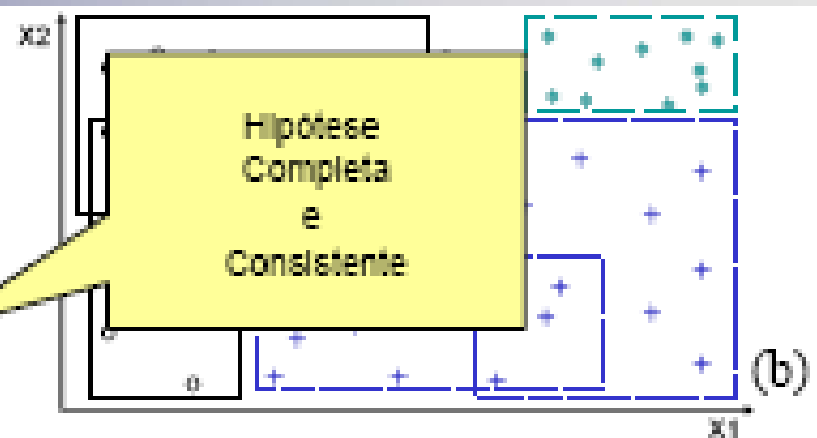
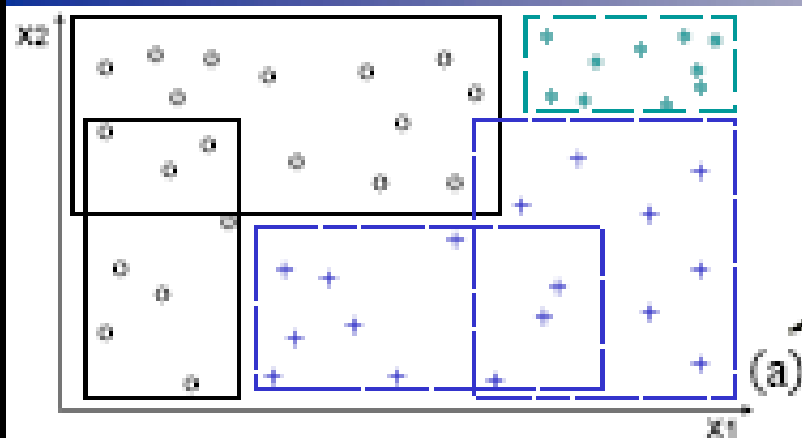
Relação entre o Tamanho do Classificador e o Erro



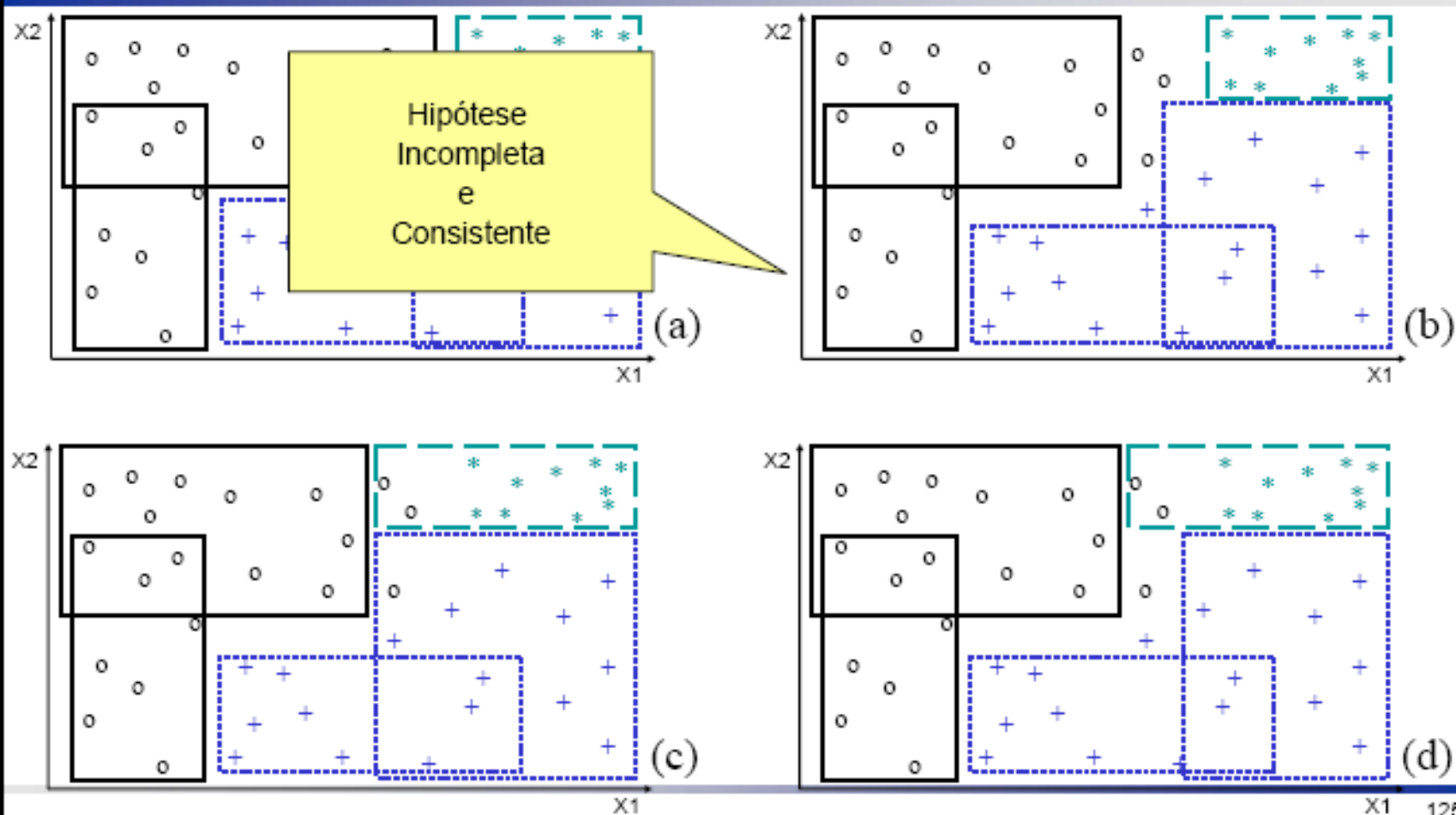
Consistência e Completude

- Depois de induzida, uma hipótese pode ser avaliada sobre
- consistência, se classifica corretamente os exemplos
- completude, se classifica todos os exemplos

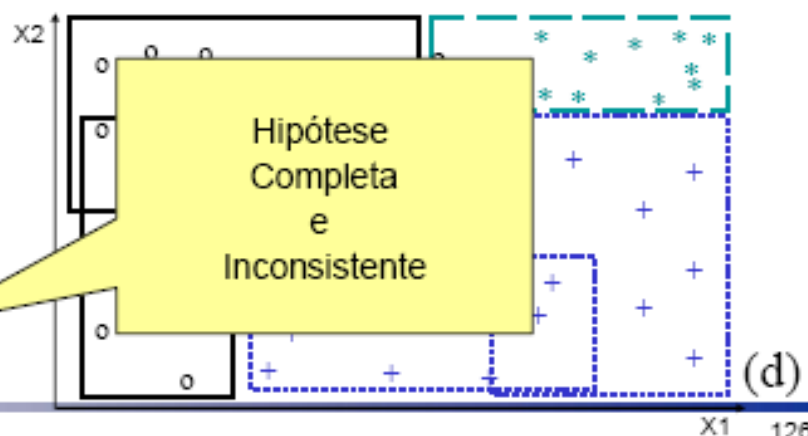
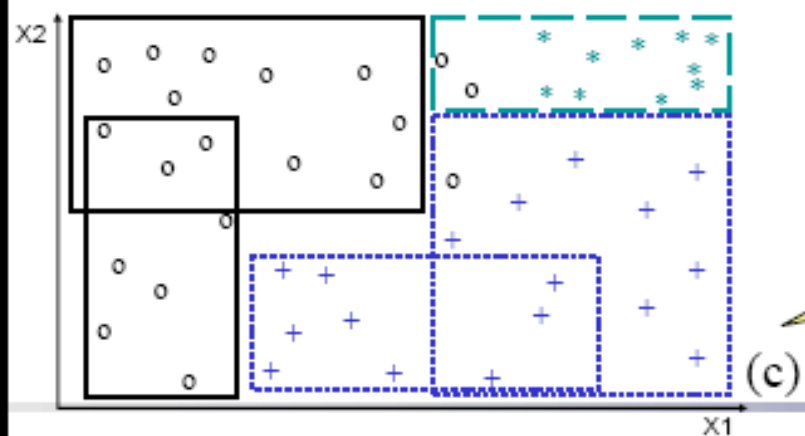
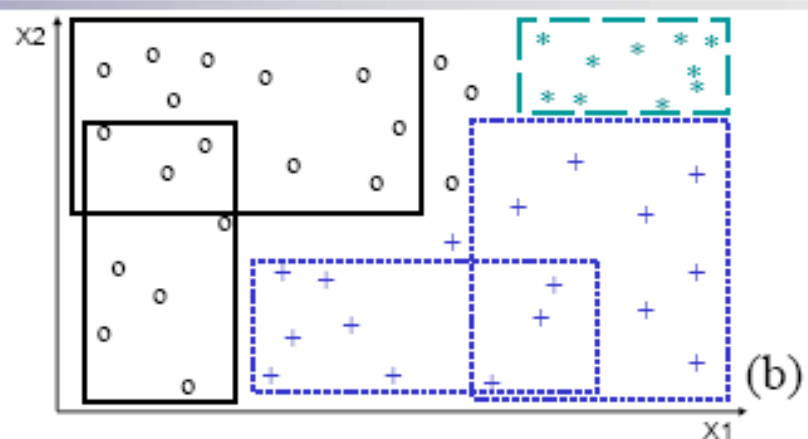
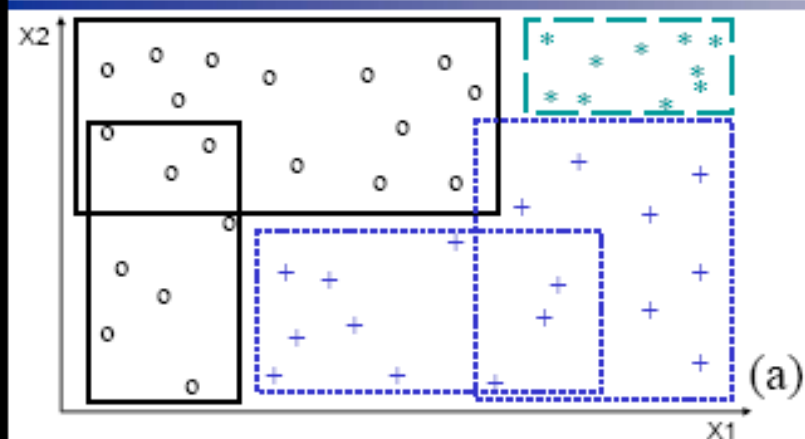
Relação entre Completude e Consistência



Relação entre Completude e Consistência

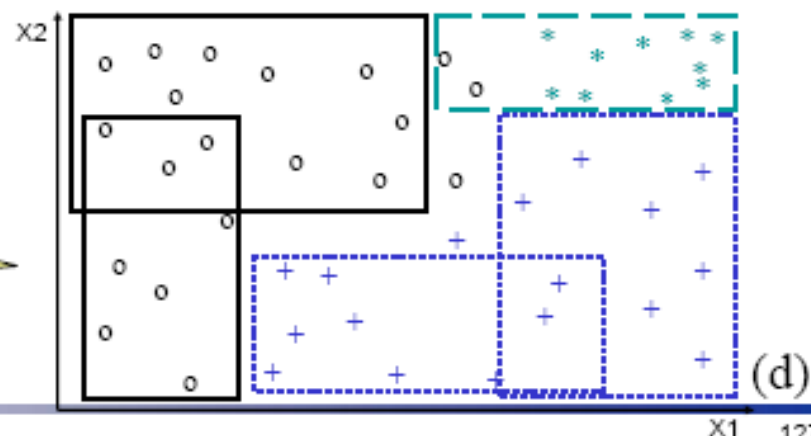
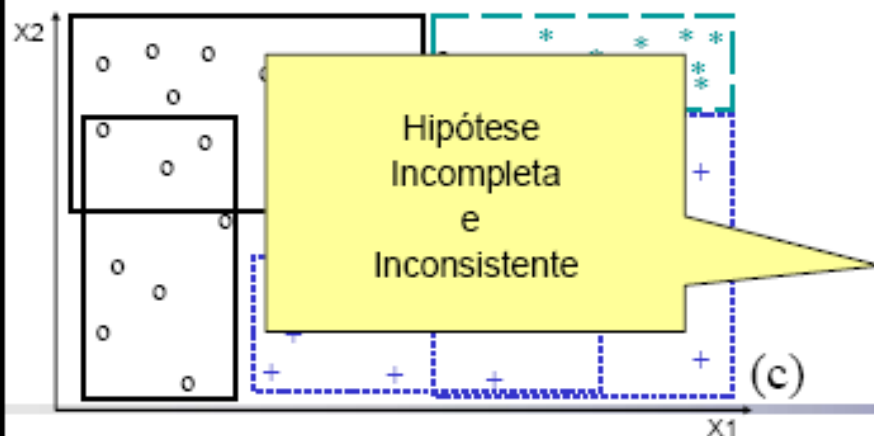
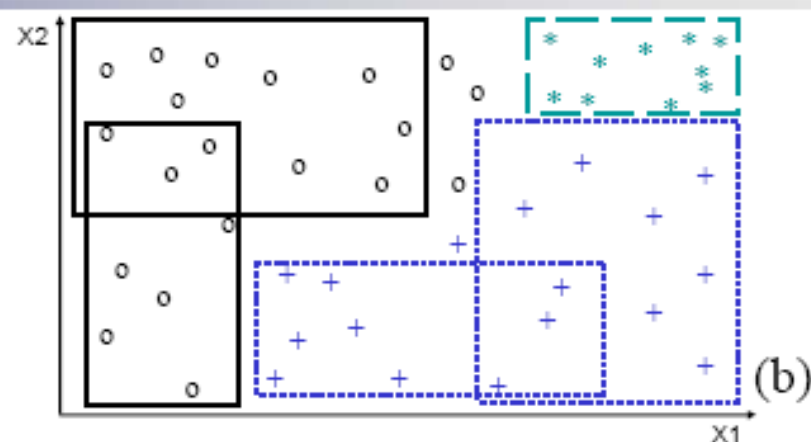
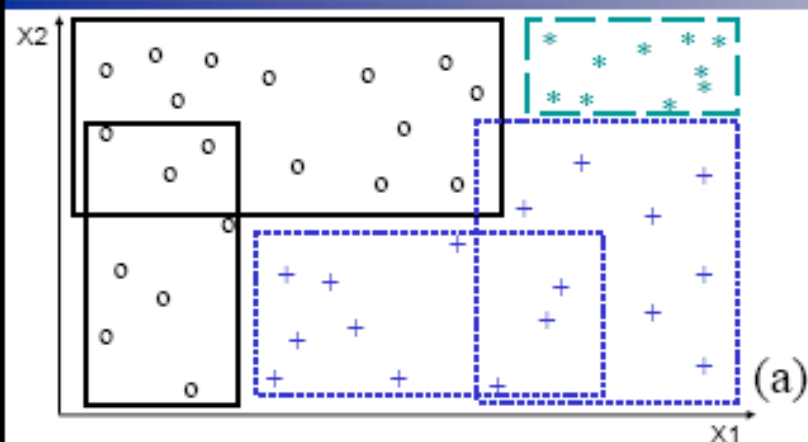


Relação entre Completude e Consistência



Hipótese
Completa
e
Inconsistente

Relação entre Completude e Consistência



Medindo a qualidade da predição

- Precisão, compreensível e interessante
- Acuracia = $\frac{\text{classificados corretamente}}{\text{total de exemplos}}$
- Erro = $1 - \text{Acuracia}$

Matriz de Confusão

- A **matriz de confusão** de uma hipótese h oferece uma medida efetiva do modelo de classificação, ao mostrar o número de classificações corretas *versus* as classificações preditas para cada classe, sobre um conjunto de exemplos T

Matriz de Confusão

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

Classe	predita C_1	predita C_2	...	predita C_k
verdadeira C_1	$M(C_1, C_1)$	$M(C_1, C_2)$	...	$M(C_1, C_k)$
verdadeira C_2	$M(C_2, C_1)$	$M(C_2, C_2)$	...	$M(C_2, C_k)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
verdadeira C_k	$M(C_k, C_1)$	$M(C_k, C_2)$	...	$M(C_k, C_k)$

Matriz de Confusão

- O número de acertos, para cada classe, se localiza na diagonal principal $M(C_i, C_i)$ da matriz
- Os demais elementos $M(C_i, C_j)$, para $i \neq j$, representam erros na classificação
- A matriz de confusão de um classificador ideal possui todos esses elementos iguais a zero uma vez que ele não comete erros

Matriz de Confusão para 2 Classes

Classe	predita C_+ predita C_-		Taxa de Erro da Classe	Taxa de Erro Total
verdadeira C_+	T_P	F_N	$\frac{F_N}{T_P + F_N}$	$\frac{F_P + F_N}{n}$
verdadeira C_-	F_P	T_N	$\frac{F_P}{F_P + T_N}$	

T_P = Verdadeiro Positivo (True Positive)

F_N = Falso Negativo (False Negative)

F_P = Falso Positivo (False Positive)

T_N = Verdadeiro Negativo (True Negative)

$n = (T_P + F_N + F_P + T_N)$

Métricas Derivadas da Matriz de Confusão para 2 Classes

- ❑ Confiabilidade positiva

$$\text{prel}(h) = \frac{T_P}{T_P + F_P}$$

- ❑ Confiabilidade negativa

$$\text{nrel}(h) = \frac{T_N}{T_N + F_N}$$

- ❑ Suporte

$$\text{sup}(h) = \frac{T_P}{n}$$

- ❑ Sensitividade

$$\text{sens}(h) = \frac{T_P}{T_P + F_N}$$

- ❑ Especificidade

$$\text{spec}(h) = \frac{T_N}{F_P + T_N}$$

- ❑ Precisão total

$$\text{tacc}(h) = \frac{T_P + T_N}{n}$$

- ❑ Cobertura

$$\text{cov}(h) = \frac{T_P + F_P}{n}$$

Prevalência de Classe

- Um ponto muito importante em AM refere-se ao desbalanceamento de classes em um conjunto de exemplos
- Por exemplo, suponha um conjunto de exemplos T com a seguinte distribuição de classes $\text{dist}(C1, C2, C3) = (99.00\%, 0.25\%, 0.75\%)$, com prevalência da classe $C1$
- Um classificador simples que classifique sempre novos exemplos como pertencentes à classe majoritária $C1$ teria uma precisão de 99,00% ($\text{maj-err}(T) = 1,00\%$)
- Isto pode ser indesejável quando as classes minoritárias são aquelas que possuem uma informação muito importante, por exemplo, supondo $C1$: paciente normal, $C2$: paciente com doença A e $C3$: paciente com doença B

Prevalência de Classe

- É importante estar ciente, quando se trabalha com conjuntos de exemplos desbalanceados, que é desejável utilizar uma medida de desempenho diferente da precisão
- Isto deve-se ao fato que a maioria dos sistemas de aprendizado é projetada para otimizar a precisão
- Com isso, normalmente os algoritmos apresentam um desempenho ruim se o conjunto de treinamento encontra-se fortemente desbalanceado, pois os classificadores induzidos tendem a ser altamente precisos nos exemplos da classe majoritária, mas freqüentemente classificam incorretamente exemplos das classes minoritárias
- Algumas técnicas foram desenvolvidas para lidar com esse problema, tais como a introdução de custos de classificação incorreta (explicada mais adiante), a remoção de exemplos redundantes ou prejudiciais ou ainda a detecção de exemplos de borda e com ruído

Custos de Erros

- Medir adequadamente o desempenho de classificadores, através da taxa de erro (ou precisão) assume um papel importante em AM, uma vez que o objetivo consiste em construir classificadores com baixa taxa de erro em novos exemplos
- Entretanto, ainda considerando o problema anterior contendo duas classes, se o custo de ter falsos positivos e falsos negativos não é o mesmo, então outras medidas de desempenho devem ser usadas
- Uma alternativa natural, quando cada tipo de classificação incorreta possui um custo diferente ou mesmo quando existe prevalência de classes, consiste em associar um custo para cada tipo de erro

Custos de Erros

- O custo $\text{cost}(C_i, C_j)$ é um número que representa uma penalidade aplicada quando o classificador faz um erro ao rotular exemplos, cuja classe verdadeira é C_i , como pertencentes à classe C_j , onde $i, j = 1, 2, \dots, k$ e k é o número de classes
- Assim, $\text{cost}(C_i, C_i) = 0$, uma vez que não constitui um erro e $\text{cost}(C_i, C_j) > 0$, $i \neq j$
- Em geral, os indutores assumem que $\text{cost}(C_i, C_j) = 1$, $i \neq j$, caso esses valores não sejam definidos explicitamente

Complexo

- ❑ É uma conjunção de disjunções dos atributos de teste, na forma:

$$X_i \text{ op valor}$$

onde X_i é um atributo, op é um operador relacional e **valor** é constante válida para o atributo X_i

- ❑ Exemplos

- Sexo = Masculino
- Idade $\geq$ 20
- Sexo = Feminino and Idade $<$ 90

Regra

- ❑ Uma regra assume a forma **if L then R** que é equivalente a $L \rightarrow R \equiv R \leftarrow L \equiv R :- L$
- ❑ Normalmente, as partes esquerda L e direita R são complexos sem atributos comuns entre eles, ou seja
 - $\text{atributos}(L) \cap \text{atributos}(R) = \emptyset$
- ❑ A parte esquerda **L** é denominada **condição**, **premissa**, **antecedente**, **cauda** ou **corpo** da regra
- ❑ A parte direita **R** é denominada **conclusão** ou **cabeça** da regra

Regra de Classificação

- ❑ Uma regra de classificação assume a forma restrita de uma regra
 - **if L then classe = C_i**
- ❑ ou simplesmente
 - **if L then C_i**
- ❑ onde C_i pertence ao conjunto de k valores de classe $\{C_1, C_2, \dots, C_k\}$
- ❑ A parte esquerda **L** é um complexo

Regra de Associação

- Uma regra de associação assume que não existe uma definição explícita de classe e qualquer atributo (ou atributos) pode ser usado como parte da conclusão da regra
- Exemplo
 - if $X_3 = 'S'$ and $X_5 > 2$ then $X_1 = 'N'$ and $X_2 < 1$

Cobertura

□ Seja regra $L \rightarrow R$

- Exemplos que satisfazem a parte L da regra são **cobertos** pela regra (ou a regra **dispara** para esses exemplos)
- Exemplos que satisfazem tanto a condição L como a conclusão R são **cobertos corretamente** pela regra
- Exemplos satisfazendo a condição L mas não a conclusão R são **cobertos incorretamente** pela regra
- Exemplos que não satisfazem a condição L **não são cobertos** pela regra

Exemplos satisfazendo...	são...
$\neg L$	Não cobertos pela regra
L	Cobertos pela regra
$L \wedge R$	Cobertos corretamente pela regra
$L \wedge \neg R$	Cobertos incorretamente pela regra

Cobertura: Exemplo

□ if $X_1 = a$ and $X_2 = s$ then classe = +

Atributos					
Exemplo	X_1	X_2	X_3	Classe	Cobertura
E_1	a	s	2	+	Coberto (corretamente)
E_2	a	s	1	-	Coberto (incorretamente)
E_3	b	n	1	+	Não coberto
E_4	b	s	2	-	Não coberto
E_5	c	n	2	*	Não coberto

Estimação da Acuracia

- 2/3 treinamento, 1/3 teste
- Validação cruzada
 - K conjuntos exclusivos e exaustivos
 - O algoritmo é executado k vezes
 - Estratificação
 - Mesmo conjunto de classes em cada conjunto

Bias Indutivo

- Qualquer critério, implícito ou explícito, utilizado para decidir entre uma hipótese e outra, sem ser a consistência com os dados.
 - Bias de representação,
 - Bias de preferência.

Linguagens de Descrição

❑ Lógica de ordem zero ou Proposicional

- O objeto é representado e descrito em termos de conjunções, disjunções e negações de constantes booleanas que representam um campo
- Ex:

$$\text{fêmea} \wedge \text{adulta} \rightarrow \text{pode_ter_filhos}$$

❑ Lógica de atributos

- Notação equivalente à LP, mas os atributos são tratados como variáveis
- Ex:

$$\text{sexo=fêmea} \wedge \text{idade=adulta} \rightarrow \text{classe=pode_ter_filhos}$$

ou

$$\text{sexo(fêmea)} \wedge \text{idade(adulta)} \rightarrow \text{classe(pode_ter_filhos)}$$

Linguagens de Descrição

❑ Lógica de 1ª ordem ou Relacional

- Pode representar objetos como predicados que especificam propriedades ou relações
- Cláusulas de Horn são um exemplo
- Ex:

$$\begin{aligned} & \text{macho}(X) \wedge \text{progenitor}(Z,X) \wedge \text{progenitor}(Z,Y) \rightarrow \text{irmão}(X,Y) \\ & \text{ou} \\ & \text{irmão}(X,Y) \leftarrow \text{macho}(X) \wedge \text{progenitor}(Z,X) \wedge \text{progenitor}(Z,Y) \\ & \text{ou} \\ & \text{irmão}(X,Y) \text{ :- macho}(X), \text{progenitor}(Z,X), \text{progenitor}(Z,Y) \end{aligned}$$

❑ Lógica de 2ª ordem

- Extensão da lógica de primeira ordem, em que os predicados podem ser considerados como variáveis
- Ex:

$$P_1(X,Y) \text{ :- } P_2(X), P_3(Z,X), P_4(Z,Y)$$

pode ser instanciado com:

$$\text{irmão}(X,Y) \text{ :- macho}(X), \text{progenitor}(Z,X), \text{progenitor}(Z,Y)$$

❑ Funções Matemáticas

Bias de Preferência

- Como o algoritmo prefere uma hipótese frente a outra.
- Qualidade da regra
- A estratégia utilizada para gerar novas regras a partir da atual.

Occam's Razor

- Entidades não devem ser multiplicadas sem necessidade
- Entre todas as hipóteses consistentes com a evidencia, a mais simples é a mais provável de ser verdadeira.

O princípio de mínimo comprimento de descrição (MDL)

- Heurística
 - Comprimento da hipótese
 - Comprimento dos dados, o comprimento dos dados quando codificado usando a hipótese como preditor
 - O comprimento do termo de codificação das instancias que são exeções