

A top-down view of a dark desk. On the left, a white smartphone is partially visible. Next to it is a black pencil. A white notebook is open, with a pair of black-rimmed glasses resting on its pages. Below the notebook is a white computer keyboard. In the bottom left corner, a white ceramic cup with a dark rim is partially visible. A large, white, torn-paper-like shape separates the desk image from the text on the right.

Introdução ao Aprendizado de Máquina

Huei Diana Lee

Inteligência Artificial
CECE/UNIOESTE-FOZ

Inteligência X Aprendizado

Aprendizado é a essência da Inteligência

Para que uma máquina tenha **Comportamento Inteligente**, deve-se aumentar a **Capacidade de Aprendizado**

Aprendizado

```
graph TD; A[Aprendizado] --> B[Refinamento de Habilidades]; A --> C([Aquisição de Conhecimentos]); C --> D[Sistemas de Inteligência Artificial];
```

Refinamento de
Habilidades

Aquisição de
Conhecimentos

Sistemas de Inteligência Artificial

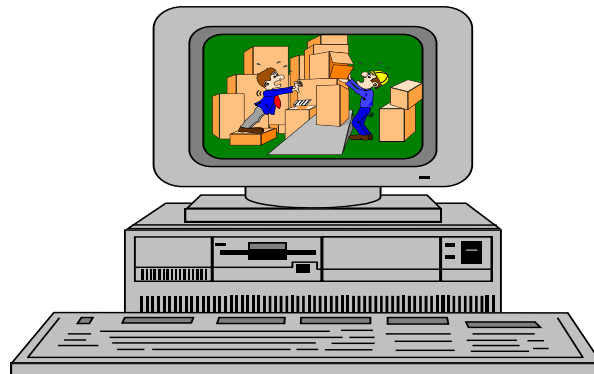
Inteligência x Aprendizado

- Ser humano **pré-programado** para o aprendizado: aprende ampliando o alcance do conhecimento que já possui, por meio de reordenações sucessivas
- Computador não possui o **programa inicial** para procurar por informações e realizar aprendizado em geral
- **Paradigmas e técnicas de Aprendizado de Máquina** possuem um alvo bem mais limitado do que o aprendizado humano

Aprendizado de Máquina (AM)

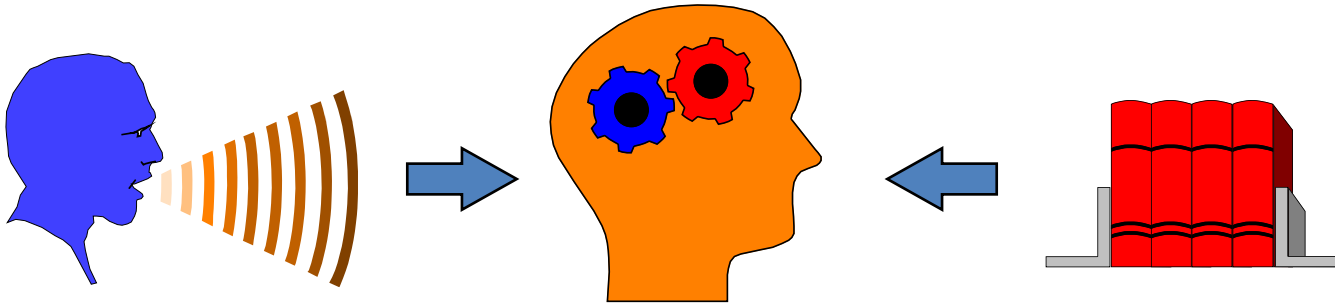
Subárea de Inteligência Artificial (IA) que pesquisa métodos computacionais relacionados à aquisição de:

- novos conhecimentos
- novas habilidades
- novas formas de organizar o conhecimento já existente

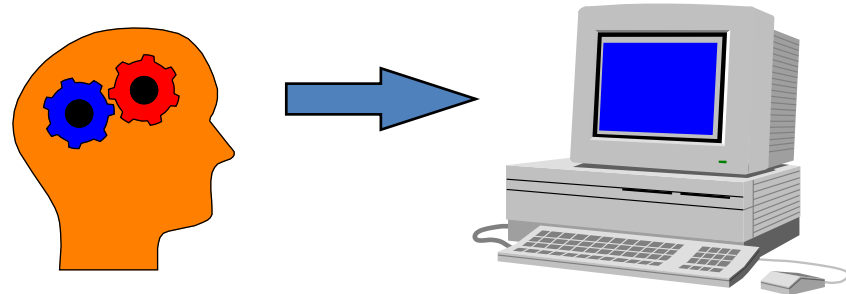


Objetivos de AM

- Um melhor entendimento dos mecanismos de aprendizado humano



- Automação da aquisição do conhecimento



Aprendizado de Máquina: definição

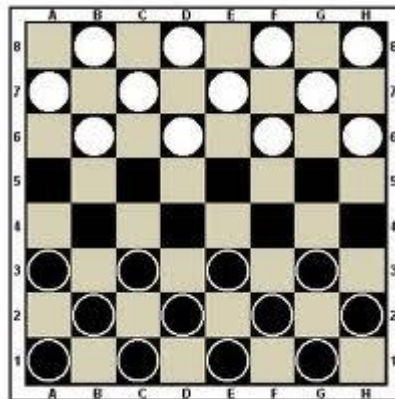
Um programa **aprende** a partir da experiência **E**, em relação a uma classe de tarefas **T**, com medida de desempenho **P**, se seu desempenho em **T**, medido por **P**, melhora com **E**

Mitchell, 1997

Algoritmos de AM: induzem uma função ou hipótese capaz de resolver o problema a partir de exemplos (casos/instâncias) do problema a ser resolvido

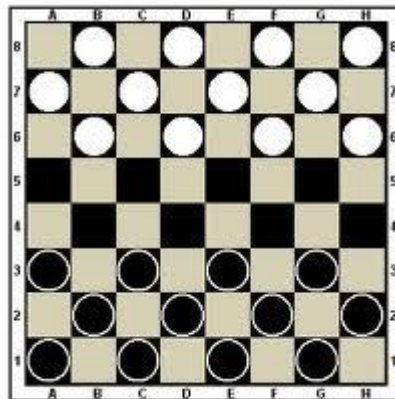
Exemplo 1

- **Problema:** aprender a **jogar** damas
 - Tarefa T: jogar damas
 - Medida de desempenho P: ?
 - Experiência E: ?



Exemplo 1

- **Problema**: aprender a **jogar** damas
 - **Tarefa T**: jogar damas
 - **Medida de desempenho P**: porcentagem de jogos vencidos contra adversários
 - **Experiência E**: praticar jogando



Exemplo 2

- **Problema:** **filtrar** mensagens de email
 - **Tarefa T:** categorizar mensagens de *email* como *spam* ou legítima
 - **Medida de desempenho P:** ?
 - **Experiência E:** ?



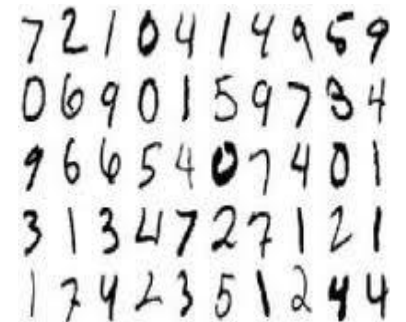
Exemplo 2

- **Problema:** **filtrar** mensagens de email
 - **Tarefa T:** categorizar mensagens de *email* como *spam* ou legítima
 - **Medida de desempenho P:** porcentagem de mensagens de *spam* corretamente identificadas
 - **Experiência E:** conjunto de exemplos de *spams*



Exemplo 3

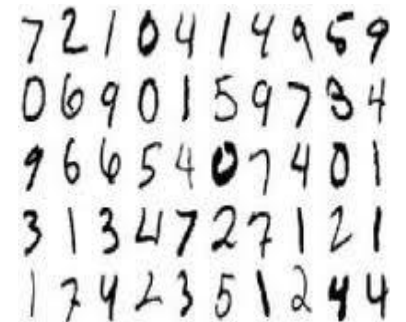
- **Problema:** reconhecer escrita manual
 - **Tarefa T:** reconhecer e classificar dígitos manuscritos dentro de imagens
 - **Medida de desempenho P:** ?
 - **Experiência E:** ?



7	2	1	0	4	1	4	9	5	9
0	6	9	0	1	5	9	7	8	4
9	6	6	5	4	0	7	4	0	1
3	1	3	4	7	2	7	1	2	1
1	7	4	2	3	5	1	2	4	4

Exemplo 3

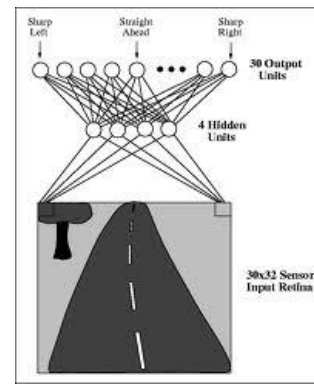
- **Problema:** reconhecer escrita manual
 - **Tarefa T:** reconhecer e classificar dígitos manuscritos dentro de imagens
 - **Medida de desempenho P:** porcentagem de dígitos corretamente identificados
 - **Experiência E:** exemplos de dígitos manuscritos com as respectivas classificações



7	2	1	0	4	1	4	9	5	9
0	6	9	0	1	5	9	7	8	4
9	6	6	5	4	0	7	4	0	1
3	1	3	4	7	2	7	1	2	1
1	7	4	2	3	5	1	2	4	4

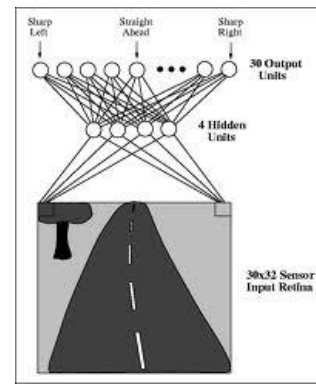
Exemplo 4

- **Problema:** carro autônomo (aprender a **dirigir**)
 - **Tarefa T:** dirigir em uma rodovia pública usando sensores de visão
 - **Medida de desempenho P:** ?
 - **Experiência E:** ?



Exemplo 4

- **Problema:** carro autônomo (aprender a **dirigir**)
 - **Tarefa T:** dirigir em uma rodovia pública usando sensores de visão
 - **Medida de desempenho P:** distância média percorrida antes de um erro
 - **Experiência E:** sequência de imagens e comandos de direção registrados observando um motorista humano



Exemplo 5

- **Problema:** diagnóstico médico
 - **Tarefa T:** diagnosticar o estado de um paciente dado um conjunto de sintomas
 - **Medida de desempenho P:** ?
 - **Experiência E:** ?



Exemplo 5

- **Problema:** diagnóstico médico
 - **Tarefa T:** diagnosticar o estado de um paciente dado um conjunto de sintomas
 - **Medida de desempenho P:** porcentagem de pacientes corretamente diagnosticados
 - **Experiência E:** prontuários médicos de pacientes com seus diagnósticos



Exemplo 6

- **Problema:** detectar bons clientes
 - **Tarefa T:** classificar potenciais clientes como bons ou maus pagadores
 - **Medida de desempenho P:** ?
 - **Experiência E:** ?



Exemplo 6

- **Problema:** detectar bons clientes
 - **Tarefa T:** classificar potenciais clientes como bons ou maus pagadores
 - **Medida de desempenho P:** porcentagem de clientes classificados
 - **Experiência E:** uma base de dados histórica em que os clientes já conhecidos são previamente classificados como bons ou maus pagadores



AM: multidisciplinar

Probabilidade e
Estatística

Teoria da
Informação

Aprendizado de Máquina

Teoria da
Computação

Neurociência

...

Estratégias de AM

Aprendizado por Hábito

Aprendizado por Instrução

Aprendizado por Dedução

Aprendizado por Analogia

Aprendizado por Indução

Estratégias de AM

Aprendizado por Hábito

Aprendizado por Instrução

- Aprendiz não precisa desempenhar nenhuma inferência sobre a informação
- Conhecimento diretamente assimilado (ex. memorização direta)

Aprendizado por Imagem

Estratégias de AM

Aprendizado por Hábito

Aprendizado por Instrução

- Adquire conceitos de uma fonte (professor ou livro por ex.)
- Não copia diretamente a informação fornecida
- Seleciona fatos mais relevantes e/ou transforma informação fonte em formas mais apropriadas

Estratégias de AM

Aprendizado por Hábito

Aprendizado por Instrução

Aprendizado por Dedução

- DEDUÇÃO: inferência logicamente correta

- Aprendiz adquire um conceito por meio da dedução de um conceito já adquirido; transformação sobre um conceito já possuído (PRESERVA veracidade)

Estratégias de AM

Aprendizado por Hábito

Aprendiz adquire conceito modificando/adaptando a definição de um conceito semelhante já conhecido

Aprendizado por Analogia

Aprendizado por Indução

Estratégias de AM

- Permite a habilidade humana realizar generalizações a partir de alguns fatos ou descobrir padrões em coleções de observações aparentemente caóticas
- Conclusões gerais obtidas a partir de exemplos específicos; raciocínio que parte do esp. p/ geral
- PODE OU NÃO PRESERVAR a veracidade

— **Aprendizado por Indução**

Estratégias de AM

Aprendizado por Hábito

Aprendizado por Instrução

Aprendizado por Dedução

Aprendizado por Analogia

Aprendizado por Indução

**Complexidade de
Aprendizado**

Inferência Indutiva

- A **Inferência Indutiva** é um dos principais meios para a aquisição de novos conhecimentos
- **Indução**: raciocínio para obter conclusões **sobre todos** os membros de uma classe pela análise de **alguns membros** da classe



Raciocínio do particular para o geral

Inferência indutiva: exemplo

Se eu noto que:

- Todos pacientes com déficit de atenção atendidos em 1986 sofriam de ansiedade
- Todos pacientes com déficit de atenção atendidos em 1987 sofriam de ansiedade
- Todos pacientes com déficit de atenção atendidos em 1988 sofriam de ansiedade
- ...
- **⇒ Posso inferir que pacientes que sofrem de déficit de atenção também sofrem de ansiedade**

Isto pode ser ou não verdade,
mas propicia uma boa generalização

Conjunto de dados

Experiência pode ser provida por um conjunto de dados (de treinamento)

Ex. base de dados de um hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2	SP	Doente
3217	Maria	18	F	67	Inexistentes	39,5	4	MG	Doente
4039	Luiz	49	M	92	Espalhadas	38,0	2	RS	Saudável
1920	José	18	M	43	Inexistentes	38,5	8	MG	Doente
4340	Cláudia	21	F	52	Uniformes	37,6	1	PE	Saudável
2301	Ana	22	F	72	Inexistentes	38,0	3	RJ	Doente
1322	Marta	19	F	87	Espalhadas	39,0	6	AM	Doente
3027	Paulo	34	M	67	Uniformes	38,4	2	GO	Saudável

Conjunto de dados

Hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2	SP	Doente
3217	Maria	18	F	67	Inexistentes	39,5	4	MG	Doente
4039	Luiz	49	M	92	Espalhadas	38,0	2	RS	Saudável
1920	José	18	M	43	Inexistentes	38,5	8	MG	Doente
4340	Cláudia	21	F	52	Uniformes	37,6	1	PE	Saudável
2301	Ana	22	F	72	Inexistentes	38,0	3	RJ	Doente
1322	Marta	19	F	87	Espalhadas	39,0	6	AM	Doente
3027	Paulo	34	M	67	Uniformes	38,4	2	GO	Saudável

Meta: induzir hipótese para fazer diagnósticos corretos para novos pacientes

Conjunto de dados

Hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2	SP	Doente
3217	Maria	18	F	67	Inexistentes	39,5	4	MG	Doente
4039	Luiz	49	M	92	Espalhadas	38,0	2	RS	Saudável
1920	José	18	M	43	Inexistentes	38,5	8	MG	Doente
4340	Cláudia	21	F	52	Uniformes	37,6	1	PE	Saudável
2301	Ana	22	F	72	Inexistentes	38,0	3	RJ	Doente
1322	Marta	19	F	87	Espalhadas	39,0	6	AM	Doente
3027	Paulo	34	M	67	Uniformes	38,4	2	GO	Saudável

Cada **linha** (paciente) é um **dado**
(objeto, exemplo, padrão ou registro)

Conjunto de dados

Hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2	SP	Doente
3217	Maria	18	F	67	Inexistentes	39,5	4	MG	Doente
4039	Luiz	49	M	92	Espalhadas	38,0	2	RS	Saudável
1920	José	18	M	43	Inexistentes	38,5	8	MG	Doente
4340	Cláudia	21	F	52	Uniformes	37,6	1	PE	Saudável
2301	Ana	22	F	72	Inexistentes	38,0	3	RJ	Doente
1322	Marta	19	F	87	Espalhadas	39,0	6	AM	Doente
3027	Paulo	34	M	67	Uniformes	38,4	2	GO	Saudável

Cada objeto é uma tupla com valores de **características** (atributos, campos ou variáveis), que descrevem seus principais aspectos

Conjunto de dados

Hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2	SP	Doente
3217	Maria	18	F	67	Inexistentes	39,5	4	MG	Doente
4039	Luiz	49	M	92	Espalhadas	38,0	2	RS	Saudável
1920	José	18	M	43	Inexistentes	38,5	8	MG	Doente
4340	Cláudia	21	F	52	Uniformes	37,6	1	PE	Saudável
2301	Ana	22	F	72	Inexistentes	38,0	3	RJ	Doente
1322	Marta	19	F	87	Espalhadas	39,0	6	AM	Doente
3027	Paulo	34	M	67	Uniformes	38,4	2	GO	Saudável

Atributo de saída (**alvo/meta**): presente em algumas tarefas, seus valores devem ser estimados usando outros atributos (de entrada/preditivos)

Importante: atributos de identificação e nome não possuem relação com a doença e não são utilizados como entradas

Conjunto de dados

Hospital

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2	SP	Doente
3217	Maria	18	F	67	Inexistentes	39,5	4	MG	Doente
4039	Luiz	49	M	92	Espalhadas	38,0	2	RS	Saudável
1920	José	18	M	43	Inexistentes	38,5	8	MG	Doente
4340	Cláudia	21	F	52	Uniformes	37,6	1	PE	Saudável
2301	Ana	22	F	72	Inexistentes	38,0	3	RJ	Doente
1322	Marta	19	F	87	Espalhadas	39,0	6	AM	Doente
3027	Paulo	34	M	67	Uniformes	38,4	2	GO	Saudável

Importante: lidar com dados imperfeitos (ruídos, ausentes, entre outros)

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

Porém deve-se ter cuidado com o **número de observações** e a **relevância dos dados**

Argumentos Dedutivos vs Indutivos

Argumento dedutivo:

- Todo homem é mortal.
- João é homem.
- Logo, João é mortal.

- Todo brasileiro é mortal.
- Todo paulista é brasileiro.
- Logo, todo paulista é mortal.

Argumento indutivo:

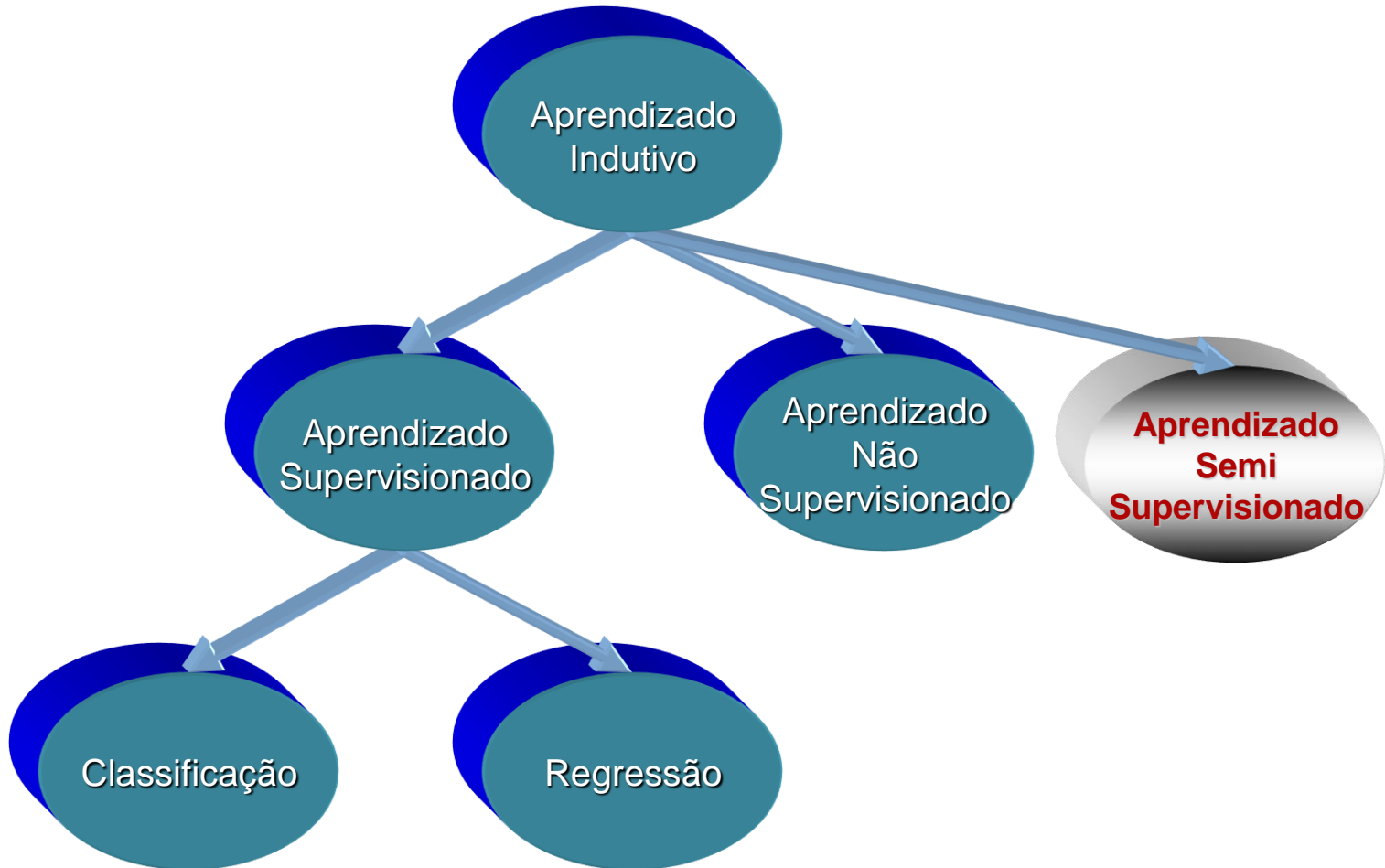
- O ferro conduz eletricidade.
- O ouro conduz eletricidade.
- O chumbo conduz eletricidade.
- A prata conduz eletricidade.
- ...
- Logo, todo metal conduz eletricidade.

- Todo cão é mortal.
- Todo gato é mortal.
- Todo peixe é mortal.
- Todo pássaro é mortal.
- ...
- Logo, todo animal é mortal.

Hierarquia do Aprendizado Indutivo



Hierarquia do Aprendizado Indutivo



Paradigmas de AM

- Simbólico
- Baseado em Casos
- Conexionista
- Genético
- Estatístico

Paradigmas de AM

Simbólico

- Explora representações de **estruturas gráficas** ou **lógicas**, ao invés de métodos estatísticos ou numéricos
- Descrições simbólicas representam um conhecimento de **alto nível**
- As representações simbólicas estão tipicamente na forma de:
 - expressão lógica
 - árvore de decisão
 - regras de produção
 - rede semântica

Paradigmas de AM

Baseado em Casos (Instance-Based)

- Exemplos de treinamento são armazenados
- Casos nunca vistos são classificados por meio de **casos similares** conhecidos
- Classificação de um caso é realizada por meio da “lembrança” de um caso similar cuja classe é conhecida e assume-se que o novo caso terá a mesma classe

Paradigmas de AM

Estatístico

- Regra geral: técnicas estatísticas tendem a focar tarefas em que todos os atributos têm valores contínuos ou ordinais
- Exemplo - um classificador linear:
 - Classes podem ser expressas como combinação linear dos valores dos atributos
 - Procura-se uma combinação linear particular que fornece a melhor aproximação sobre o conjunto de dados

Paradigma Estatístico

Aprendizado Bayesiano

- Provê uma abordagem **probabilística** à inferência
- Baseia-se na suposição de que:
 - Quantidades de interesse são governadas pela distribuição de probabilidades
 - Decisões ótimas podem ser tomadas por meio de raciocínio sobre estas probabilidades juntamente com os dados observados

Teorema de Bayes

$$P(B|A) = P(A|B)P(B)/P(A)$$

Paradigmas de AM

Conexionista

- Redes Neurais Artificiais inspiradas, em parte, na observação de que sistemas de aprendizado biológico são compostos por redes muito complexas de neurônios interconectados
- Redes Neurais Artificiais são redes construídas a partir de conjuntos de **unidades simples altamente interconectadas** (conexionismo)

Paradigmas de AM

Genético

- Classificador genético: uma população de elementos de classificação que competem para fazer a predição
- Elementos com performance “fraca” são descartados; elementos mais “fortes” proliferam, produzindo variações de si mesmos
- Analogia direta com a teoria de Darwin - sobrevivem os mais bem adaptados ao ambiente

Características do Aprendizado Indutivo

- Aprendizado Incremental
- Aprendizado Não Incremental
- Linguagens de Descrição

Características do Aprendizado Indutivo

- Aprendizado Incremental

Em geral, no aprendizado incremental, o processo de aprendizado procede através de uma sequência de hipóteses, H_1 , H_2 , ... etc., sobre o conceito que está sendo aprendido

Quando um exemplo é processado, a hipótese corrente é atualizada, se necessário, resultando na próxima hipótese

Características do Aprendizado Indutivo

- Aprendizado Incremental
- 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, no qual todos os exemplos estão disponíveis e, provavelmente, não irão ocorrer mudanças

Características do Aprendizado Indutivo

Linguagens de Descrição (LD)

1. LD de exemplos $L_{\mathcal{E}}$ (Objetos)
2. LD de conceitos L_H (Hipóteses)
3. LD da teoria do domínio L_K (Conhecimento de fundo)

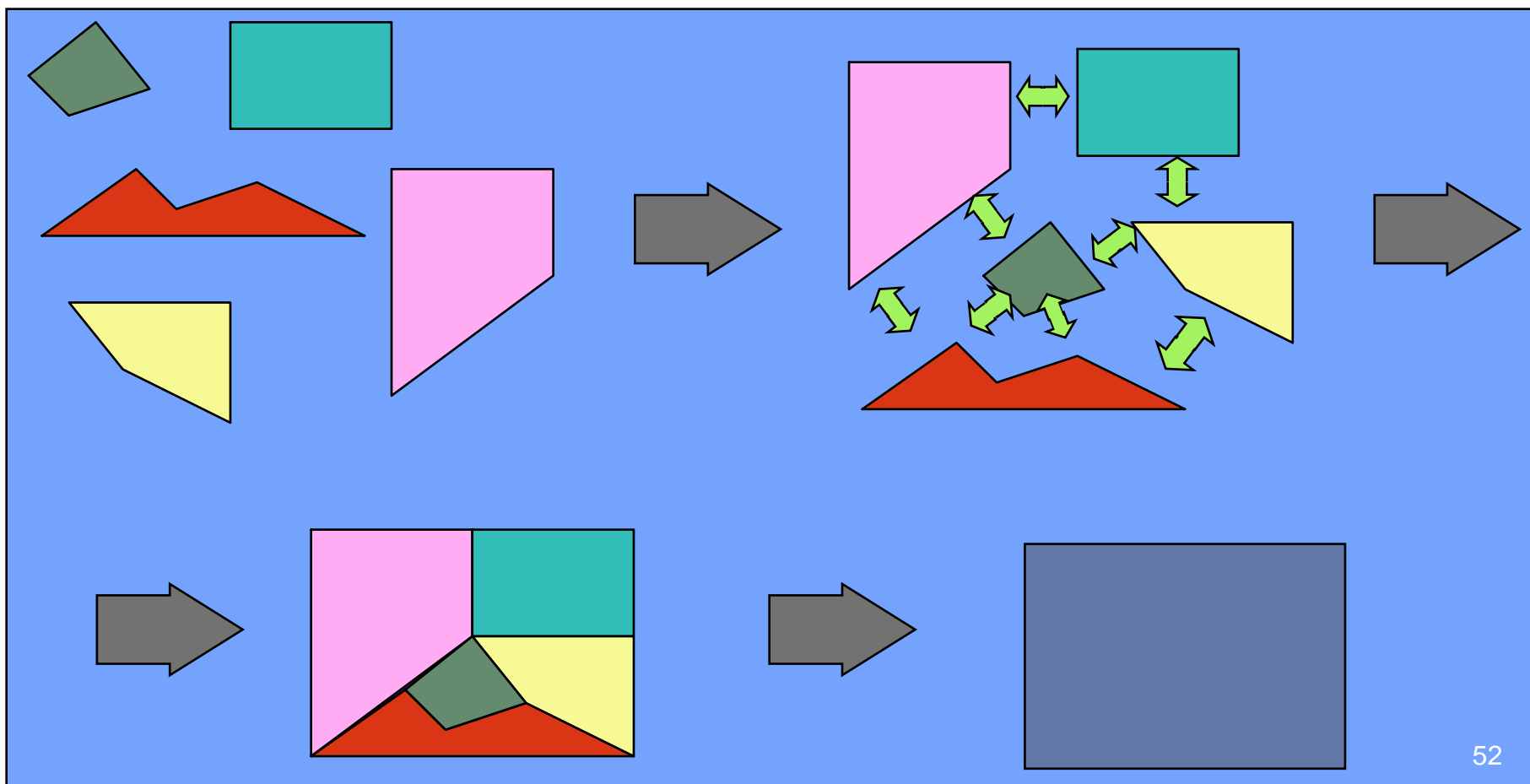
Características do Aprendizado Indutivo

LD de exemplos $L_{\mathcal{E}}(\text{Objetos})$

- 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:
 - Descrições estruturais
 - Descrições de atributos

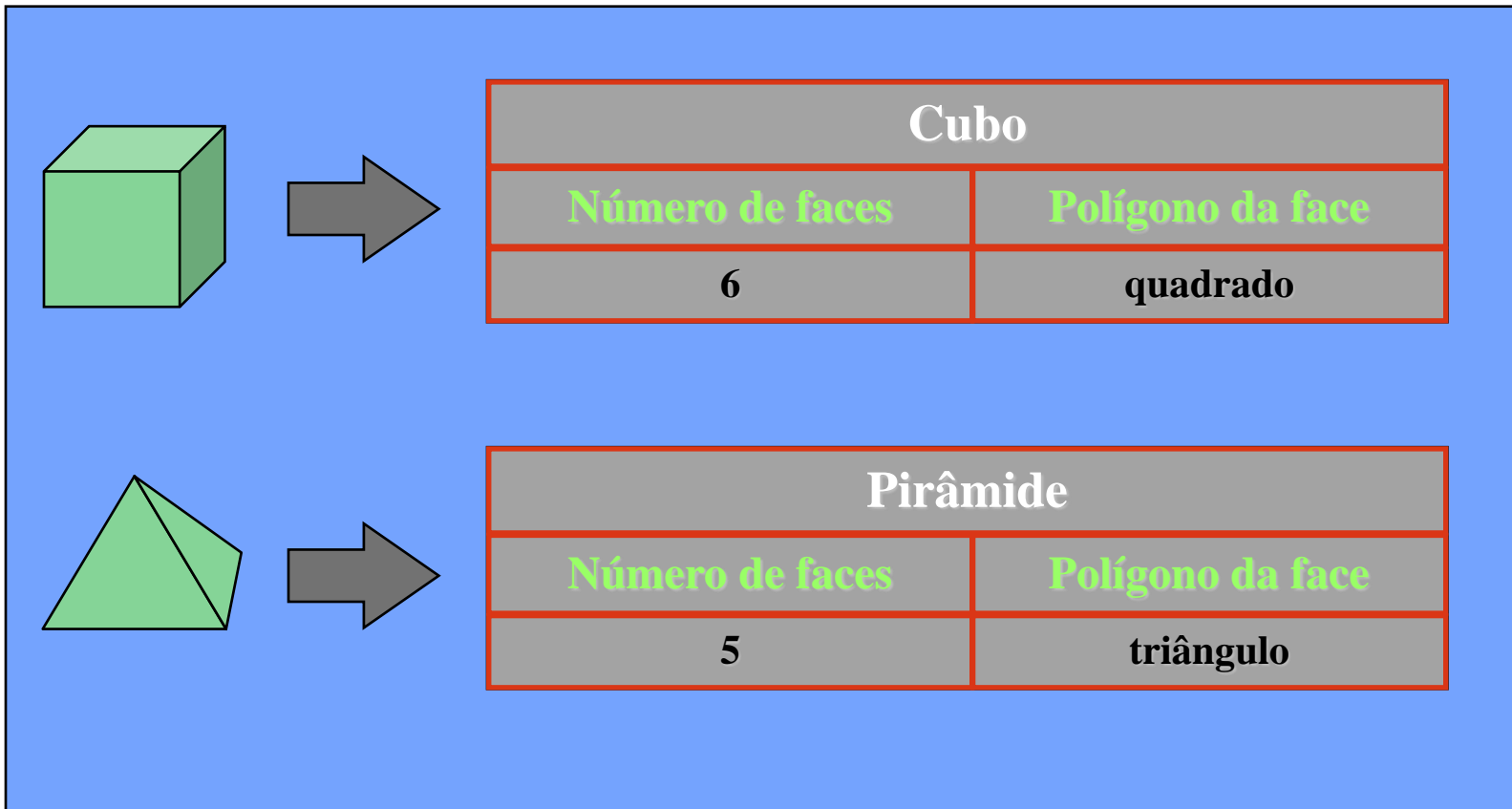
Descrições Estruturais

Um objeto é descrito em termos de seus **componentes** e a **relação** entre eles.



Descrições de Atributos

Um objeto é descrito em termos de suas **características** globais como um vetor de **valores** de atributos.



Características do Aprendizado Indutivo

LD de conceitos L_H (Hipóteses)

- 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*

Características do Aprendizado Indutivo

LD de conceitos L_H (Hipóteses)

- Árvores de decisão para representar conceitos



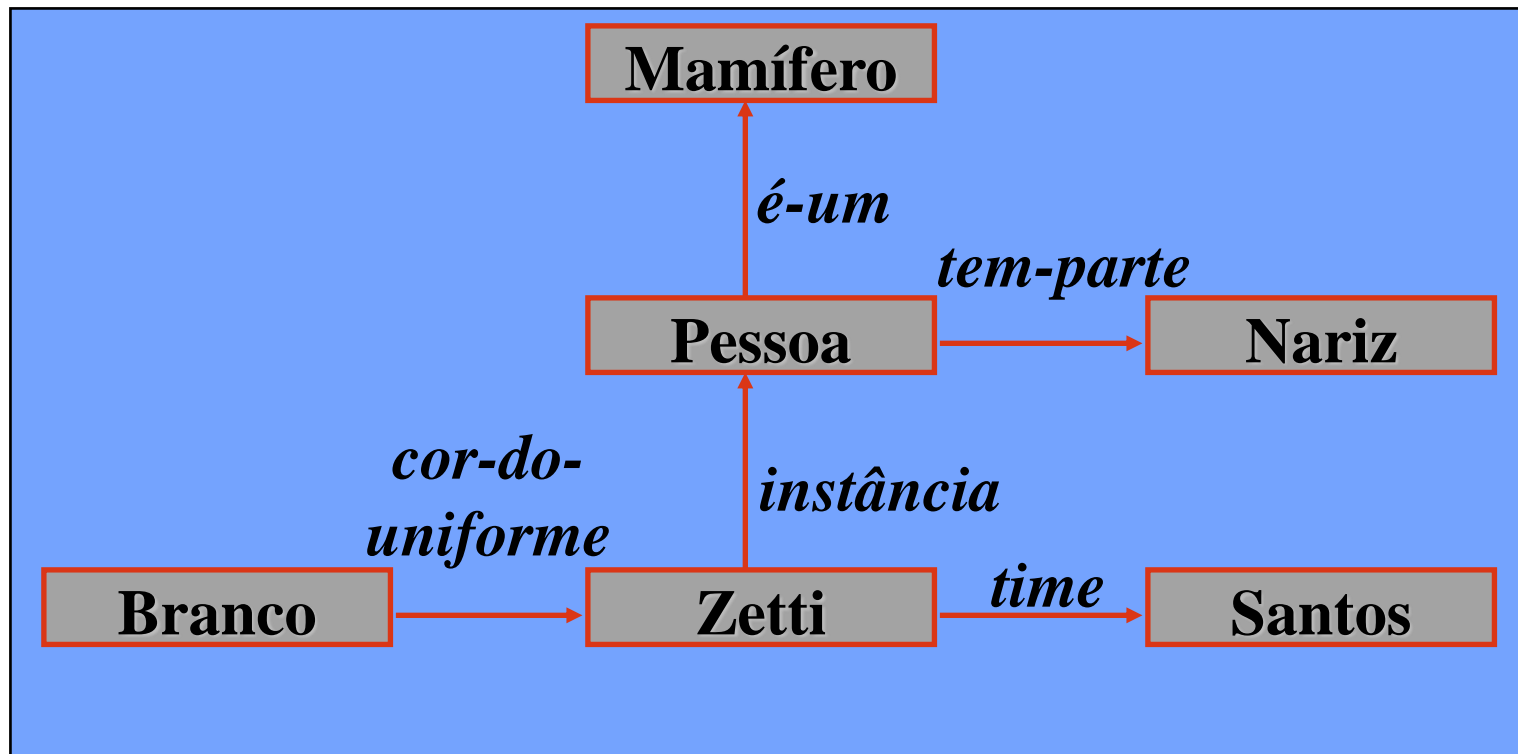
- Lógica de predicados

```
filha(X, Y) <-- mulher(X), pais(X, Y).
```

Características do Aprendizado Indutivo

LD de conceitos L_H (Hipóteses)

- Redes semânticas



Características do

Aprendizado Indutivo

LD da teoria do domínio L_K
(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

Aprendizado Indutivo de Conceitos

- Dados

$\varepsilon = \varepsilon^+ \cup \varepsilon^-$ conjunto de exemplos de treinamento de um conceito C ,

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

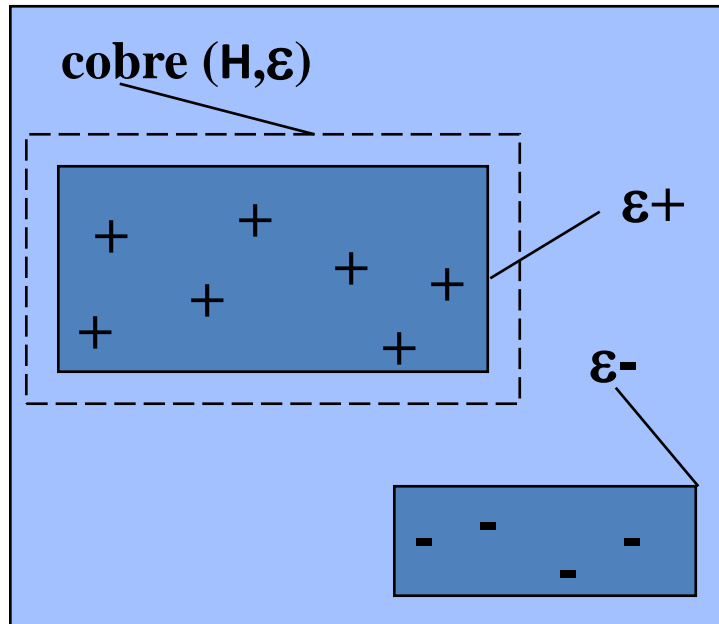
- cada exemplo $e \in \varepsilon^+$ é coberto por H
- nenhum exemplo negativo $e \in \varepsilon^-$ é coberto por H

Aprendizado Indutivo de Conceitos

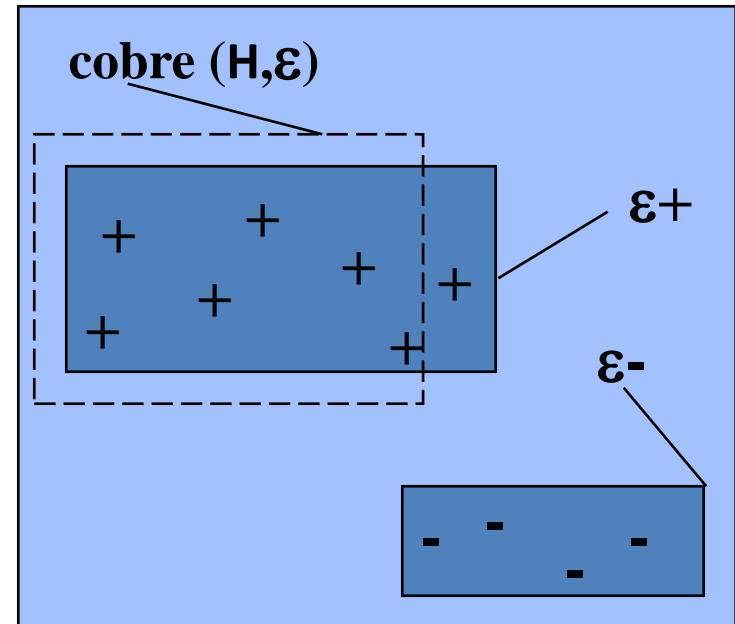
- $cobre(H, \varepsilon) = \{e \in \varepsilon^+ \mid cobre(H, e) = true\}$
(instância positiva)
- $cobre(H, \varepsilon) = \{e \in \varepsilon^- \mid cobre(H, e) = false\}$
(instância negativa)

Completeza e Consistência de H

H: completa, consistente

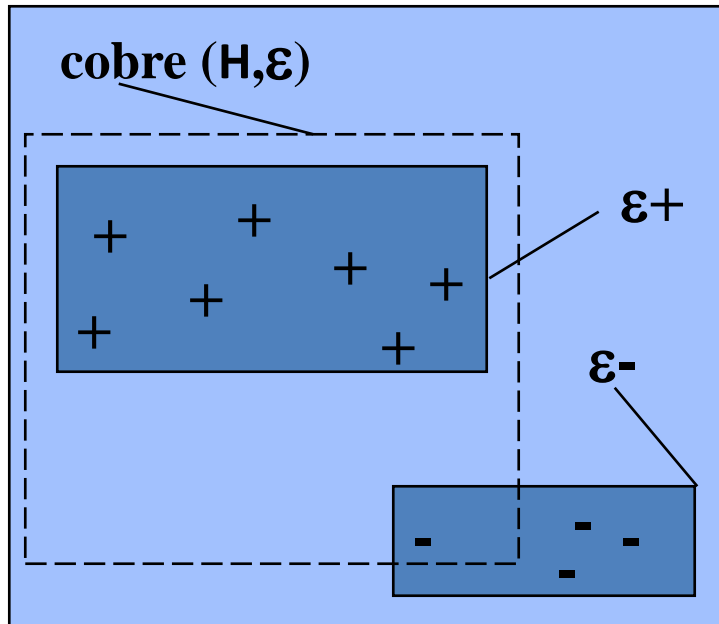


H: incompleta, consistente

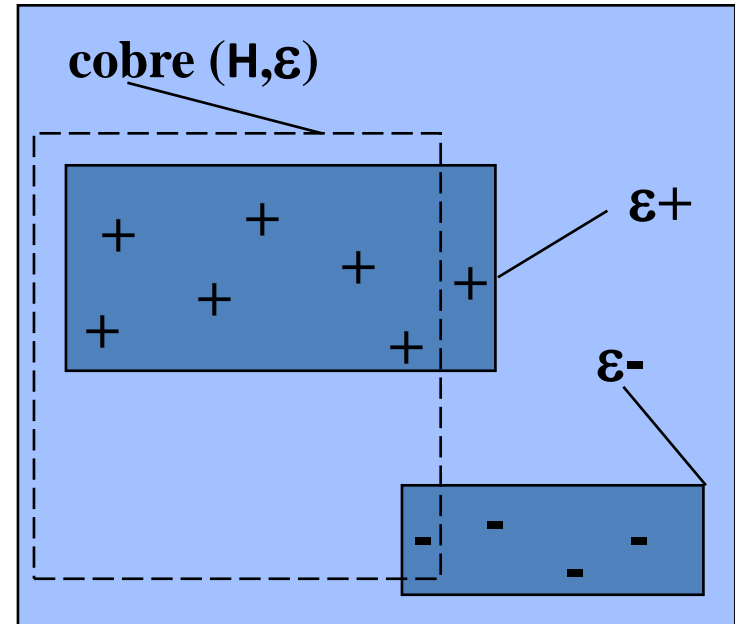


Completeza e Consistência de H (Cont)

H: completa, inconsistente



H: incompleta, inconsistente



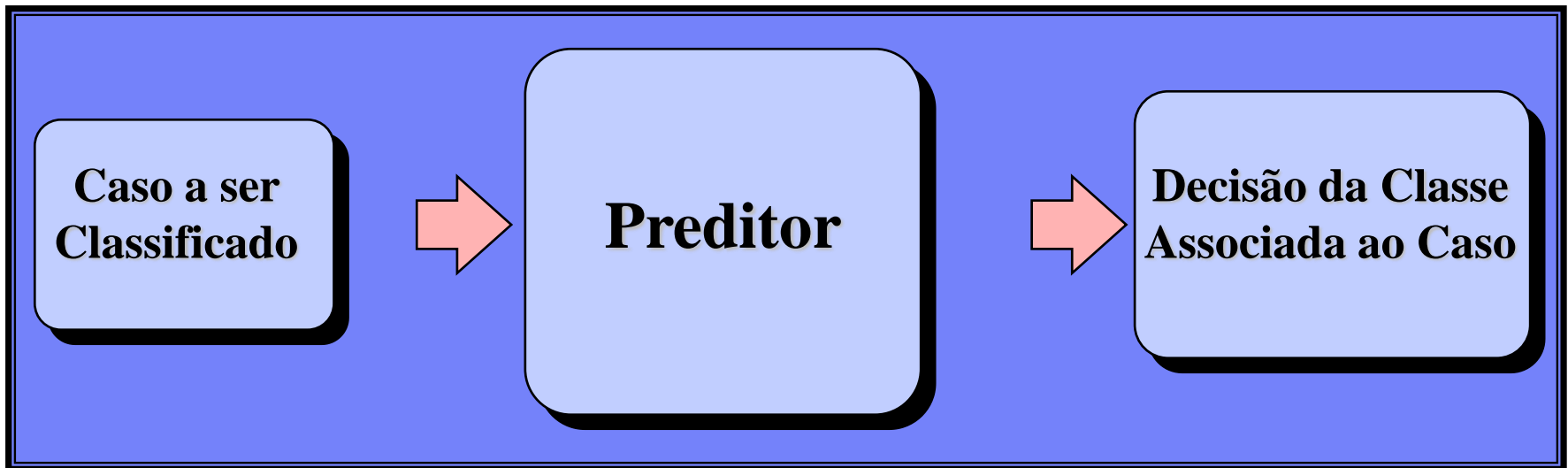
Sistemas de Aprendizizado de Máquina Indutivo

O paradigma de aprendizado indutivo busca aprender conceitos por meio de instâncias destes conceitos

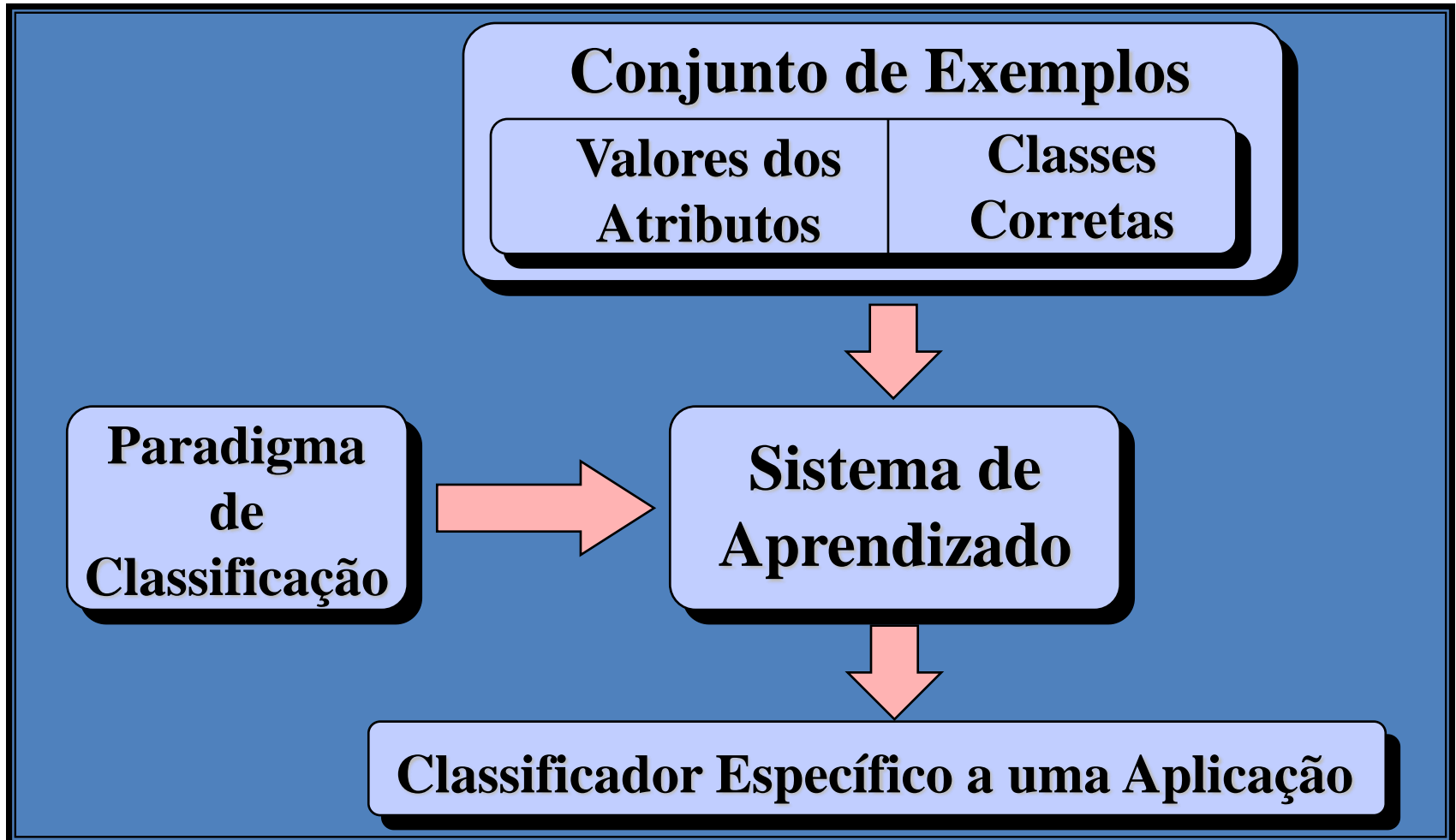


Sistemas de Aprendizado de Máquina Indutivo (Cont)

O classificador utiliza os conceitos aprendidos para classificar novos exemplos

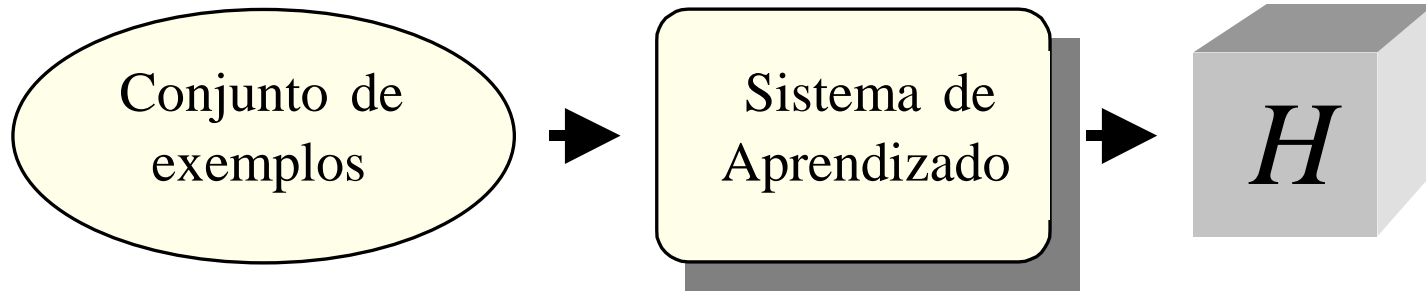


Representação da Classificação

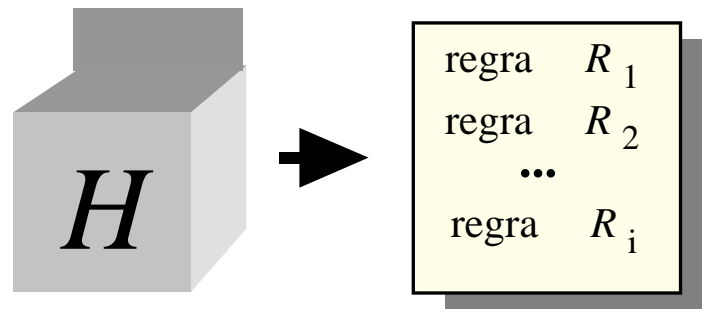


Conhecimento Adquirido (Hipótese H)

- H vista como classificador



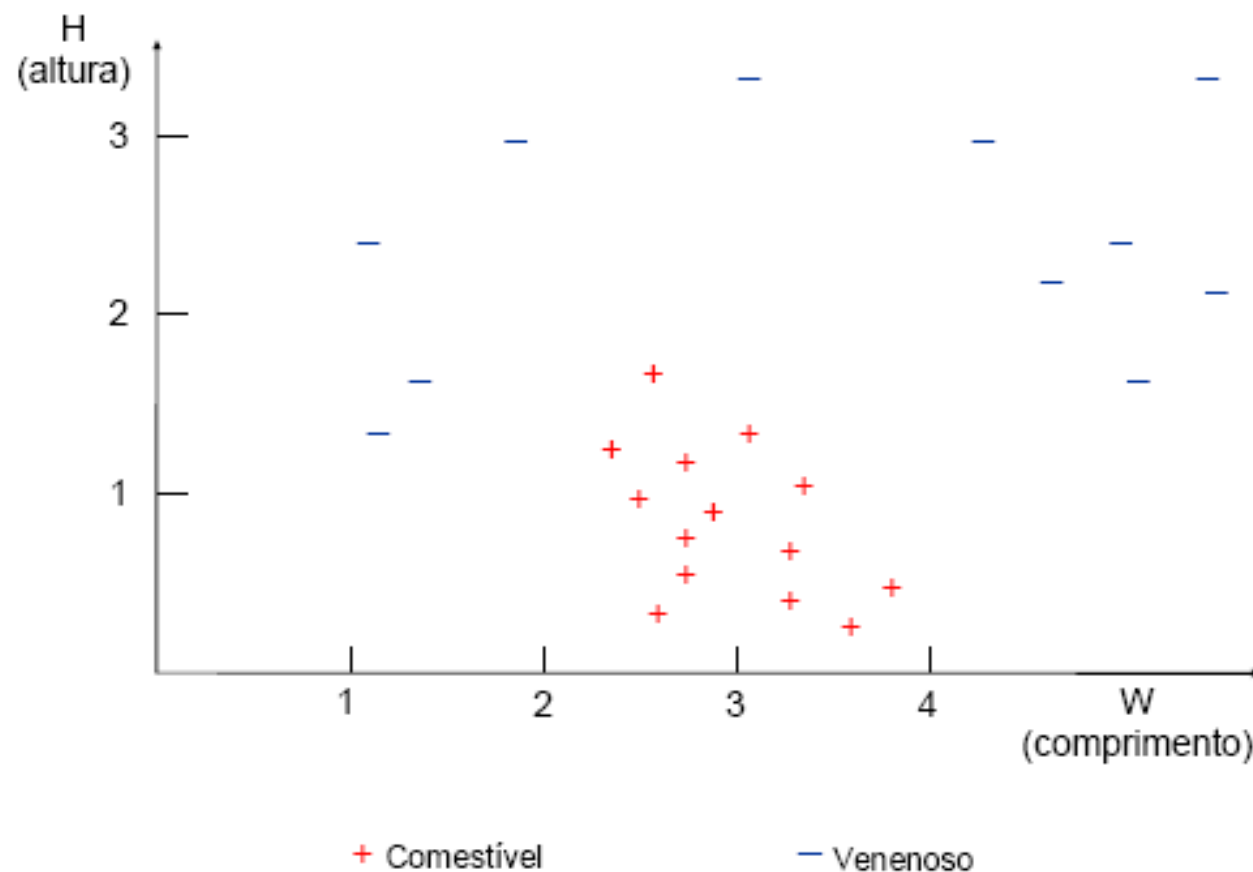
- H vista como conjunto de regras



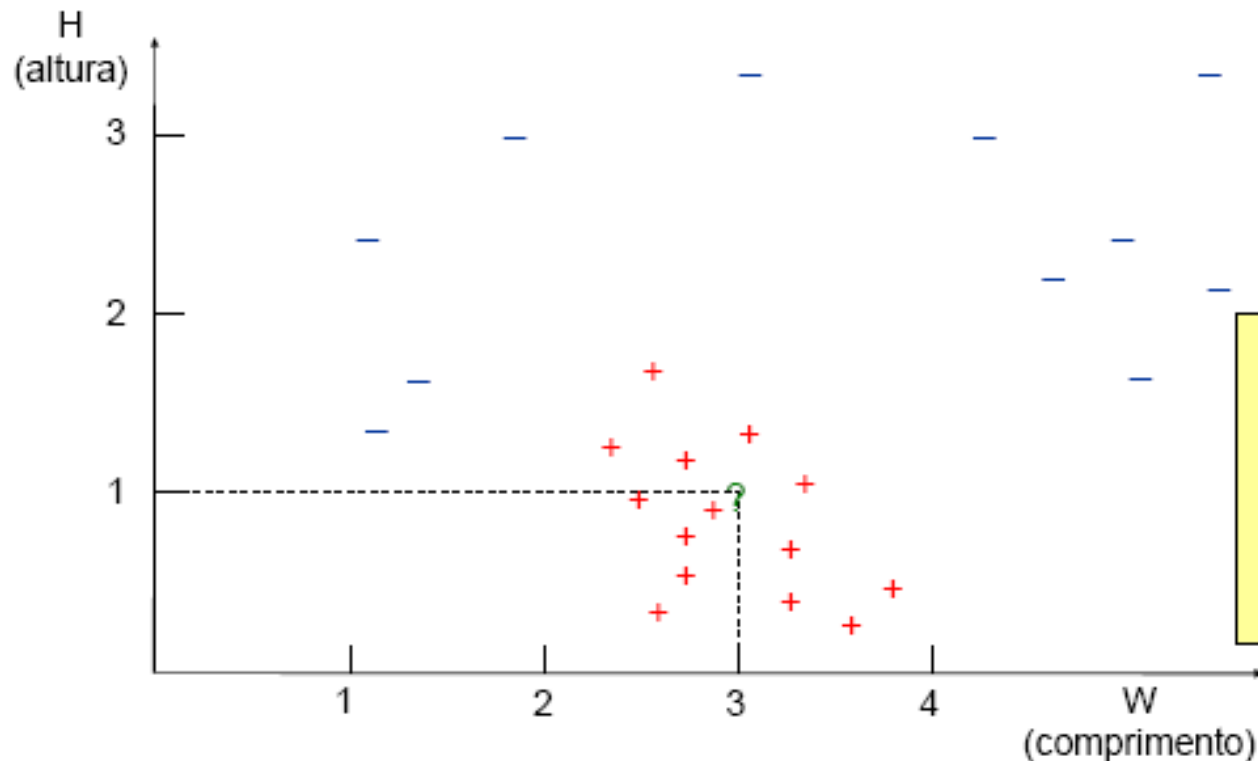
Sistemas de Aprendizado de Máquina

<i>Modo de Aprendizado</i>	<i>Paradigmas de Aprendizado</i>	<i>Linguagens de Descrição</i>	<i>Formas de Aprendizado</i>
<ul style="list-style-type: none">- Supervisionado- Não Supervisionado- Semi Supervisionado	<ul style="list-style-type: none">- Simbólico- Estatístico- Instance-Based- Conexionista- Genético	<ul style="list-style-type: none">- Instâncias ou Exemplos- Conceitos Aprendidos ou Hipóteses- Teoria de Domínio ou Conhecimento de Fundo	<ul style="list-style-type: none">- Incremental- Não Incremental

Exemplo: Cogumelos Comestíveis x Venenosos



Exemplo: Cogumelos Comestíveis x Venenosos

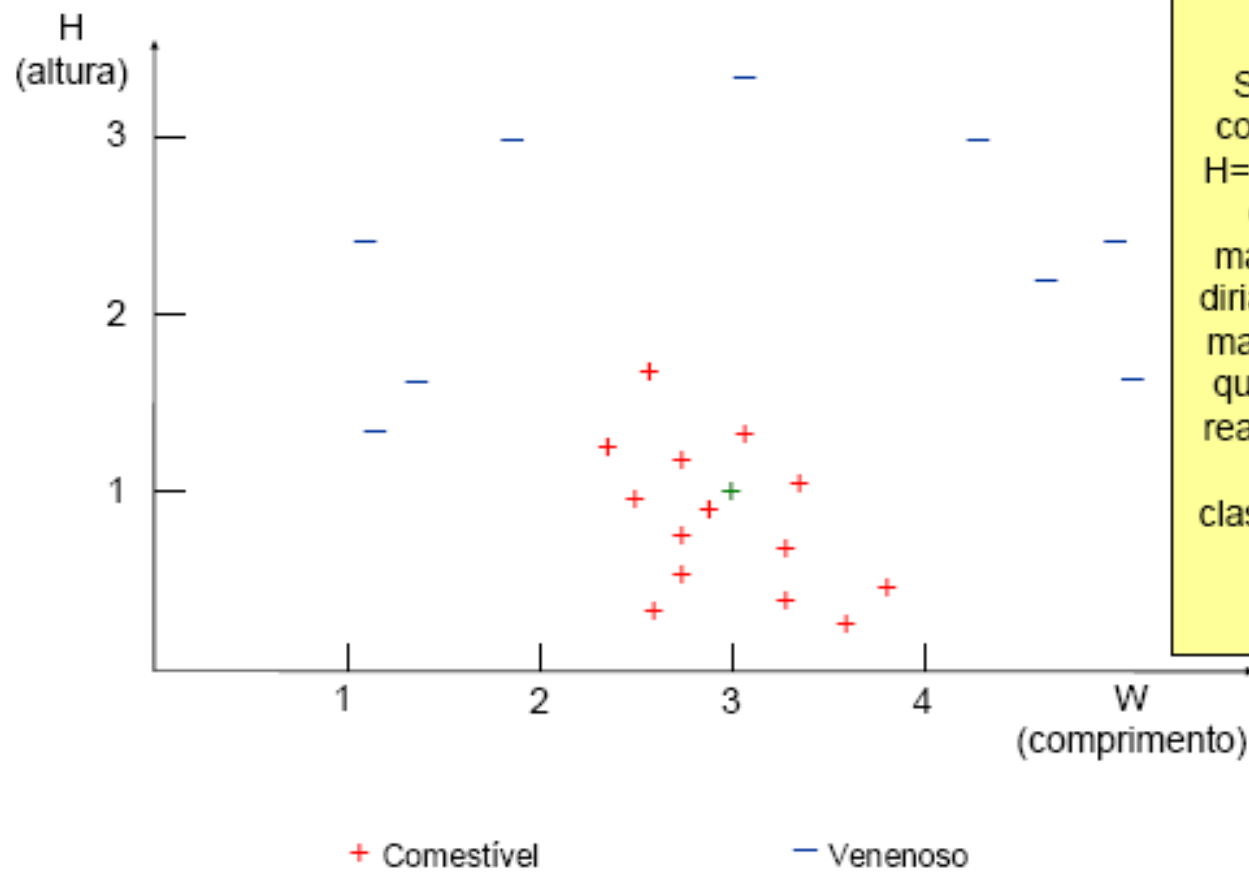


Suponha um novo cogumelo com $W=3$, $H=1$. Ele é comestível ou venenoso?

+ Comestível

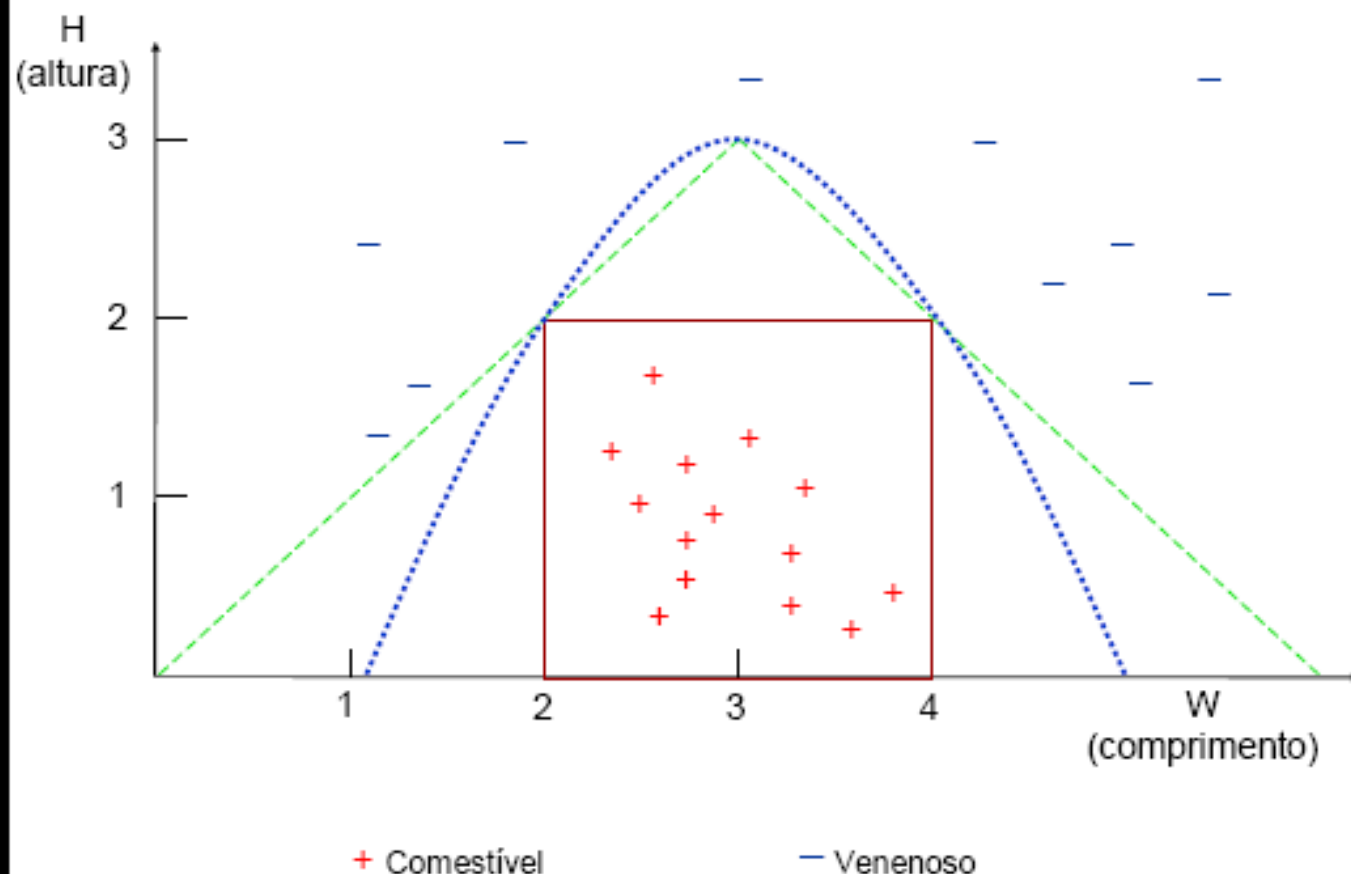
- Venenoso

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

Colaboraram com a preparação deste material, pesquisadores, alunos de pós-graduação e graduação do Labic/ICMC-USP São Carlos e Labi/Unioeste:

- ✍ Profª. Maria Carolina Monard
 - ✍ Profª. Huei Diana Lee
 - ✍ Prof. José Augusto Baranauskas
 - ✍ Prof. Gustavo E.A.P.A. Batista
 - ✍ Chandler Caulkins
 - ✍ Prof. Cláudio Alex Rocha
 - ✍ Marcos Roberto Geromini
 - ✍ Profª. Thereza Patrícia Padilha
 - ✍ Prof. Paulo Horst
 - ✍ Vangrei Fonseca Simão
-
- ✍ Livro Inteligência Artificial: uma abordagem de Aprendizado de Máquina, 2011.
Autores Katti Faceli, Ana Carolina Lorena, João Gama, André C.P.L.F. de Carvalho
Editora LTC

Revisado por Huei Diana Lee