

Aprendizado de Máquina

Aprendizado Supervisionado e Não Supervisionado

Visão Geral e Conceitos Básicos

Inteligência Artificial – 2020/1

Aprendizado de Máquina

- Subárea da IA relacionada à construção de sistemas capazes de adquirir conhecimento de forma automática (sem a interferência humana)
- É parte da aquisição de conhecimento:

“Aquisição de Conhecimento é a transferência e transformação do conhecimento especializado com potencial para a resolução de problemas de alguma fonte de conhecimento para um programa.”

Aprendizado de Máquina

- Classificação de técnicas de Aquisição de conhecimento:
 - Manuais - processo conduzido por engenheiros do conhecimento
 - Semi-automáticas – especialistas utilizam ferramentas que minimizam a participação do engenheiro do conhecimento
 - Automáticas – objetivam minimizar a participação humana, extraíndo conhecimento de fontes de dados extensas

Aprendizado de Máquina

- Paradigmas de aprendizado
 - **simbólico** - analisam exemplos e contra-exemplos de um conceito e constroem representações simbólicas (expressão lógica, árvore de decisão, regras, redes semânticas)
 - **estatístico** - utilizam modelos estatísticos para encontrar uma boa aproximação do conceito induzido
 - **baseado em exemplos (instance based)** - classificam exemplos nunca vistos por meio de exemplos similares conhecidos. Os exemplos representativos são armazenados. (lazy learning)
 - **conexionista** - relacionado ao treinamento das redes neurais
 - **evolutivo** - baseado na teoria da evolução natural

Aprendizado de Máquina

- Aprendizado de máquina indutivo
 - Independente do paradigma utilizado, a grande maioria das estratégias de aprendizado de máquina realiza o que é chamado de **aprendizado indutivo**
- **Indução**: forma de inferência lógica que permite obter conclusões genéricas sobre um conjunto particular de exemplos.
- Um **conceito** é aprendido efetuando-se inferência indutiva sobre os exemplos apresentados.
- **Hipóteses** geradas podem ou não preservar a verdade.

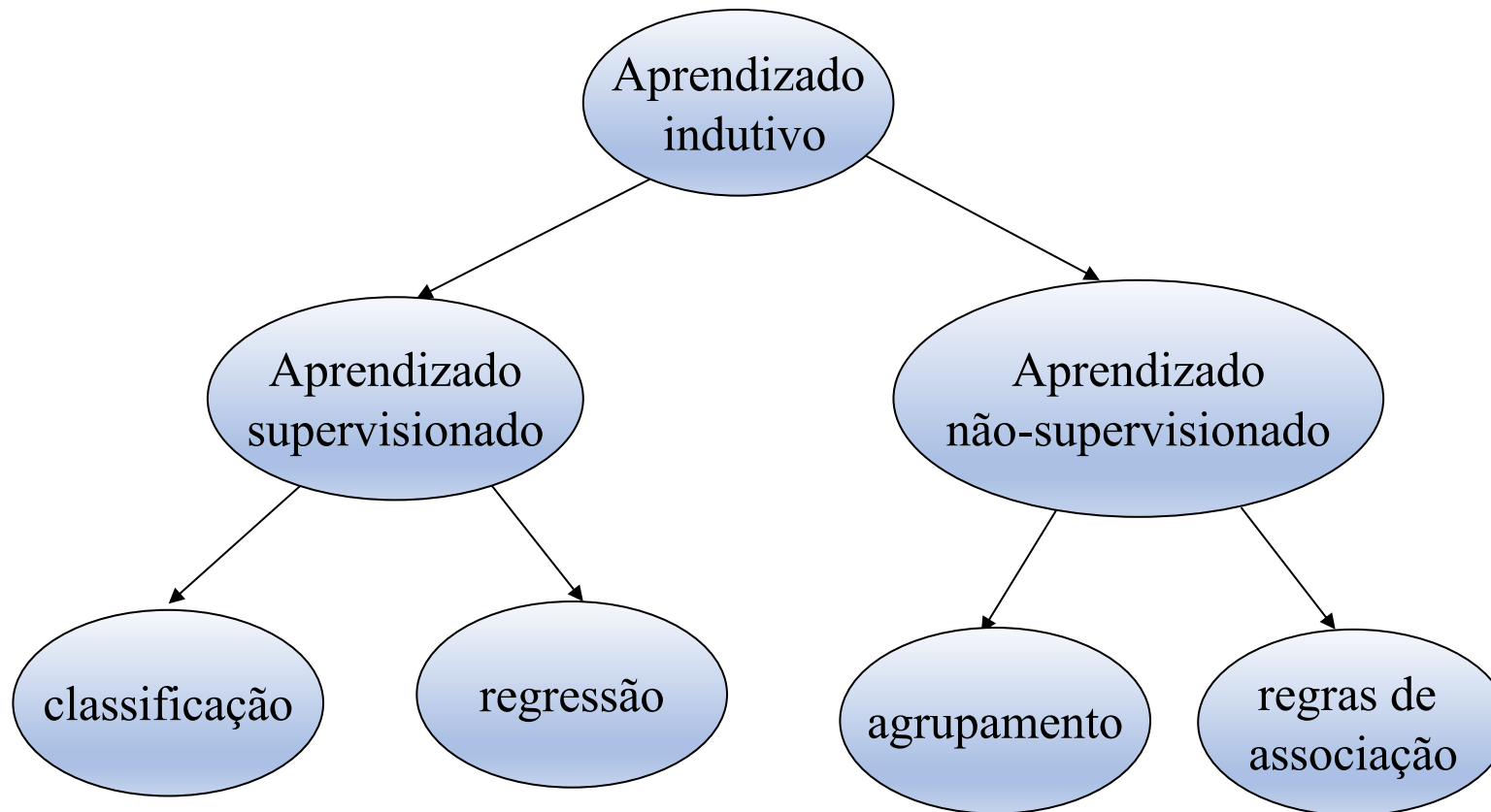
Aprendizado de Máquina

- Aprendizado de máquina indutivo
- Aprende fazendo generalizações sobre casos específicos



Aprendizado de Máquina

- Hierarquia do Aprendizado Indutivo



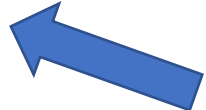
Aprendizado de Máquina

- Aprendizado supervisionado
 - Utiliza informação prévia (supervisão) a respeito da classe a que pertence o exemplo
- Aprendizado não supervisionado
 - Não utiliza a informação prévia a respeito da classe
- Aprendizado semissupervisionado
 - Utiliza apenas uma parte dos exemplos com informação prévia e os demais sem essa informação

Aprendizado de Máquina

- Formato dos Dados para Aprendizado Indutivo – Tabela Atributo-Valor
 - Os dados devem estar em forma de tabela, com um exemplo por linha e um atributo por coluna
 - **Exemplo** (instância, caso, registro, dado) - tupla de valores de atributos que descreve o objeto de interesse (um paciente, histórico de clientes de um banco)
 - **Classe** (rótulo) : atributo especial que descreve o conceito-meta que se deseja aprender

Exemplos	X1	X2	Xm	Y
E1	x_{11}	x_{12}	x_{1m}	y_1
E2	x_{21}	x_{22}	x_{2m}	y_2
...
En	x_{n1}	x_{n2}	x_{nm}	y_m



Classe: pode ser conhecida ou não

Aprendizado de Máquina Supervisionado

- Os exemplos consistem de um conjunto de atributos e um atributo **meta** (classe, conceito que vai ser aprendido).
- O valor do atributo meta é chamado de **rótulo** do exemplo
- Os exemplos são ditos rotulados.
- Induz um modelo que prevê o valor do atributo meta de dados não conhecidos, usando o conhecimento extraído de dados (rotulados) conhecidos.
- Pode ser usado diretamente na tomada de decisões

Aprendizado de Máquina Supervisionado

- Formato dos dados para aprendizado supervisionado
- Tabela atributo-valor
- Os exemplos são ditos rotulados
- A classe é conhecida

Classe



Exemplos	X1	X2	Xm	Y
E1	x_{11}	x_{12}	x_{1m}	y_1
E2	x_{21}	x_{22}	x_{2m}	y_2
En	x_{n1}	x_{n2}	x_{nm}	y_m

Aprendizado de Máquina Supervisionado

- Tipos de atributos
 - Atributos nominais (categóricos, discretos)
 - Tem um número pequeno de valores, em geral representam categorias
 - Ex: cores, atributos do tipo sim/não, V/F, grau de escolaridade, etc.
 - Atributos contínuos
 - Tem valores numéricos dentro de um intervalo:
 - Ex: temperatura, altura, medidas em geral.

Exemplos	X1	X2	X3	...	Y
E1	0	A	25	...	y ₁
E2	1	A	21,6	...	y ₂
E3	0	B	32,5	...	y ₃
E4	0	B	15	...	y ₄
...

Aprendizado de Máquina Supervisionado

- Tipos de problemas do AM supervisionado
 - **Classificação**: quando o atributo classe é discreto
 - **Regressão**: quando o atributo classe é contínuo.

Classificação

Exemplos	X1	X2	X3	...	Y
E1	0	A	25	...	Baixa
E2	1	A	21,6	...	Alta
E3	0	B	32,5	...	Baixa
E4	0	B	15	...	Média
...

Regressão

Exemplos	X1	X2	X3	...	Y
E1	0	A	25	...	15,1
E2	1	A	21,6	...	28,3
E3	0	B	32,5	...	17
E4	0	B	15	...	31,5
...

Classificação

- Consiste na generalização de exemplos ou na extração de um padrão de classificação, em uma linguagem de representação capaz de reconhecer a classe de um novo exemplo.
- Atributo meta é **discreto** – determina uma categoria
- Extrair um padrão de classificação significa ser capaz de descrever um número grande de casos de uma maneira concisa
- Usualmente a linguagem de representação é baseada em **árvores** ou **regras**

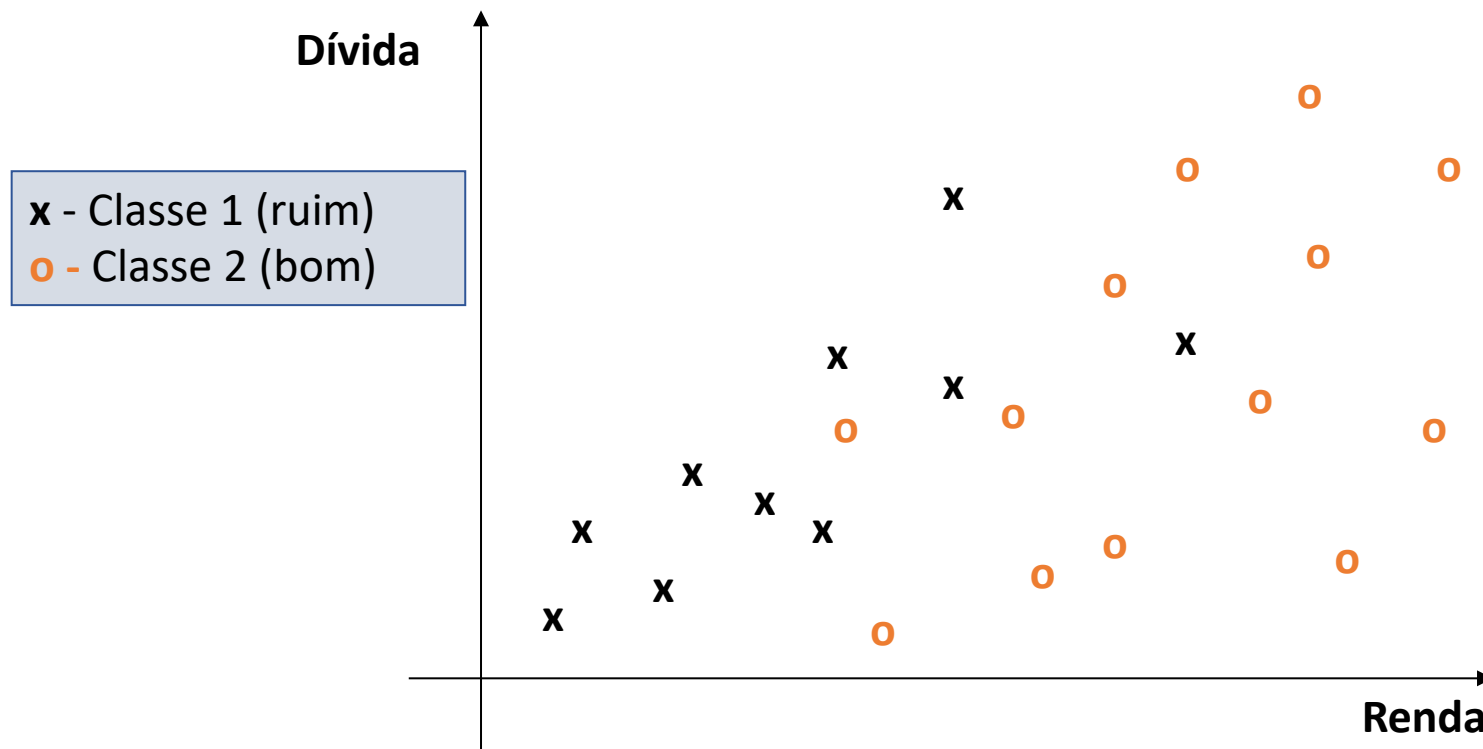
Classificação

- Exemplo
 - Encontrar padrões que permitam distinguir os clients da classe 1 (devedor) dos clients da classe 2 (bom pagador)
 - Uma regra extraída desse conjunto de dados poderia ser:
Se renda acima de 40 e dívida abaixo de 30 então classe 2.

Exemplos	Renda	Dívida	Y
E1	30	40	1
E2	40	20	2
E3	80	60	2
E4	35	30	1

Classificação

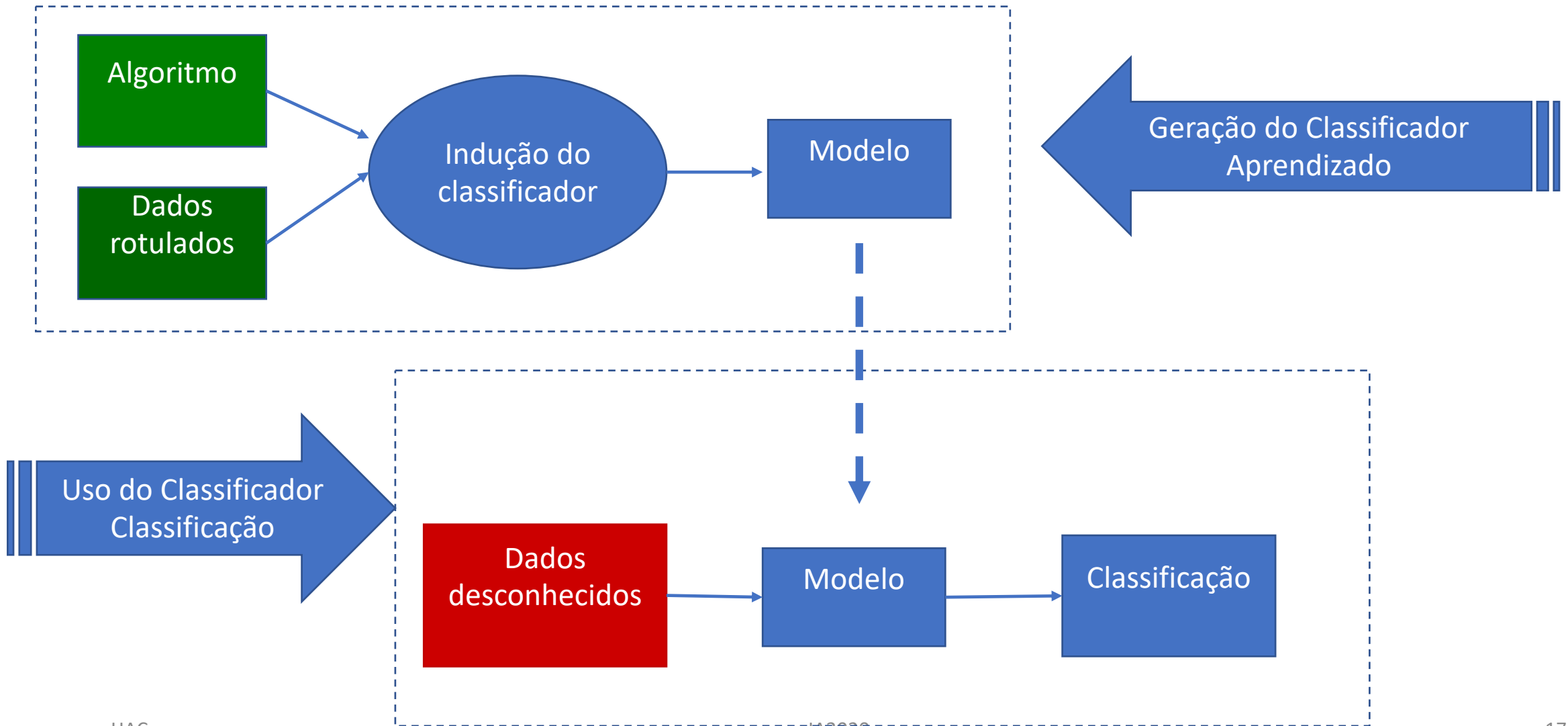
- Representação Gráfica de dados para classificação



Dados no formato atributo-valor:

Exemplos	Renda	Dívida	Y
E1	30	40	1
E2	40	20	2
E3	80	60	2
E4	35	30	1
...

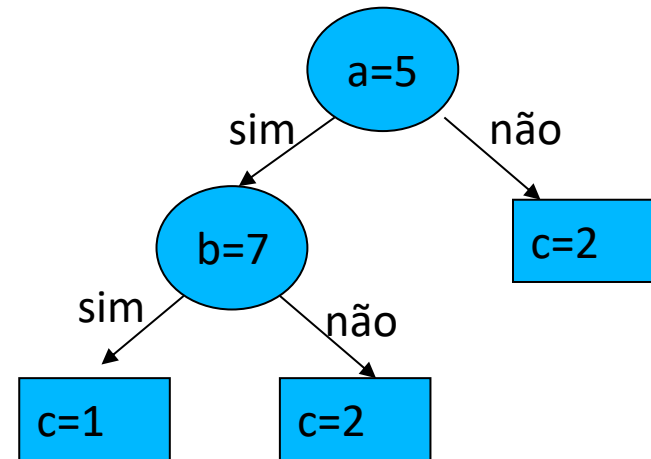
Classificação



Classificação

Em que formato o classificador é representado e como ele é usado para classificação?

- Árvores de decisão
- Regras de decisão



Se $a = 5$ e $b = 7$ então $c = 1$
senão $c = 2$

Regressão

- Consiste em extrair padrões a partir de dados conhecidos rotulados, que permitam estimar o valor do atributo meta de um dado desconhecido.
- Atributo meta é **contínuo**

Encontrar uma relação entre um conjunto de atributos de entrada e um atributo-meta contínuo da seguinte forma:

$$y = f(x_1, x_2, \dots, x_d)$$

Regressão

- Exemplo

- Encontrar padrões que permitam prever o consume de um carro
- Uma regra extraída desse conjunto de dados poderia ser:

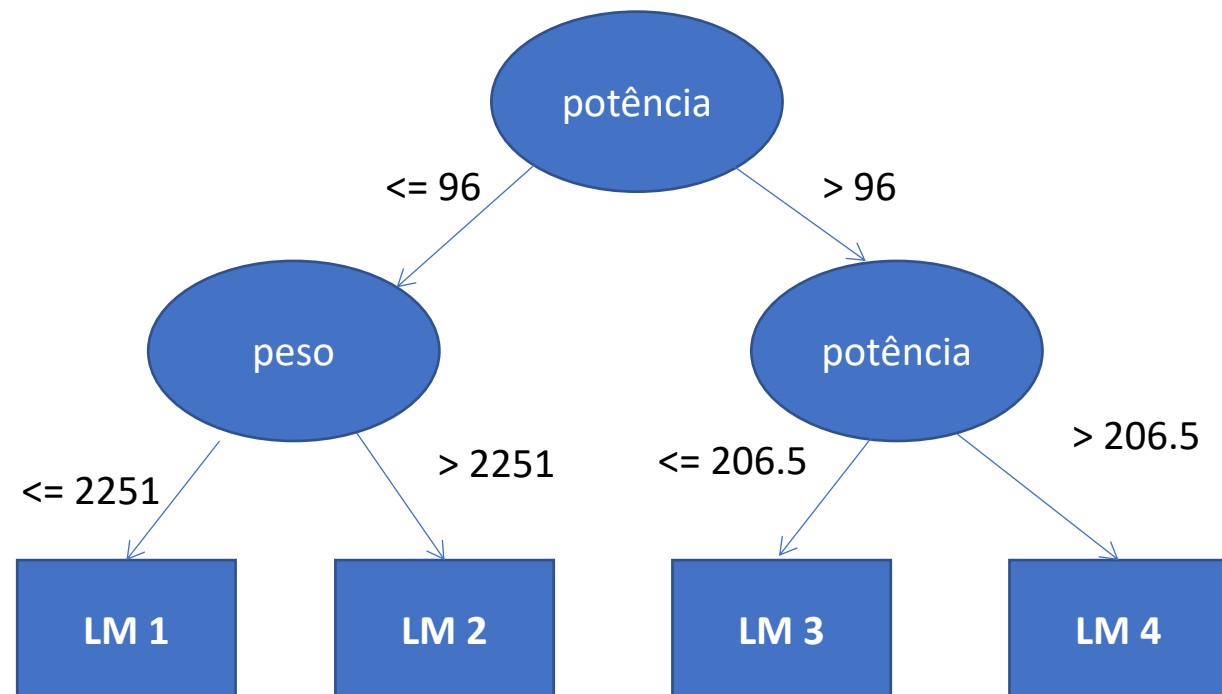
Se potência > 170 e aceleração entre 12 e 13,9 então consumo = 32.

Exemplos	Potencia	Peso	Aceleração	Consumo
E1	180	3852	13,5	33
E2	175	3010	14,4	32
E3	82	2720	19,4	31
E4	168	3120	16,7	32,5

Regressão

Formas de representação do “classificador” que realiza regressão:

- Árvore de regressão
- Regras de regressão



Aprendizado de Máquina não Supervisionado

- Os exemplos consistem de um conjunto de atributos (sem atributo-meta).
- Os exemplos são **não-rotulados**
- Descreve, de forma concisa, estruturas ou relacionamentos encontrados nos dados
- Consiste na identificação de comportamentos intrínsecos do conjunto de dados

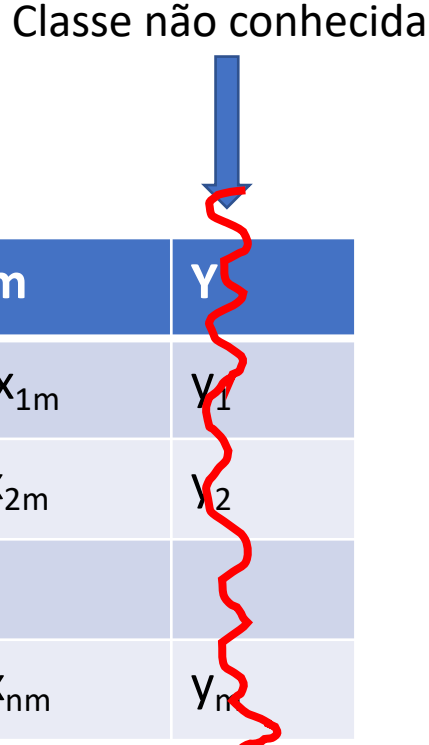
Aprendizado de Máquina não Supervisionado

- Formato dos dados para aprendizado não supervisionado

- Tabela atributo-valor

- Os exemplos são ditos **não** rotulados

- A classe **não** é conhecida



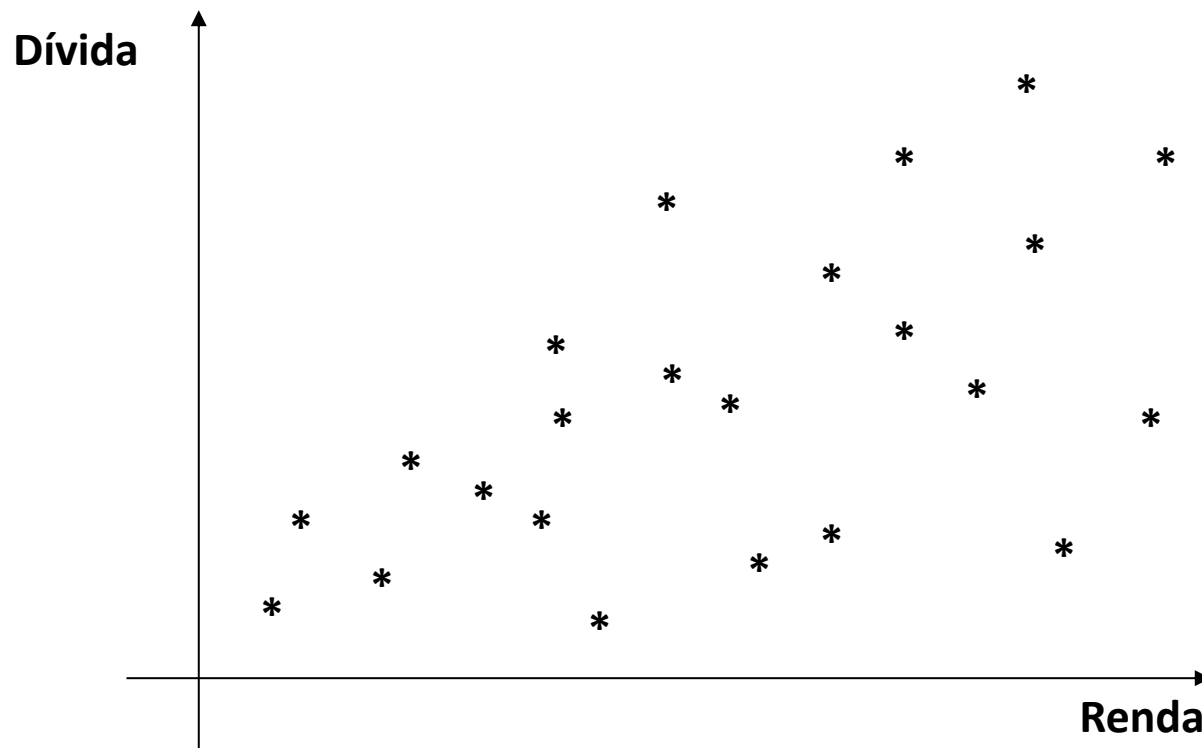
Exemplos	X1	X2	Xm	Y
E1	x_{11}	x_{12}	x_{1m}	y_1
E2	x_{21}	x_{22}	x_{2m}	y_2
En	x_{n1}	x_{n2}	x_{nm}	y_n

Agrupamento

- Agrupamento é a organização não supervisionada de dados (observações, instâncias, objetos, exemplos) em grupos (clusters).
- Intuitivamente, exemplos semelhantes devem estar no mesmo grupo e exemplos distintos devem estar em grupos distintos.
- Agrupamento é uma técnica dirigida por dados, isto é, não se conhece previamente a que grupo os dados pertencem.

Agrupamento

- Representação Gráfica de dados para agrupamento

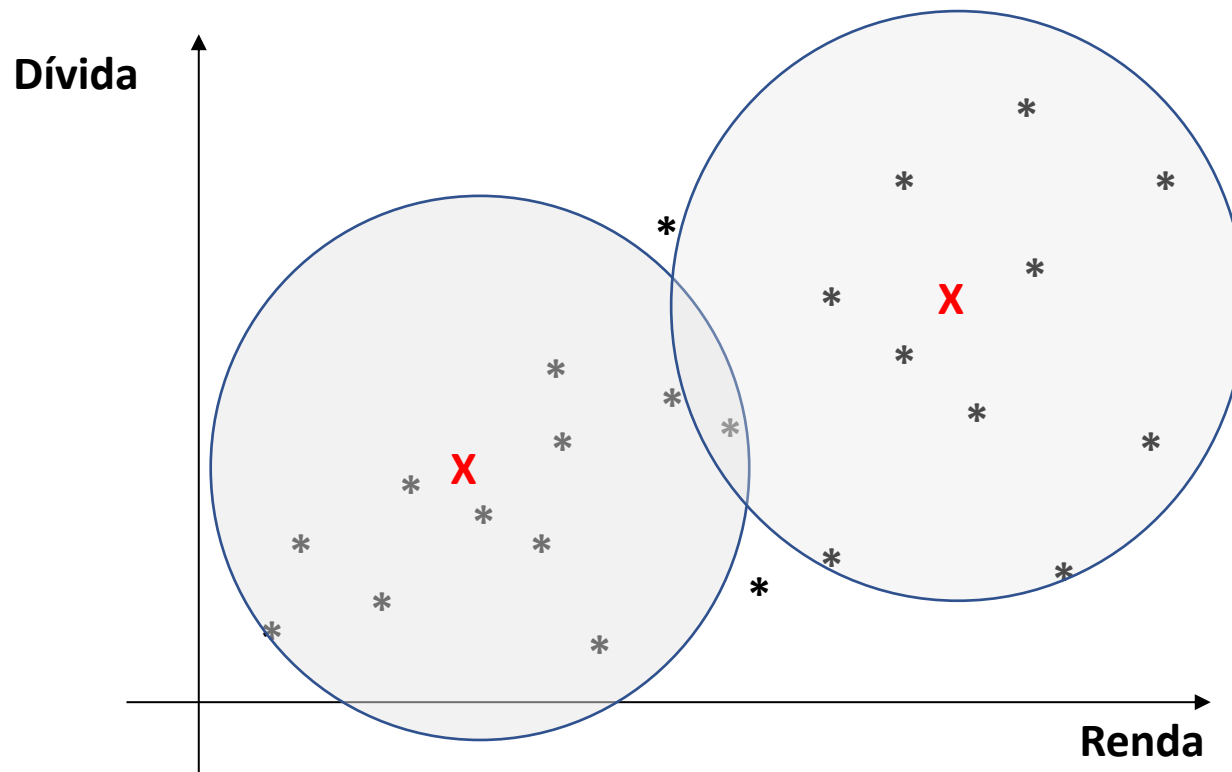


Dados no formato atributo-valor:

Exemplos	Renda	Dívida
E1	30	40
E2	40	20
E3	80	60
E4	35	30
...

Agrupamento

- Representação Gráfica de dados para agrupamento



Dados no formato atributo-valor:

Exemplos	Renda	Dívida
E1	30	40
E2	40	20
E3	80	60
E4	35	30
...

Regras de associação

- Objetivo:
 - Encontrar tendências que possam ser usadas para entender e explorar padrões de comportamento dos dados.
- Exemplo:
 - Uma Base de Dados de um supermercado teria como regra o fato de que 80% dos clientes que compram um produto Q, também adquirem, na mesma ocasião o produto W. Em que 80% é o fator de confiabilidade da regra.
- Problema:
 - Analisar um grande volume de conhecimento extraído no formato de regras.

Regras de associação

$A = \{\text{bermuda, calça, camiseta, sandália, tênis}\}$ e

$T = \{1, 2, 3, 4\}$.

Suporte Mínimo = 50% (2 transações).

Confiança Mínima = 50%.

Transações	Itens Comprados
1	calça, camiseta, tênis
2	camiseta, tênis
3	bermuda, tênis
4	calça, sandália

<i>Itemsets</i> Frequentes	Suporte
{tênis}	75%
{calça}	50%
{camiseta}	50%
{camiseta, tênis}	50%

Regras de associação

Suporte Mínimo = 50% (2 transações)

Confiança Mínima = 50%

***tênis* → *camiseta*,**
em que:

<i>Itemsets</i> Freqüentes	Suporte
{tênis}	75%
{calça}	50%
{camiseta}	50%
{camiseta, tênis}	50%

$$\text{suporte} = \text{suporte}(\{\text{tênis}, \text{camiseta}\}) = 50\%$$

$$\text{confiança} = \frac{\text{suporte}(\{\text{tênis}, \text{camiseta}\})}{\text{suporte}(\{\text{tênis}\})} = \frac{50}{75} = 66,6\%$$

- Próxima aula:
- Aprendizado Supervisionado: Árvores de Decisão