



Introdução à Inteligência Artificial

Módulo 03: Aprendizado de Máquina (*Machine Learning* - ML)

Parte 3 – Aprendizado Supervisionado

Professores: Rafael de Magalhães Dias Frinhani
Rodrigo Silva Lima



UNIFEI



Softex



MCTI
FUTURO

FUTURO DO TRABALHO, TRABALHO DO FUTURO



Sumário do Módulo 03

Parte 3 – Aprendizado Supervisionado

1. Taxonomia dos Algoritmos de Aprendizado de Máquina
2. Aprendizado Supervisionado
3. Algoritmos de Aprendizado Supervisionado
 - 3.1. *Decision Tree*
 - 3.2. *Ensemble Learning – Random Forest, GBDT*
 - 3.3. *K-Nearest Neighbors*
 - 3.4. *Näive Bayes*
 - 3.5. *Regressão Linear*
 - 3.6. *Regressão Logística*
 - 3.7. *Suport Vector Machine*



UNIFEI



Softex



MCTI
FUTURO

FUTURO DO TRABALHO. TRABALHO DO FUTURO

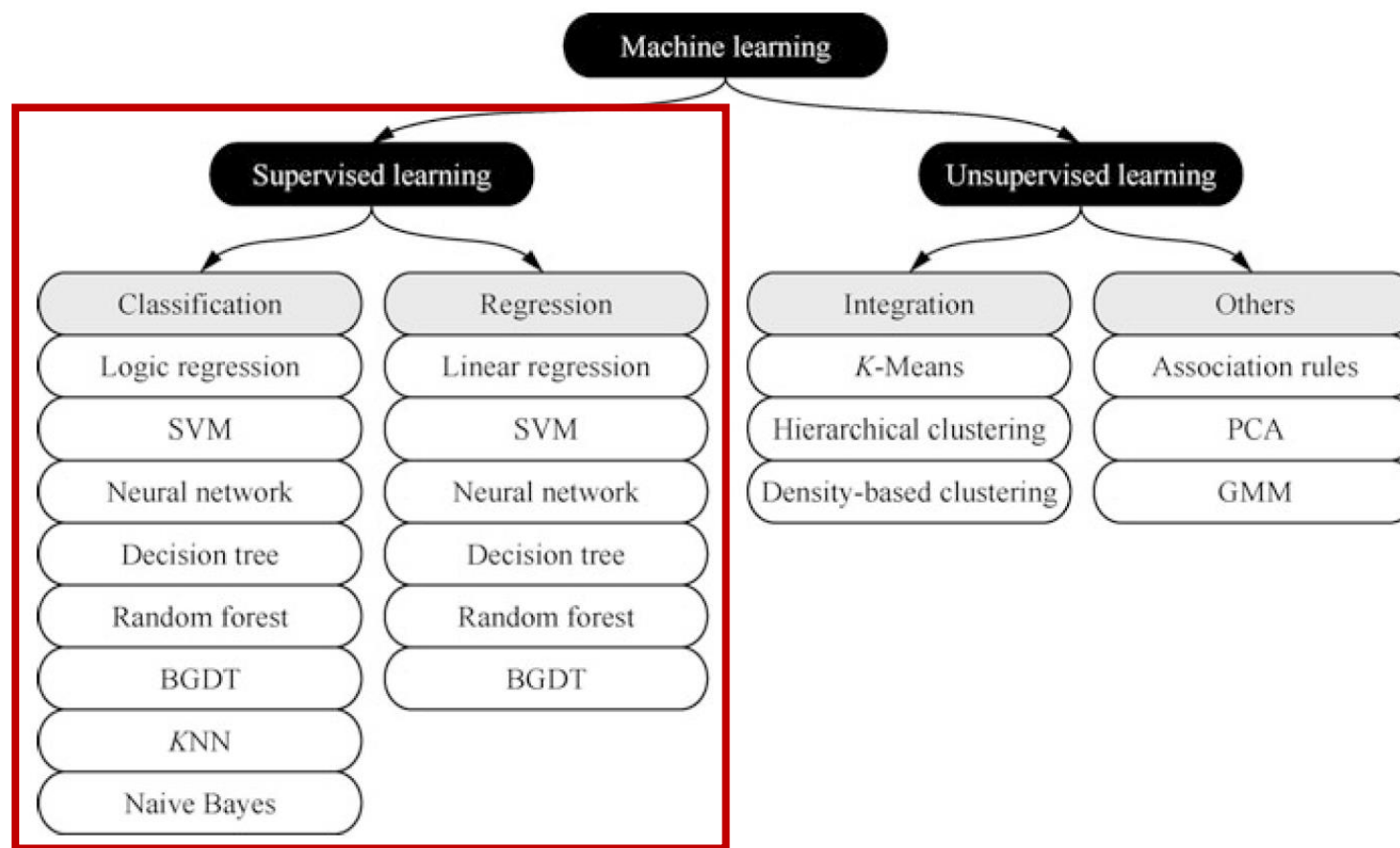


Objetivos do Módulo

- Este módulo tem por objetivo apresentar uma **visão geral do aprendizado Supervisionado** para a tarefa de classificação e regressão.
- Entender a **importância dos dados na obtenção de modelos supervisionados de qualidade**, os problemas que podem comprometer o aprendizado e como contorná-los, as formas de divisão do conjunto de dados para o treinamento e teste do modelo.
- Descrever o **funcionamento básico dos algoritmos Supervisionados** *Decision Tree*, *Ensemble Learning* (*Random Forest*, *GBDT*), *K-Nearest Neighbors*, *Näive Bayes*, *Regressão Linear*, *Regressão Logística* e *Support Vector Machine* (SVM).

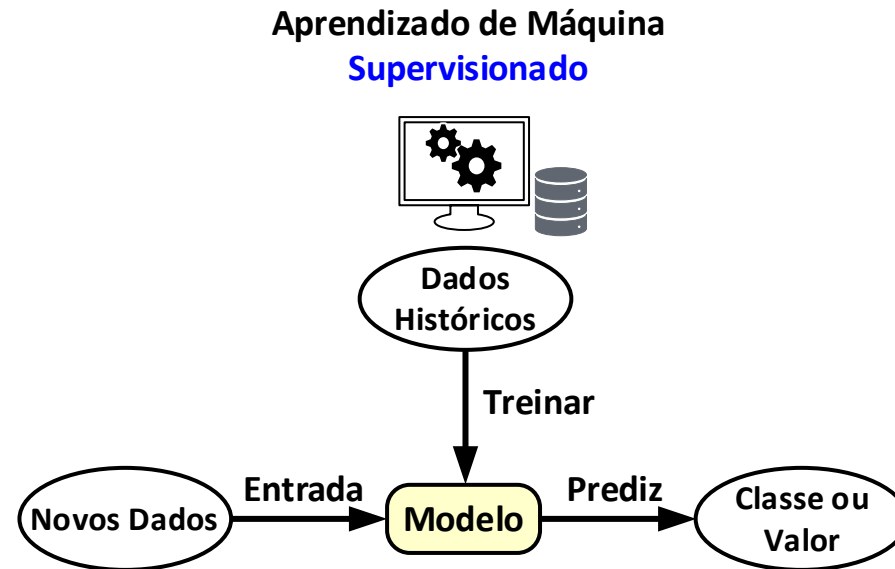
1. Taxonomia dos Algoritmos de Aprendizado de Máquina

O **aprendizado** demanda uma **interação entre um agente** que receberá o conhecimento (aluno) e o **ambiente** (professor). Os paradigmas de aprendizagem variam conforme o papel desempenhado pelo agente, cujos **métodos podem ser organizados conforme a natureza dessa interação**.



2. Aprendizado Supervisionado

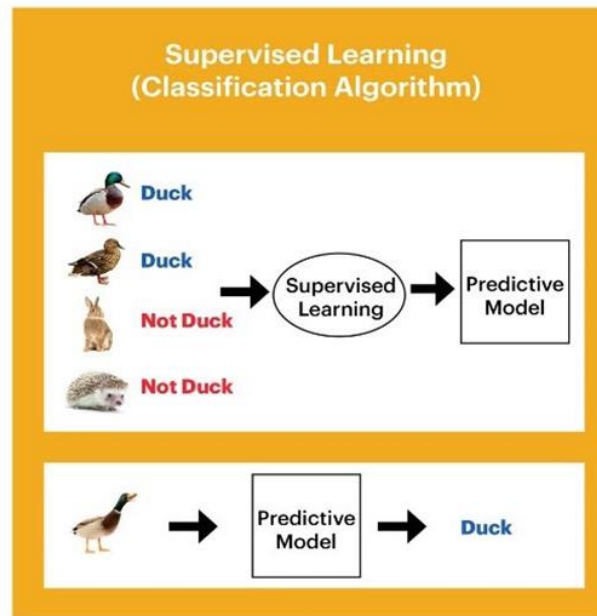
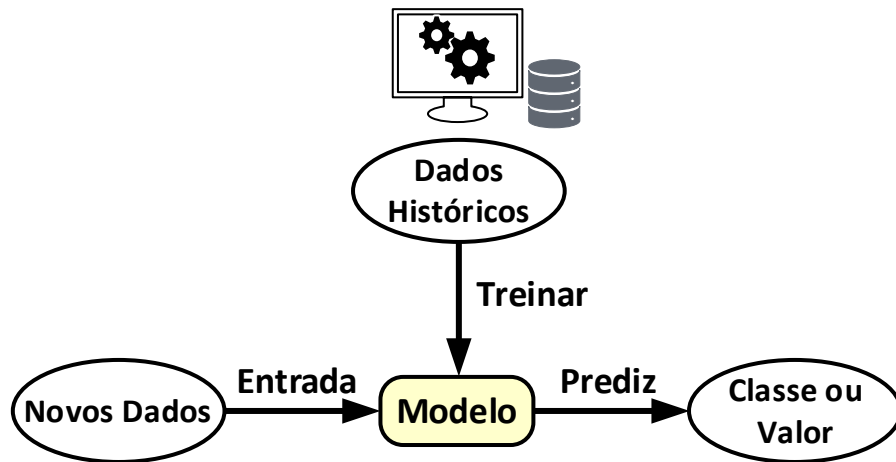
Considera um **agente** que **se baseia na experiência** (dados históricos), que possui os padrões que devem ser considerados **para prever valores faltantes** nos atributos de novos dados.



2. Aprendizado Supervisionado

Considera um agente que se baseia na experiência (dados históricos), que possui os padrões que devem ser considerados para prever valores faltantes nos atributos de novos dados.

Aprendizado de Máquina Supervisionado



PREVISÃO

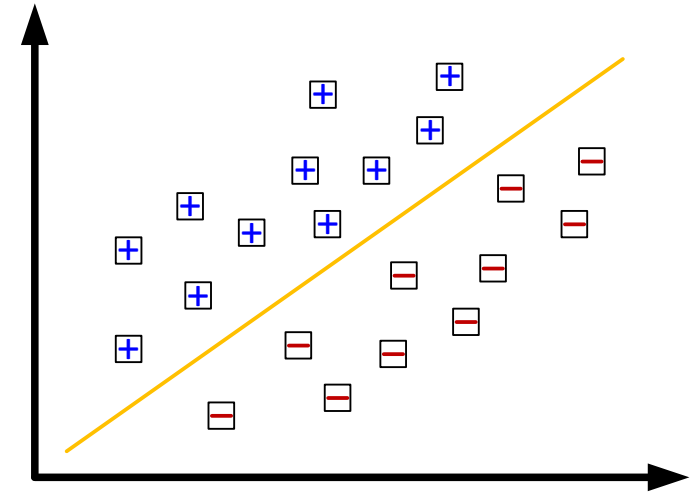
Classificação: valores nominais (rótulos de classe).

Regressão: valores numéricos.

Algoritmo Supervisionado: Métodos preditivos capazes de deduzir valores (nominais ou numéricos) a partir de dados históricos, que realizam a tarefa de classificação ou regressão.

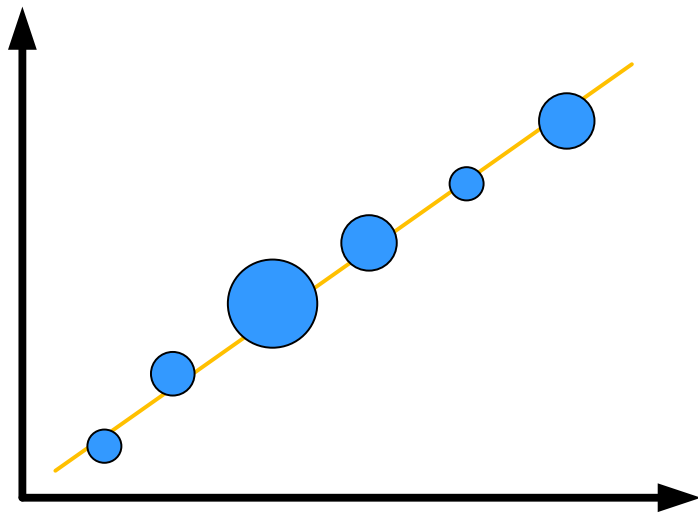
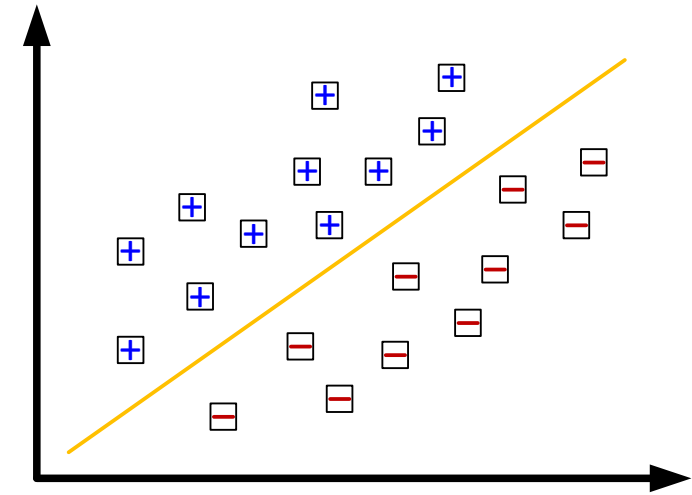
2. Aprendizado Supervisionado

Classificação – identifica qual das classes pré-existent em um conjunto de dados, um novo objeto estaria associado. O resultado são números discretos que correspondem as classes de objetos (rótulos ou categorias). Ex. classificação de imagens, análises de risco financeiro de clientes, detecção de lixo etc.



2. Aprendizado Supervisionado

Classificação – identifica qual das classes pré-existent em um conjunto de dados, um novo objeto estaria associado. O resultado são números discretos que correspondem as classes de objetos (rótulos ou categorias). Ex. classificação de imagens, análises de risco financeiro de clientes, detecção de lixo etc.



Regressão – prevê valores indisponíveis de atributos numéricos e contínuos de objetos de dados. O modelo prevê uma variável numérica (**dependente**) a partir de uma ou mais **variáveis independentes** (atributos), com base na **correlação** entre os valores conhecidos. Ex. identificação de tendências (séries temporais), prever o valor do sinistro de um segurado com base no seu perfil.

2. Aprendizado Supervisionado

Características

- Abrange os métodos que realizam as tarefas de **classificação** ou **regressão** de dados.
- Útil quando **se conhece** quais são as **classes** de objetos e se deseja prever os valores de atributos nominais ou numéricos.
- A classificação e a regressão **tem propósito preditivo**.

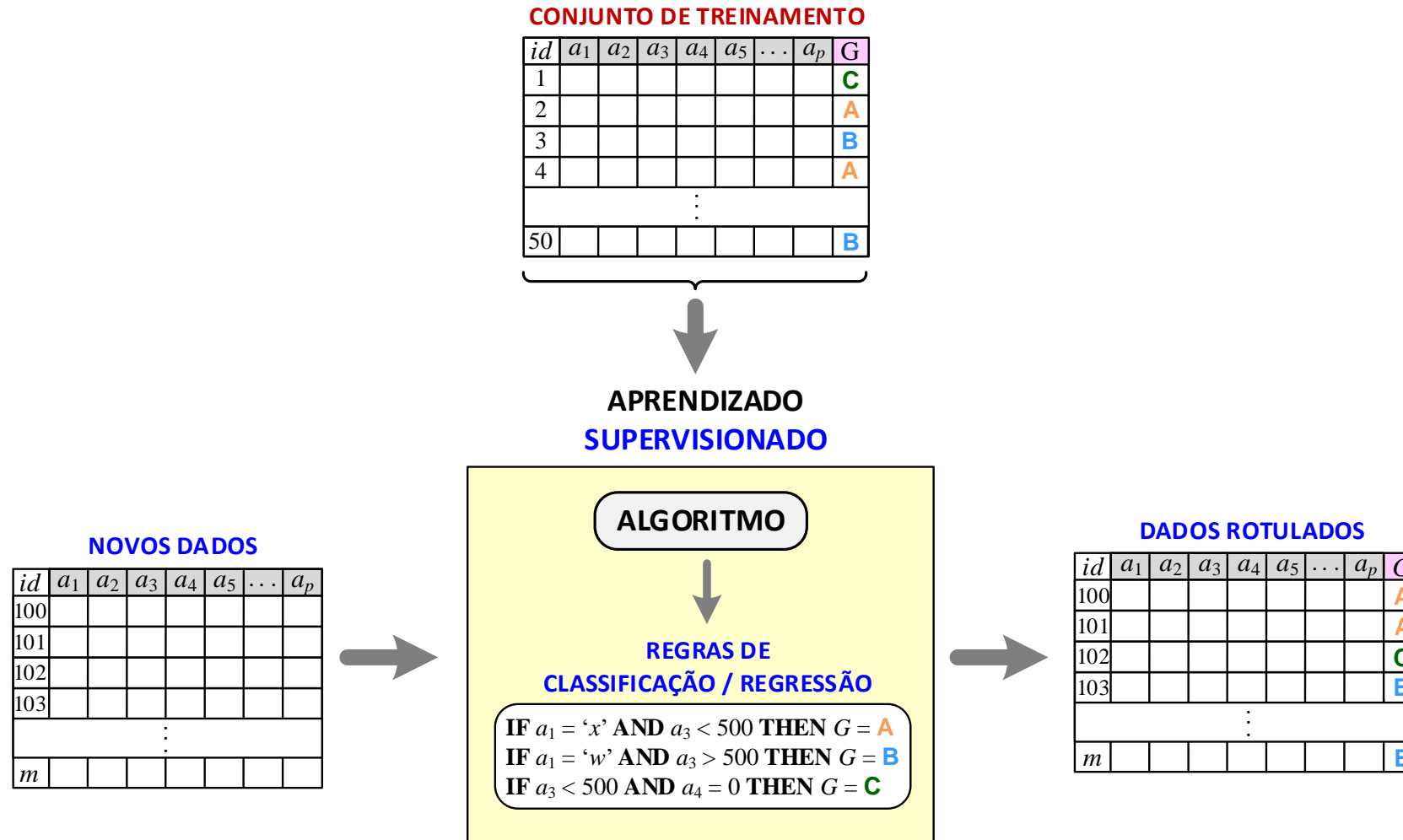
2. Aprendizado Supervisionado

Características

- Abrange os métodos que realizam as tarefas de classificação ou regressão de dados.
- Útil quando se conhece quais são as classes de objetos e se deseja prever os valores de atributos nominais ou numéricos.
- A classificação e a regressão tem propósito preditivo.
- Os algoritmos supervisionados examinam os dados para entender os padrões, utilizando essa informação para inferir a classe ou valor de um atributo de um novo objeto de dados.
- Quando treinado com mais dados de qualidade, o modelo baseado em um algoritmo supervisionado tende a melhorar seu desempenho preditivo.
- Problemas de classificação podem ser reduzidos a um de regressão. Ex. modelo pode classificar um objeto considerando a classe de um atributo com maior probabilidade dele pertencer .

2. Aprendizado Supervisionado

Modelo genérico de um método de Classificação / Regressão de Dados



2. Aprendizado Supervisionado

Modelo genérico de um método de Classificação / Regressão de Dados

No treinamento do algoritmo são usados dados em que se conhece a classe do objeto ou valor de seus atributos.

CONJUNTO DE TREINAMENTO

<i>id</i>	a_1	a_2	a_3	a_4	a_5	\dots	a_p	G
1								C
2								A
3								B
4								A
\vdots								
50								B

Tem como entrada novos dados cujos objetos não possuem rótulos de classe ou valor em um dado atributo.

NOVOS DADOS

<i>id</i>	a_1	a_2	a_3	a_4	a_5	\dots	a_p
100							
101							
102							
103							
\vdots							
m							

APRENDIZADO
SUPERVISIONADO

ALGORITMO

REGRAS DE
CLASSIFICAÇÃO / REGRESSÃO

IF $a_1 = 'x'$ AND $a_3 < 500$ THEN $G = A$
IF $a_1 = 'w'$ AND $a_3 > 500$ THEN $G = B$
IF $a_3 < 500$ AND $a_4 = 0$ THEN $G = C$

DADOS ROTULADOS

<i>id</i>	a_1	a_2	a_3	a_4	a_5	\dots	a_p	G
100								A
101								A
102								C
103								B
\vdots								
m								B

2. Aprendizado Supervisionado

Modelo genérico de um método de Classificação / Regressão de Dados

No treinamento do algoritmo são usados dados em que se conhece a classe do objeto ou valor de seus atributos.

CONJUNTO DE TREINAMENTO

id	a_1	a_2	a_3	a_4	a_5	\dots	a_p	G
1								C
2								A
3								B
4								A
\vdots								
50								B

Tem como entrada novos dados cujos objetos não possuem rótulos de classe ou valor em um dado atributo.

NOVOS DADOS

id	a_1	a_2	a_3	a_4	a_5	\dots	a_p
100							
101							
102							
103							
\vdots							
m							

APRENDIZADO
SUPERVISIONADO

ALGORITMO

REGRAS DE
CLASSIFICAÇÃO / REGRESSÃO

IF $a_1 = 'x'$ AND $a_3 < 500$ THEN $G = A$
IF $a_1 = 'w'$ AND $a_3 > 500$ THEN $G = B$
IF $a_3 < 500$ AND $a_4 = 0$ THEN $G = C$

O algoritmo descobre as regras de classificação ou regressão considerando os atributos dos objetos.

Padrões identificados pelo algoritmo são usados para inferir valores (nominais ou numéricos) dos novos dados.

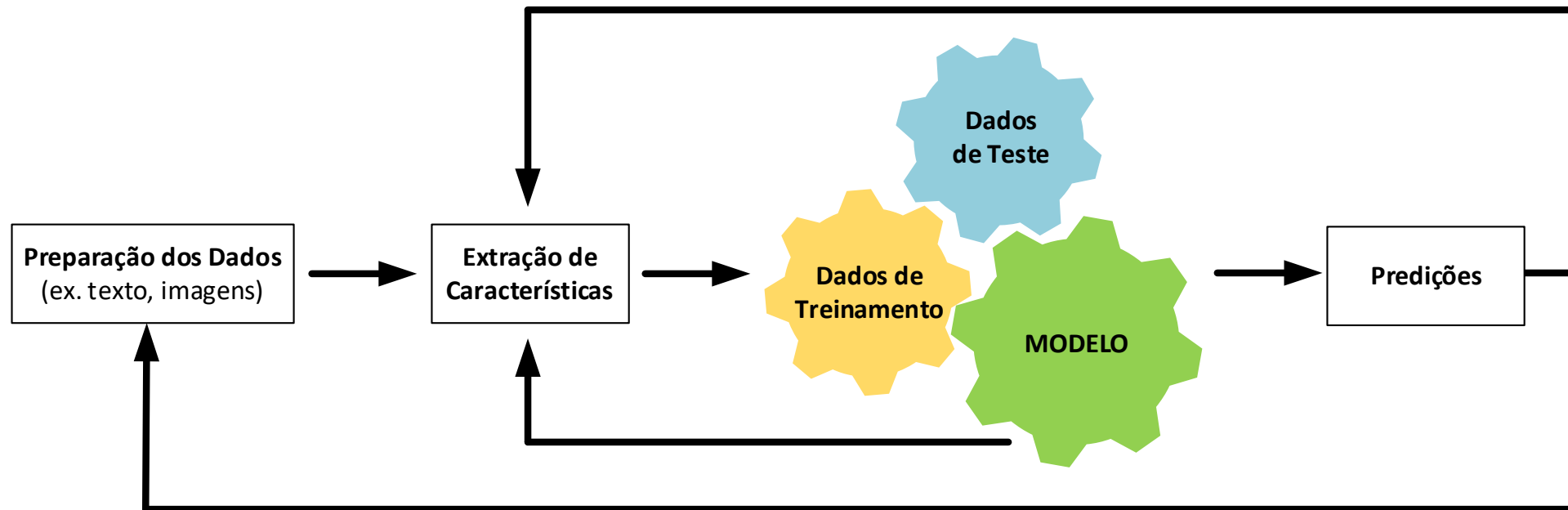
DADOS ROTULADOS

id	a_1	a_2	a_3	a_4	a_5	\dots	a_p	G
100								A
101								A
102								C
103								B
\vdots								
m								B

2. Aprendizado Supervisionado

Pipeline de um Aprendizado Supervisionado Genérico

Um *pipeline* do aprendizado supervisionado corresponde a **execução de uma sequência linear de módulos especializados** (ex. preparação de dados, dados de treinamento) para a obter previsões. **Novos dados podem** ser usados para o treinamento do modelo para **melhorar seu desempenho**.



2. Aprendizado Supervisionado

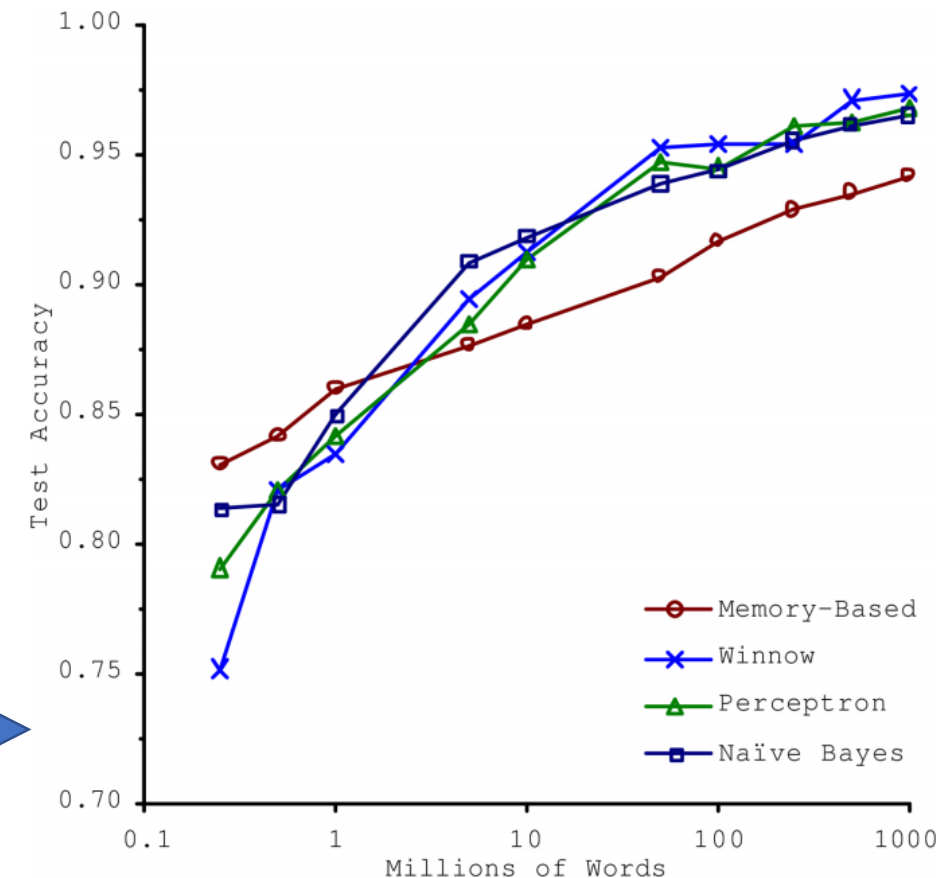
Importância dos Dados

Dados de qualidade são fundamentais para obtenção de modelos de qualidade. **Alguns problemas observados nos dados:**

Dados insuficientes:

Dependendo do problema, **pode ser necessário uma grande quantidade de dados** para que algoritmo de aprendizado de máquina funcione adequadamente (estatisticamente válido).

O gráfico ao lado mostra a **relação entre o aumento da quantidade de dados e o aumento da acurácia** de quatro modelos de aprendizado supervisionado .



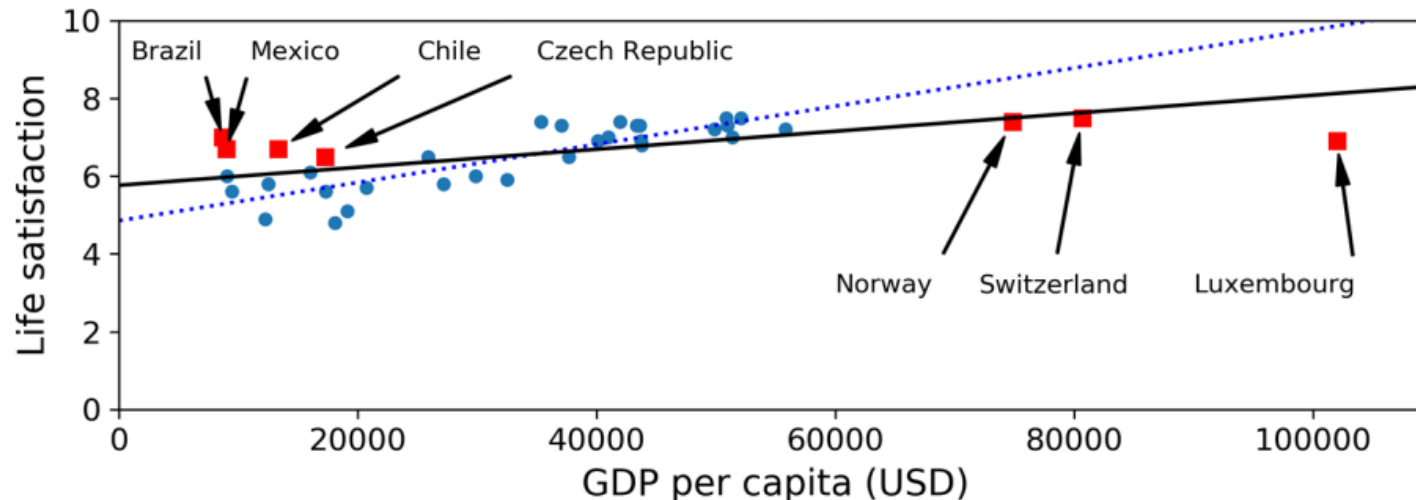
2. Aprendizado Supervisionado

Importância dos Dados

Dados de qualidade são fundamentais para obtenção de modelos de qualidade. **Alguns problemas observados nos dados:**

Dados não-representativos:

Os dados de treinamento devem ser boas representações dos novos casos.



◀ O gráfico ao lado mostra a relação entre o Produto Interno Bruto (*Gross Domestic Product*, GDP) per capita e o nível de satisfação de vida de determinados países.

Dependendo do objetivo do problema, o uso de determinados dados pode impossibilitar a obtenção de tendências desejadas.

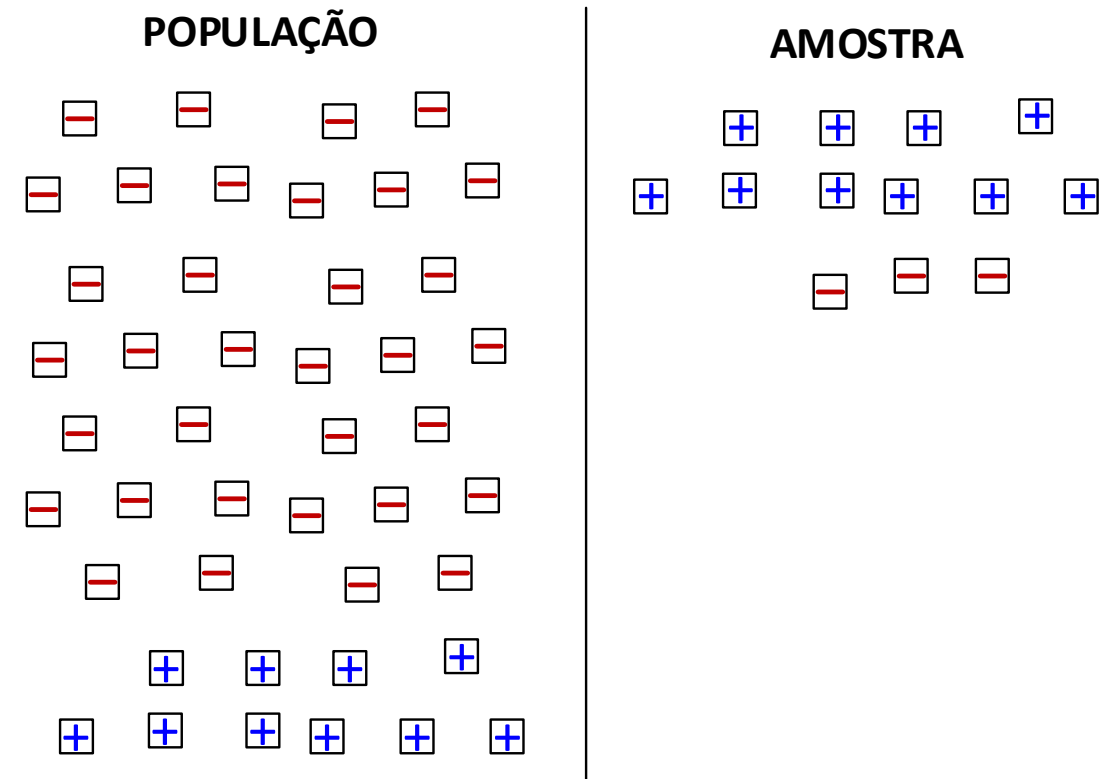
2. Aprendizado Supervisionado

Importância dos Dados

Dados de qualidade são fundamentais para obtenção de modelos de qualidade. **Alguns problemas observados nos dados:**

Viés de Amostragem:

Ocorre quando há uma **incoerência das observações** de um segmento da população, implicando em uma maioria (ou minoria) de indivíduos de uma determinada classe que **não condiz com a realidade da população original**.



2. Aprendizado Supervisionado

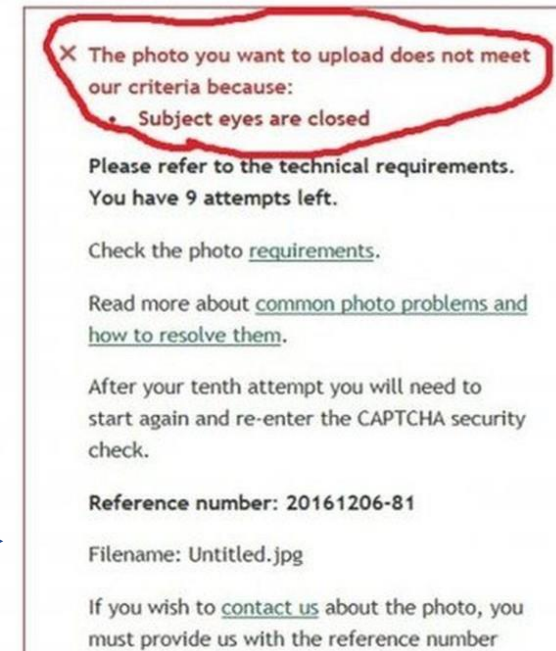
Importância dos Dados

Dados de qualidade são fundamentais para obtenção de modelos de qualidade. **Alguns problemas observados nos dados:**

Viés de Amostragem:

Ocorre quando há uma incoerência das observações de um segmento da população, implicando em uma maioria (ou minoria) de indivíduos de uma determinada classe que não condiz com a realidade da população original.

Na situação ao lado, o modelo provavelmente não foi treinado com imagens de asiáticos (ou a amostra era baixa). O modelo identifica um olho fechado, mas não identifica um asiático.



2. Aprendizado Supervisionado

Divisão do Conjunto de Dados

A divisão do conjunto em treinamento e teste pode influenciar nos ajustes de generalização do modelo. A % de divisão não é rígida, sendo comum 80/20 (depende da quantidade de dados).

2. Aprendizado Supervisionado

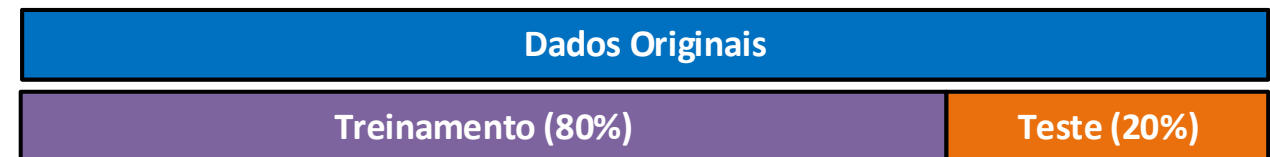
Divisão do Conjunto de Dados

A divisão do conjunto em treinamento e teste pode influenciar nos ajustes de generalização do modelo. A % de divisão não é rígida, sendo comum 80/20 (depende da quantidade de dados).

Hold Out

- O conjunto original é dividido em treinamento e teste.
- O modelo é ajustado conforme o conjunto de treinamento.
- O conjunto de teste permite mensurar o quão bem o modelo atua em dados desconhecidos.

Problema: Difícil validar a acurácia do treinamento devido a amostra mais restrita, e o modelo pode apresentar baixa generalização.



2. Aprendizado Supervisionado

Divisão do Conjunto de Dados

A divisão do conjunto em treinamento e teste pode influenciar nos ajustes de generalização do modelo. A % de divisão não é rígida, sendo comum 80/20 (depende da quantidade de dados).

Validation

- O conjunto original é dividido em treinamento, **validação** e teste.
- Útil para **otimizar os hiperparâmetros**, ou para **comparar o desempenho** de diferentes modelos.
- O conjunto de **validação** não é usado para treinar os modelos, mas para **calcular os erros** e **auxiliar no ajuste dos hiperparâmetros** visando diminuir esses erros.

Problema: O modelo pode apresentar boa acurácia em novos dados com atributos similares aos usados na validação.



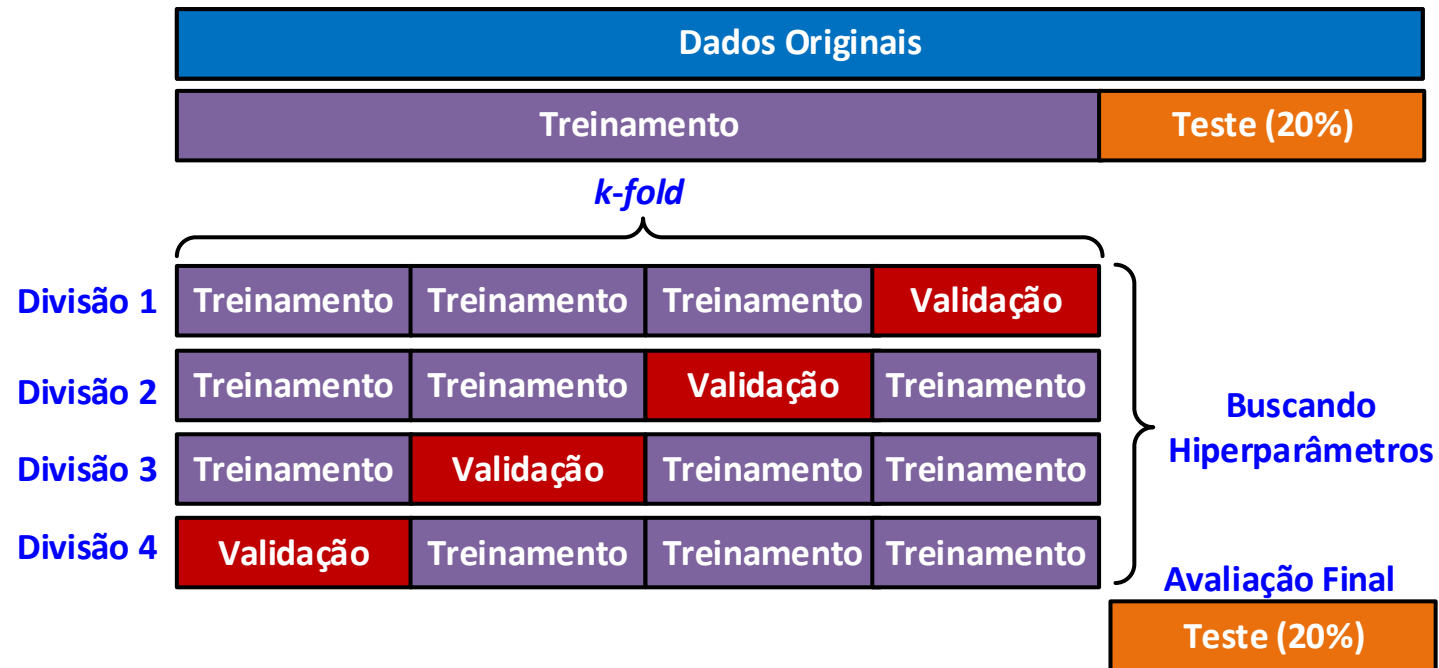
2. Aprendizado Supervisionado

Divisão do Conjunto de Dados

A divisão do conjunto em treinamento e teste pode influenciar nos ajustes de generalização do modelo. A % de divisão não é rígida, sendo comum 80/20 (depende da quantidade de dados).

k-fold Cross Validation (Validação Cruzada)

- O conjunto original é dividido em treinamento, validação e teste.
- a cada iteração um *fold* diferente é usado para validação e os demais para treinamento.
- cada amostra é usada o mesmo número de vezes para treinamento e apenas uma vez para validação.



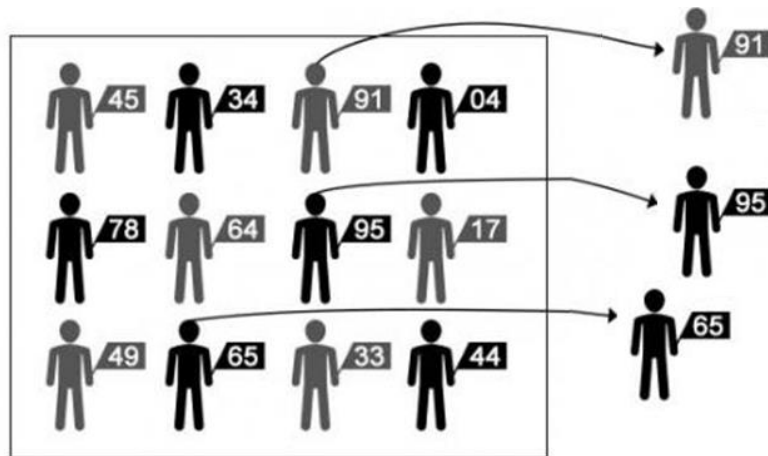
2. Aprendizado Supervisionado

Seleção dos Dados de Treinamento e Teste

A forma de selecionar os objetos de dados que irão compor cada conjunto **pode contribuir** na obtenção de um modelo com **maior** capacidade de **generalização**.

Amostragem Aleatória Simples

Os objetos de dados são aleatoriamente escolhidos (ex. teste) considerando **todos os dados do conjunto**.



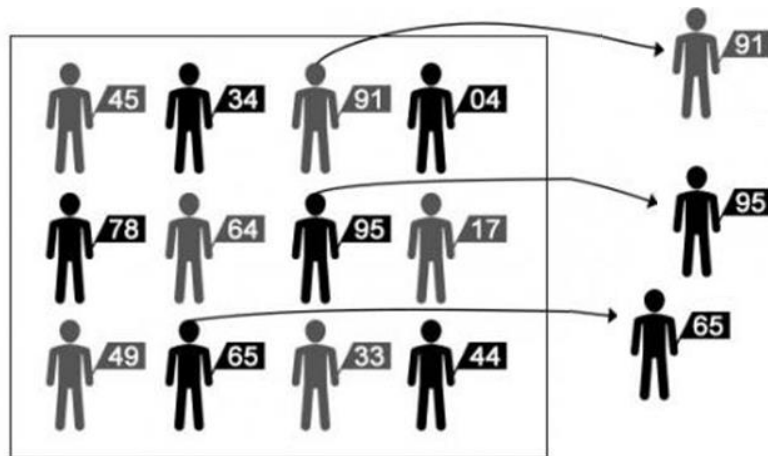
2. Aprendizado Supervisionado

Seleção dos Dados de Treinamento e Teste

A forma de selecionar os objetos de dados que irão compor cada conjunto pode contribuir na obtenção de um modelo com maior capacidade de generalização.

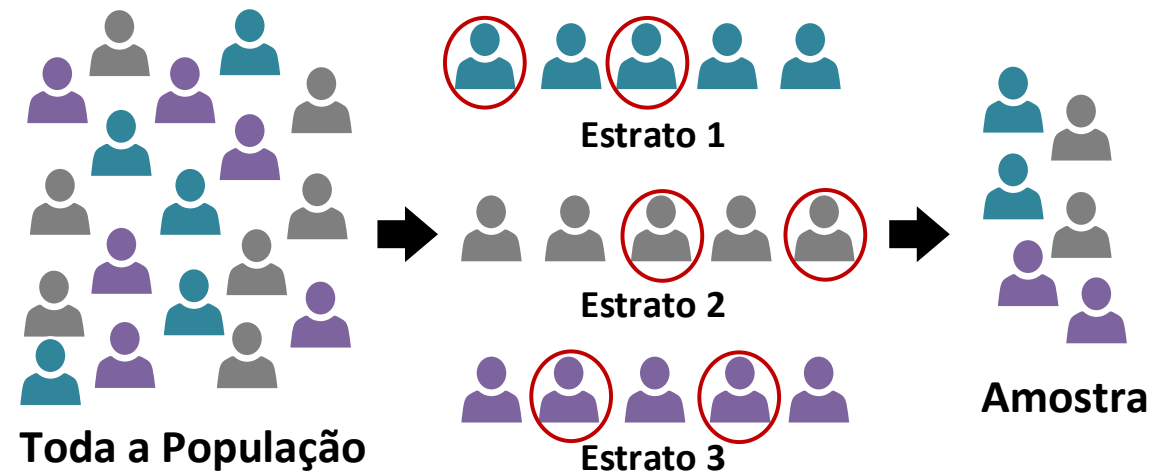
Amostragem Aleatória Simples

Os objetos de dados são aleatoriamente escolhidos (ex. teste) considerando todos os dados do conjunto.



Amostragem Aleatória Estratificada

Objetos similares são organizados em grupos (*estratos*), escolhe-se aleatoriamente de forma **proporcional ao tamanho do estrato**.

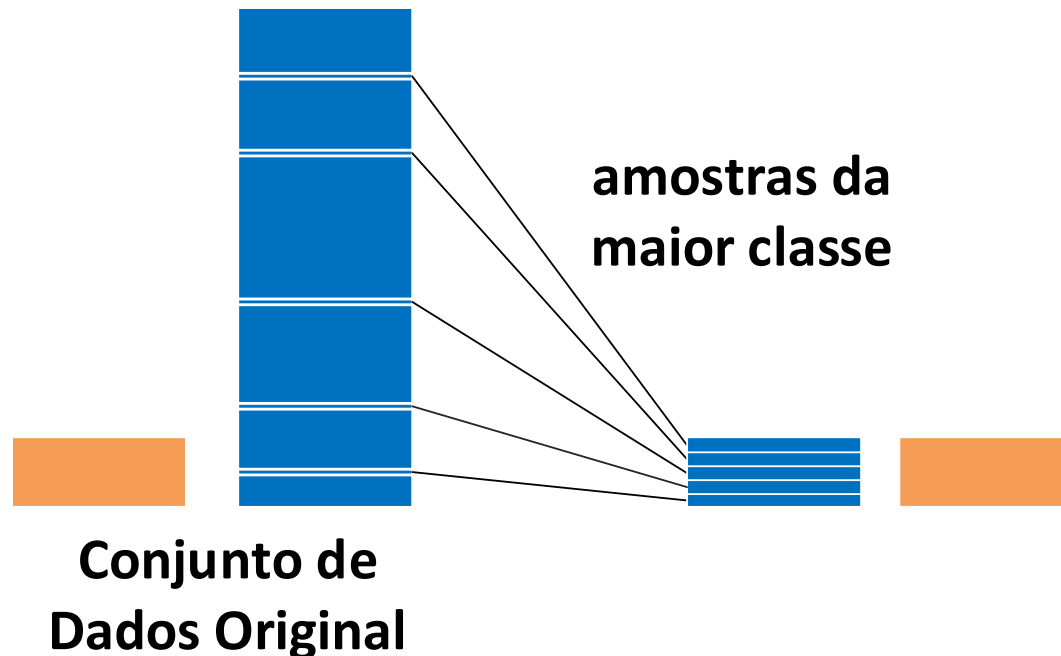


2. Aprendizado Supervisionado

Desbalanceamento de Classes

Ocorre quando a quantidade de objetos entre as classes é discrepante, podendo enviesar a previsão ao favorecer as classes maiores. Os métodos para o balanceamento de classes:

Undersampling

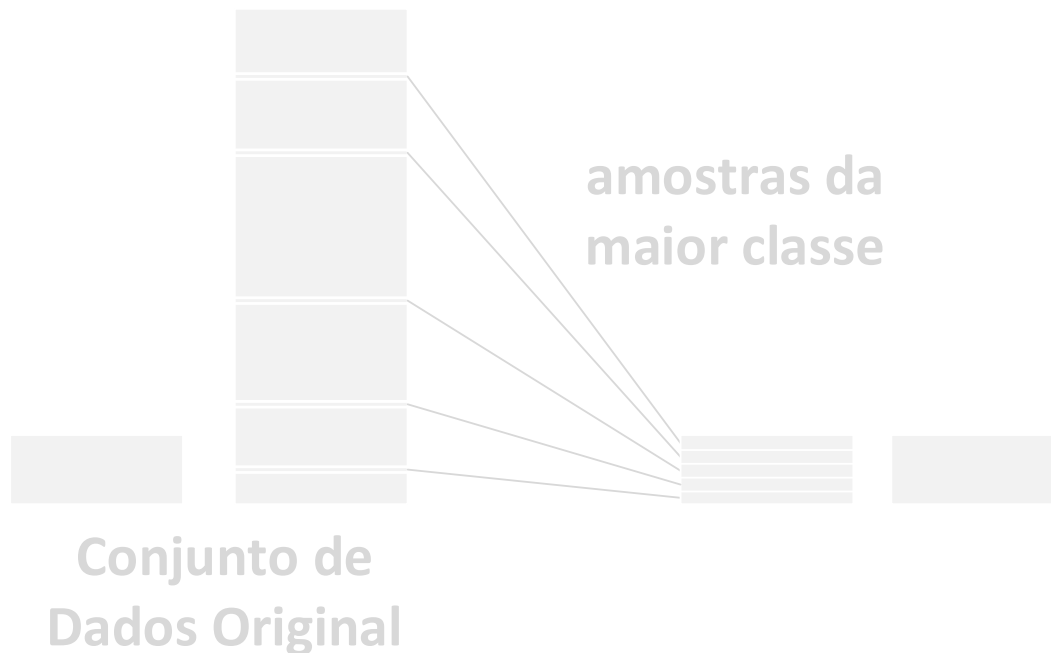


2. Aprendizado Supervisionado

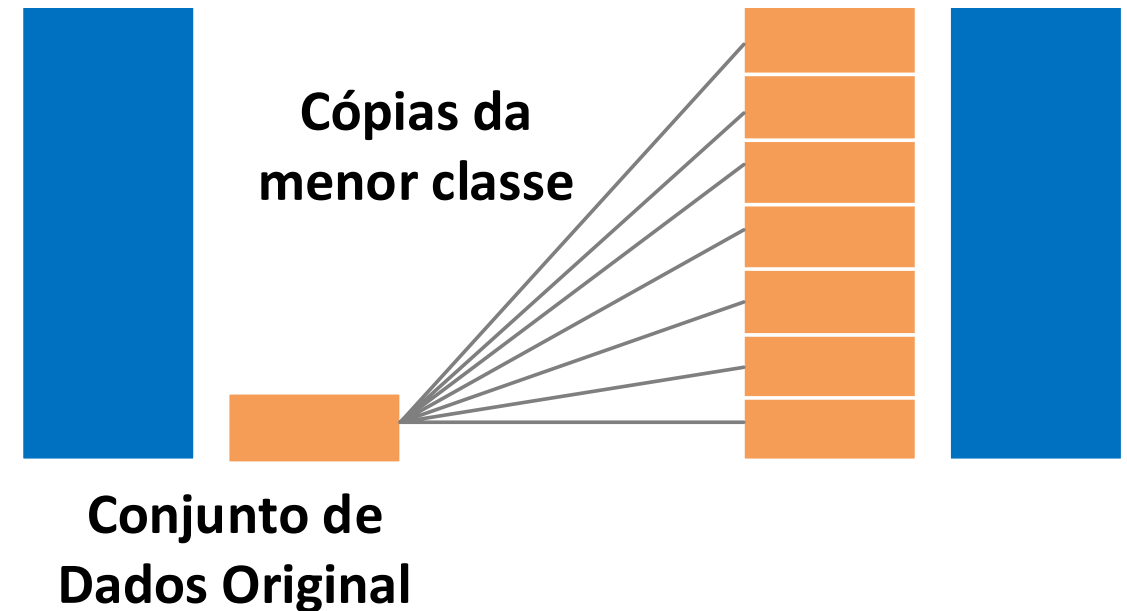
Desbalanceamento de Classes

Ocorre quando a quantidade de objetos entre as classes é discrepante, podendo enviesar a previsão ao favorecer as classes maiores. Os **métodos para o balanceamento de classes**:

Undersampling



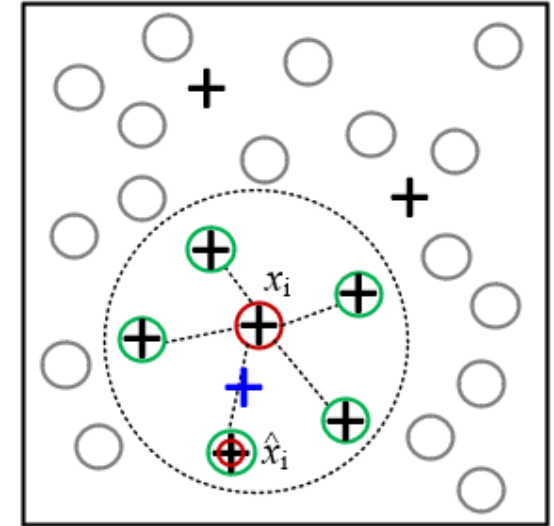
Oversampling



2. Aprendizado Supervisionado

SMOTE (Synthetic Minority Oversampling Technique)

- Gera novas **instâncias sintéticas** a partir **de casos minoritários**, não altera os casos majoritários.
- As novas instâncias não são simples cópias dos dados originais. **São geradas** instâncias virtuais de treinamento **por interpolação linear** à partir da classe minoritária.
- Os registros sintéticos são gerados selecionando aleatoriamente um ou mais dos **k-vizinhos mais próximos** para cada exemplo na classe minoritária.



LEGENDA

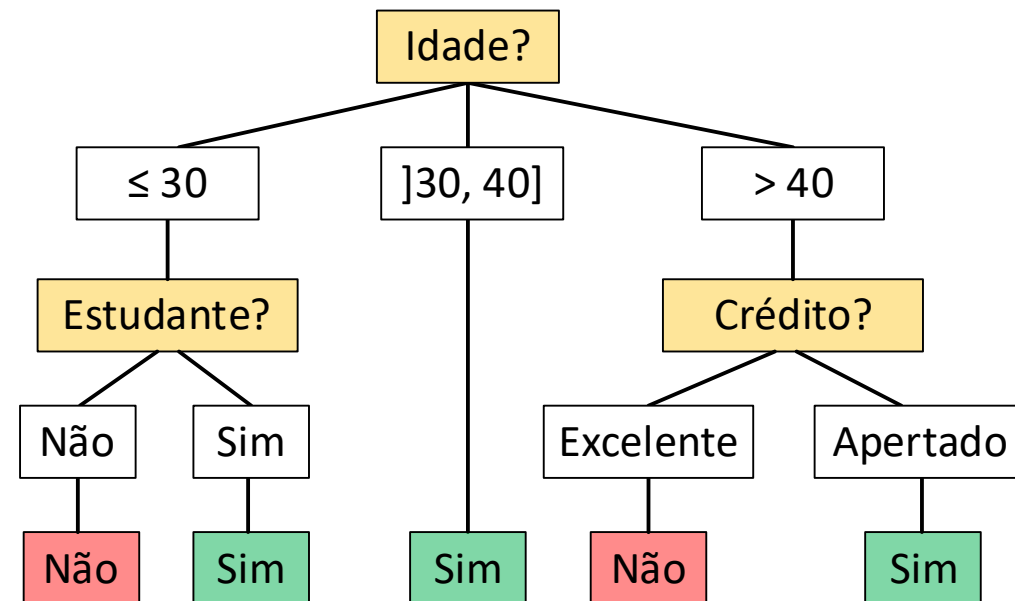
- Amostras da classe majoritária
- ⊕ Amostras da classe minoritária
- ⊕ (red border) Amostra x_i selecionada aleatoriamente da classe minoritária
- ⊕ (green border) Os 5 vizinhos mais próximos de x_i
- ⊕ (green border, red border) Amostra \hat{x}_i selecionada aleatoriamente dos 5 vizinhos.
- ⊕ (blue) Instância minoritária sintética gerada

3. Algoritmos de Aprendizado Supervisionado

DECISION TREE

Aprendizado baseado em **decisões hierárquicas sobre os atributos**, dispostas em uma estrutura em forma de árvore (binária ou não-binária). Pode ser usada para **tarefas de classificação e regressão**.

Exemplos da **algoritmos** CART, C4.5, C5.0, ID3.



Exemplo de árvore de decisão para previsão de potencial cliente para compra de computador com base no seu perfil.

3. Algoritmos de Aprendizado Supervisionado

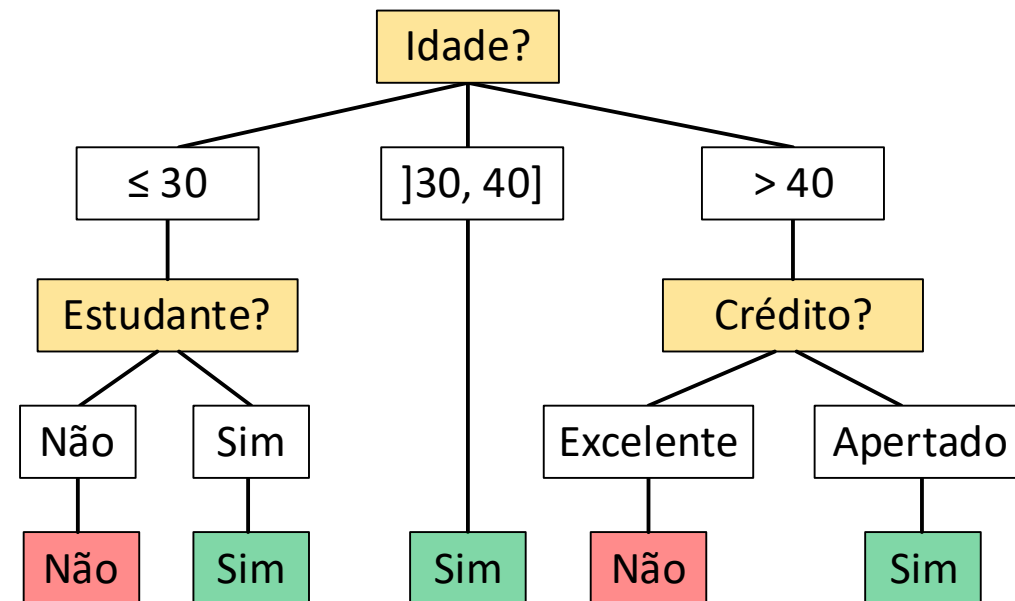
DECISION TREE

Aprendizado baseado em decisões hierárquicas sobre os atributos, dispostas em uma estrutura em forma de árvore (binária ou não-binária). Pode ser usada para tarefas de classificação e regressão.

Exemplos de algoritmos CART, C4.5, C5.0, ID3.

Construção da Árvore de Decisão:

1. São **selecionados os atributos** e determinada a estrutura da árvore com base na relação entre eles.
2. Os **dados são divididos** conforme os atributos. Nós filhos são gerados recursivamente (*top-down*), tendo o atributo mais “puro” como critério de divisão.
3. A **poda** (*prunning*) reduz o tamanho da árvore para diminuir o *overfitting* do modelo.

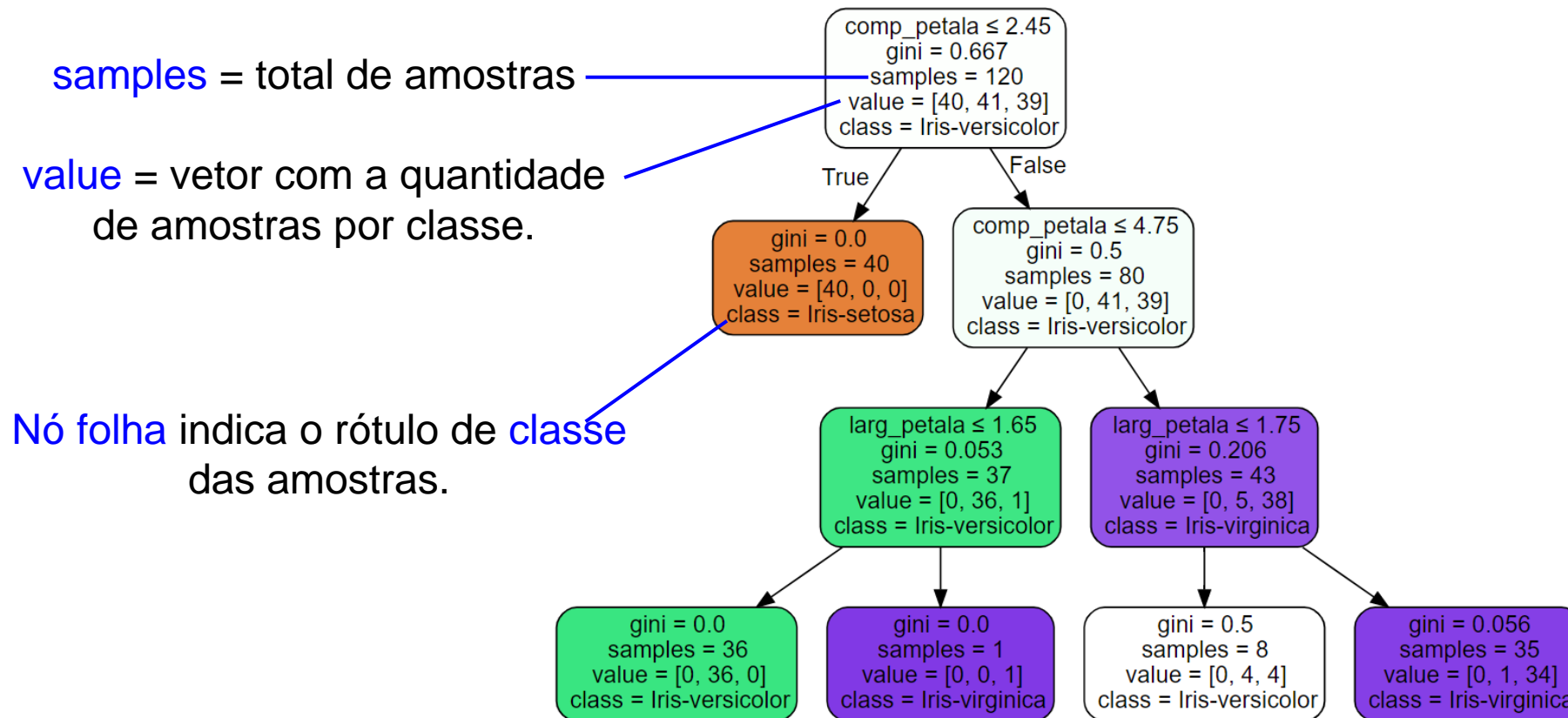


Exemplo de árvore de decisão para previsão de potencial cliente para compra de computador com base no seu perfil.

3. Algoritmos de Aprendizado Supervisionado

DECISION TREE

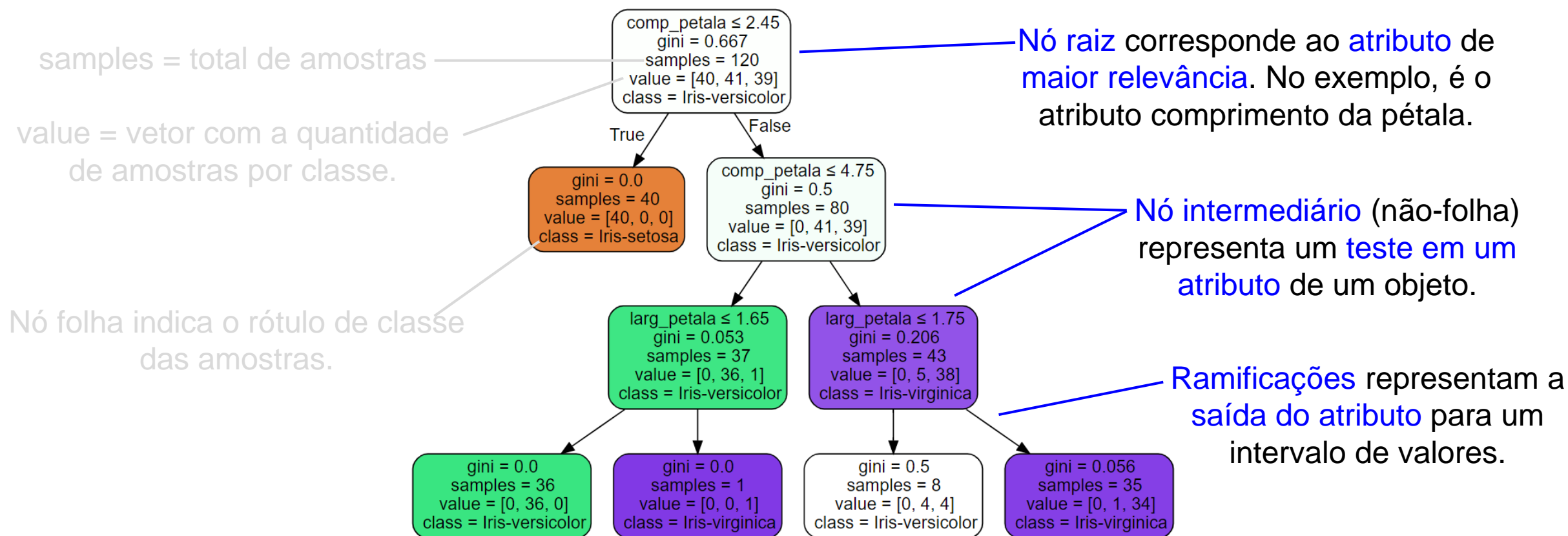
Exemplo de Árvore de decisão para o conjunto de dados Iris (UCI *Machine Learning Repository*)



3. Algoritmos de Aprendizado Supervisionado

DECISION TREE

Exemplo de Árvore de decisão para o conjunto de dados Iris (UCI *Machine Learning Repository*)



3. Algoritmos de Aprendizado Supervisionado

DECISION TREE – Critério de Divisão

A decisão de **como dividir um nó da árvore** é baseada no critério de divisão (*split criteria*), que reflete uma condição presente em uma ou mais variáveis preditivas dos dados de treinamento.

Seu objetivo é **maximizar a separação das** diferentes **classes entre os nós filhos**. É necessário escolher entre vários atributos e várias alternativas para dividi-los.

3. Algoritmos de Aprendizado Supervisionado

DECISION TREE – Critério de Divisão

A decisão de como dividir um nó da árvore é baseada no critério de divisão (*split criteria*), que reflete uma condição presente em uma ou mais variáveis preditivas dos dados de treinamento.

Seu objetivo é maximizar a separação das diferentes classes entre os nós filhos. É necessário escolher entre vários atributos e várias alternativas para dividi-los.

Métricas para análise da qualidade da divisão:

ENTROPIA

$$H(X) = - \sum_{k=1}^K p_k \log_2 (p_k)$$

K → quantidade de classes
 p_k → probabilidade da amostra pertencer à classe k

ÍNDICE GINI

$$Gini(X) = 1 - \sum_{k=1}^K p_k^2$$

3. Algoritmos de Aprendizado Supervisionado

Critério de Divisão – Exemplo de cálculo do Índice Gini

Cálculo do índice Gini para seleção do atributo mais relevante (nó raiz).

Decisão de jogar tênis considerando as condições climáticas.

DIA	AMBIENTE	TEMPERATURA	HUMIDADE	VENTO	DECISÃO
1	ensolarado	quente	alta	fraco	não
2	ensolarado	quente	alta	forte	não
3	nublado	quente	alta	fraco	sim
4	chuva	moderado	alta	fraco	sim
5	chuva	frio	normal	fraco	sim
6	chuva	frio	normal	forte	não
7	nublado	frio	normal	forte	sim
8	ensolarado	moderado	alta	fraco	não
9	ensolarado	frio	normal	fraco	sim
10	chuva	moderado	normal	fraco	sim
11	ensolarado	moderado	normal	forte	sim
12	nublado	moderado	alta	forte	sim
13	nublado	quente	normal	fraco	sim
14	chuva	moderado	alta	forte	não

3. Algoritmos de Aprendizado Supervisionado

Critério de Divisão – Exemplo de cálculo do Índice Gini

Cálculo do índice Gini para seleção do atributo mais relevante (nó raiz).

Decisão de jogar tênis considerando as condições climáticas.

DIA	AMBIENTE	TEMPERATURA	HUMIDADE	VENTO	DECISÃO
1	ensolarado	quente	alta	fraco	não
2	ensolarado	quente	alta	forte	não
3	nublado	quente	alta	fraco	sim
4	chuva	moderado	alta	fraco	sim
5	chuva	frio	normal	fraco	sim
6	chuva	frio	normal	forte	não
7	nublado	frio	normal	forte	sim
8	ensolarado	moderado	alta	fraco	não
9	ensolarado	frio	normal	fraco	sim
10	chuva	moderado	normal	fraco	sim
11	ensolarado	moderado	normal	forte	sim
12	nublado	moderado	alta	forte	sim
13	nublado	quente	normal	fraco	sim
14	chuva	moderado	alta	forte	não

Gini do atributo “AMBIENTE” é dado pela soma do índice para cada um de seus valores.

AMBIENTE (Nominal)

VALOR	SIM	NÃO	INSTÂNCIAS
ensolarado	2	3	5
nublado	4	0	4
chuva	3	2	5

$$Gini(X) = 1 - \sum_{k=1}^K p_k^2$$

3. Algoritmos de Aprendizado Supervisionado

Critério de Divisão – Exemplo de cálculo do Índice Gini

Cálculo do índice Gini para seleção do atributo mais relevante (nó raiz).

Decisão de jogar tênis considerando as condições climáticas.

DIA	AMBIENTE	TEMPERATURA	HUMIDADE	VENTO	DECISÃO
1	ensolarado	quente	alta	fraco	não
2	ensolarado	quente	alta	forte	não
3	nublado	quente	alta	fraco	sim
4	chuva	moderado	alta	fraco	sim
5	chuva	frio	normal	fraco	sim
6	chuva	frio	normal	forte	não
7	nublado	frio	normal	forte	sim
8	ensolarado	moderado	alta	fraco	não
9	ensolarado	frio	normal	fraco	sim
10	chuva	moderado	normal	fraco	sim
11	ensolarado	moderado	normal	forte	sim
12	nublado	moderado	alta	forte	sim
13	nublado	quente	normal	fraco	sim
14	chuva	moderado	alta	forte	não

Gini do atributo “AMBIENTE” é dado pela soma do índice para cada um de seus valores.

AMBIENTE (Nominal)

VALOR	SIM	NÃO	INSTÂNCIAS
ensolarado	2	3	5
nublado	4	0	4
chuva	3	2	5

$$Gini(X) = 1 - \sum_{k=1}^K p_k^2$$

$$Gini(ensolarado) = 1 - (2/5)^2 + (3/5)^2 = 1 - (0,16 + 0,36) = 0,48$$

3. Algoritmos de Aprendizado Supervisionado

Critério de Divisão – Exemplo de cálculo do Índice Gini

Cálculo do índice Gini para seleção do atributo mais relevante (nó raiz).

Decisão de jogar tênis considerando as condições climáticas.

DIA	AMBIENTE	TEMPERATURA	HUMIDADE	VENTO	DECISÃO
1	ensolarado	quente	alta	fraco	não
2	ensolarado	quente	alta	forte	não
3	nublado	quente	alta	fraco	sim
4	chuva	moderado	alta	fraco	sim
5	chuva	frio	normal	fraco	sim
6	chuva	frio	normal	forte	não
7	nublado	frio	normal	forte	sim
8	ensolarado	moderado	alta	fraco	não
9	ensolarado	frio	normal	fraco	sim
10	chuva	moderado	normal	fraco	sim
11	ensolarado	moderado	normal	forte	sim
12	nublado	moderado	alta	forte	sim
13	nublado	quente	normal	fraco	sim
14	chuva	moderado	alta	forte	não

Gini do atributo “AMBIENTE” é dado pela soma do índice para cada um de seus valores.

AMBIENTE (Nominal)

VALOR	SIM	NÃO	INSTÂNCIAS
ensolarado	2	3	5
nublado	4	0	4
chuva	3	2	5

$$Gini(X) = 1 - \sum_{k=1}^K p_k^2$$

$$Gini(ensolarado) = 1 - (2/5)^2 + (3/5)^2 = 1 - (0,16 + 0,36) = 0,48$$

$$Gini(nublado) = 1 - (4/4)^2 + (0/4)^2 = 0$$

$$Gini(chuva) = 1 - (3/5)^2 + (2/5)^2 = 1 - (0,36 + 0,16) = 0,48$$

$$\begin{aligned} Gini(AMBIENTE) &= \{(5/14) \times 0,48\} + \{(4/14) \times 0\} + \{(5/14) \times 0,48\} \\ &= 0,171 + 0 + 0,171 = \mathbf{0,342} \end{aligned}$$

3. Algoritmos de Aprendizado Supervisionado

Critério de Divisão – Exemplo de cálculo do Índice Gini

Cálculo do índice Gini para seleção do atributo mais relevante (nó raiz).

Decisão de jogar tênis considerando as condições climáticas.

DIA	AMBIENTE	TEMPERATURA	HUMIDADE	VENTO	DECISÃO
1	ensolarado	quente	alta	fraco	não
2	ensolarado	quente	alta	forte	não
3	nublado	quente	alta	fraco	sim
4	chuva	moderado	alta	fraco	sim
5	chuva	frio	normal	fraco	sim
6	chuva	frio	normal	forte	não
7	nublado	frio	normal	forte	sim
8	ensolarado	moderado	alta	fraco	não
9	ensolarado	frio	normal	fraco	sim
10	chuva	moderado	normal	fraco	sim
11	ensolarado	moderado	normal	forte	sim
12	nublado	moderado	alta	forte	sim
13	nublado	quente	normal	fraco	sim
14	chuva	moderado	alta	forte	não

É calculado o Índice Gini de cada atributo, sendo **selecionado aquele que apresentar melhor qualidade** (valor mínimo).

ATRIBUTO	ÍNDICE GINI
Ambiente	0,342
Temperatura	0,439
Humidade	0,367
Vento	0,428

3. Algoritmos de Aprendizado Supervisionado

DECISION TREE

Síntese das características dos algoritmos de Árvore de Decisão mais conhecidos:

	ID3	C4.5	C5.0	CART
Tipo de Dados	categórico	categórico, contínuo	categórico, contínuo, timestamps	contínuo, nominal
Velocidade	baixa	maior que ID3	muito alta	média
Poda	não	<i>Pre-Pruning</i>	<i>Pre-Pruning</i>	<i>Post-Pruning</i>
Boosting	não	não	suportado	suportado
Valores Ausentes	não lida	não lida	lida com	lida com
Split Criteria	Entropia, <i>Information Gain</i>	<i>Split Info</i> , <i>Gain Ratio</i>	igual C4.5	Índice Gini

Fonte: Sharma, Himani, and Sunil Kumar. "A survey on decision tree algorithms of classification in data mining." International Journal of Science and Research (IJSR) 5.4 (2016): 2094–2097.

3. Algoritmos de Aprendizado Supervisionado

DECISION TREE

Síntese das características dos algoritmos de Árvore de Decisão mais conhecidos:

	ID3	C4.5	C5.0	CART
Tipo de Dados	categórico	categórico, contínuo	categórico, contínuo, timestamps	contínuo, nominal
Velocidade	baixa	maior que ID3	muito alta	média
Poda	não	<i>Pre-Pruning</i>	<i>Pre-Pruning</i>	<i>Post-Pruning</i>
Boosting	não	não	suportado	suportado
Valores Ausentes	não lida	não lida	lida com	lida com
Split Criteria	Entropia, <i>Information Gain</i>	<i>Split Info</i> , <i>Gain Ratio</i>	igual C4.5	Índice Gini

Pre-Pruning: poda da árvore visando parar seu crescimento (durante execução).

Post-Pruning: poda da árvore completa (ao final da execução).

Fonte: Sharma, Himani, and Sunil Kumar. "A survey on decision tree algorithms of classification in data mining." International Journal of Science and Research (IJSR) 5.4 (2016): 2094–2097.

VÍDEOS { **Visual Guide to Decision Trees.**
Disponível: <https://www.youtube.com/watch?v=zs6yHVtxyv8>, Acessado: 22/02/2023
Decision Tree CART - Machine Learning Fun and Easy
Disponível: <https://www.youtube.com/watch?v=dJhGvXc8sHE>, Acessado: 22/02/2023

3. Algoritmos de Aprendizado Supervisionado

ENSEMBLE LEARNING

Paradigma de aprendizado de máquina que **combina modelos individuais** para melhorar o desempenho geral do modelo e aumentar sua generalização. Usa técnicas de votação, peso e combinação de previsões, **obtendo previsões por vezes melhores que as de métodos tradicionais**.

3. Algoritmos de Aprendizado Supervisionado

ENSEMBLE LEARNING

Paradigma de aprendizado de máquina que combina modelos individuais para melhorar o desempenho geral do modelo e aumentar sua generalização. Usa técnicas de votação, peso e combinação de previsões, obtendo previsões por vezes melhores que as de métodos tradicionais.

TIPOS DE MÉTODOS ENSEMBLE

Bagging (Bootstrapped Aggregating)

- constrói vários **modelos independentes** e calcula a **média de suas previsões** para um resultado geral.
- contribui para uma **menor variância** do que o modelo tradicional.
- ex. *Random Forest* e variantes *Bootstrapped* (ex. Naive Bayes, *Linear Regression*, KNN).

3. Algoritmos de Aprendizado Supervisionado

ENSEMBLE LEARNING

Paradigma de aprendizado de máquina que combina modelos individuais para melhorar o desempenho geral do modelo e aumentar sua generalização. Usa técnicas de votação, peso e combinação de previsões, obtendo previsões por vezes melhores que as de métodos tradicionais.

TIPOS DE MÉTODOS ENSEMBLE

Bagging (Bootstrapped Aggregating)

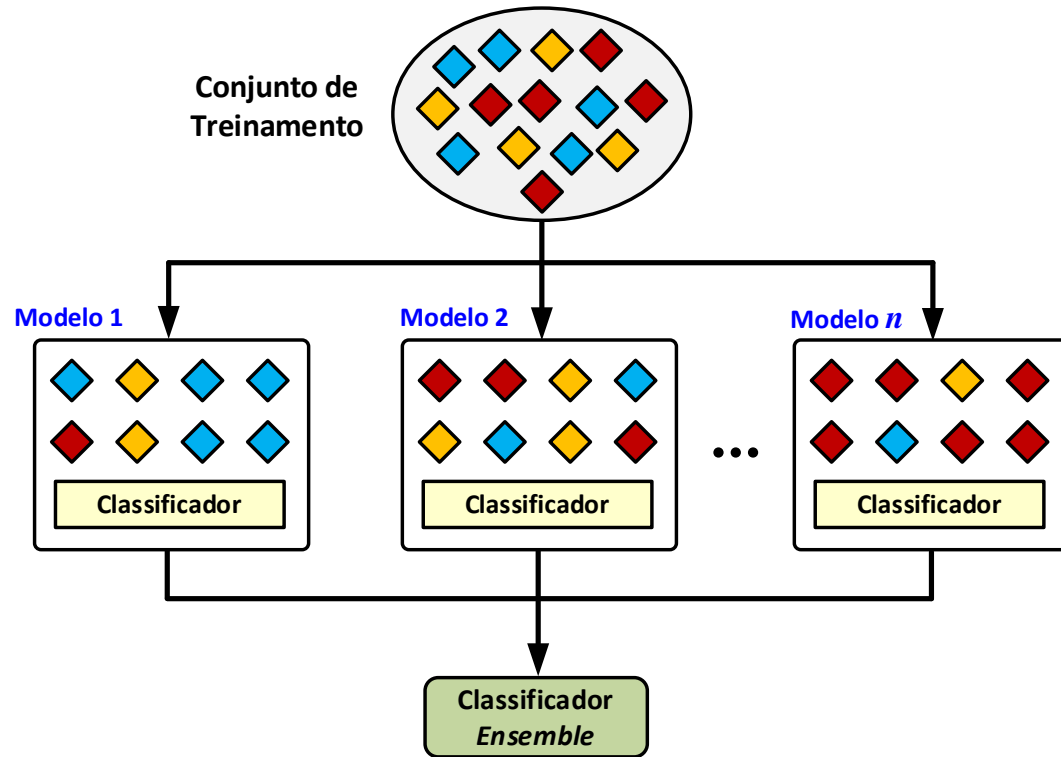
- constrói vários modelos independentes e calcula a média de suas previsões para um resultado geral.
- contribui para uma menor variância do que o modelo tradicional.
- ex. *Random Forest* e variantes *Bootstrapped* (ex. Naive Bayes, *Linear Regression*, KNN).

Boosting

- **constrói um modelo básico sequencialmente.** Atribui pesos aos dados para aumentar a probabilidade dos que foram classificados incorretamente serem escolhidos para o próximo subconjunto de treinamento.
- ex. Adaboost, GBDT (*Gradiente Boosting Decision*) e XGboost.

3. Algoritmos de Aprendizado Supervisionado

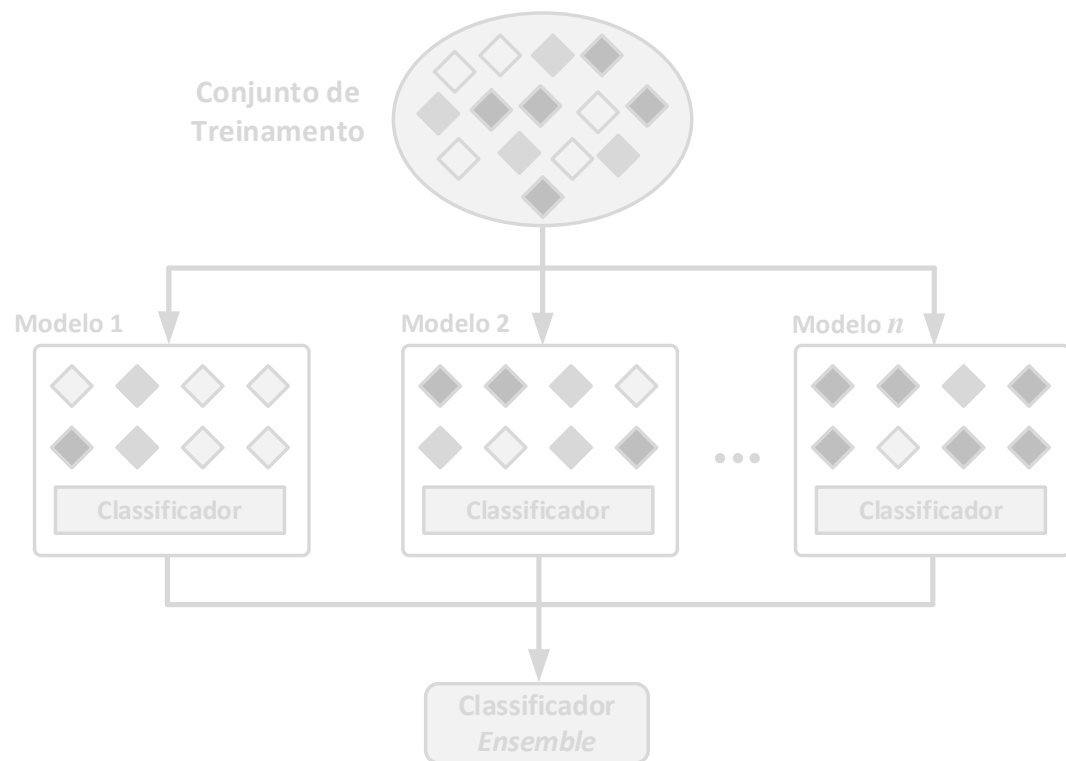
ENSEMBLE LEARNING



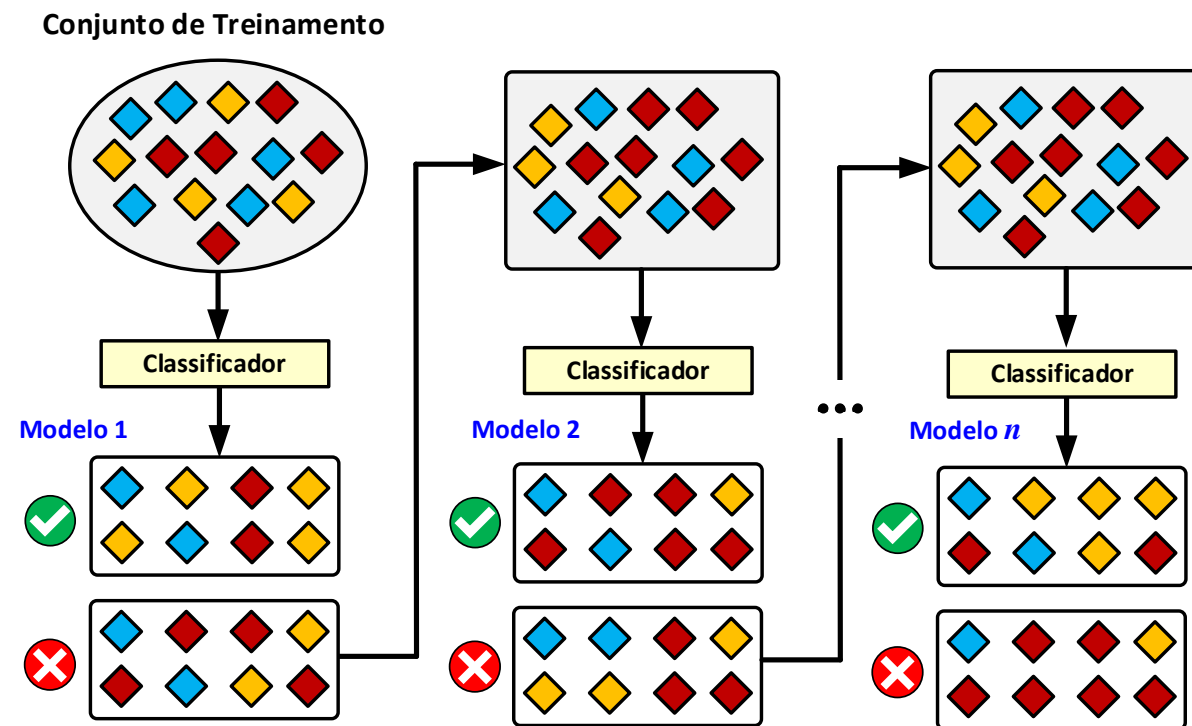
(a) Bagging

3. Algoritmos de Aprendizado Supervisionado

ENSEMBLE LEARNING



(a) Bagging

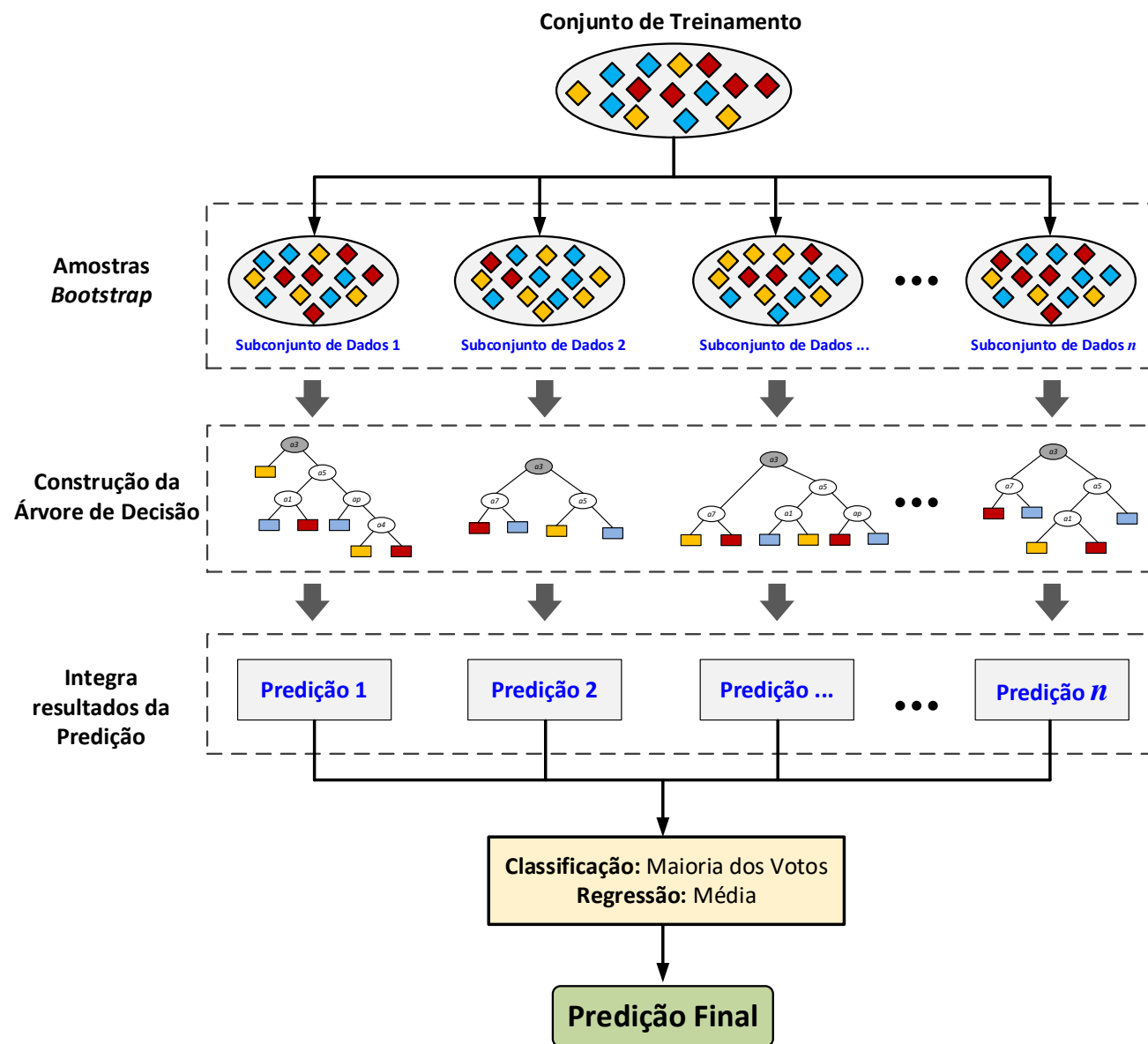


(b) Boosting

3. Algoritmos de Aprendizado Supervisionado

RANDOM FOREST

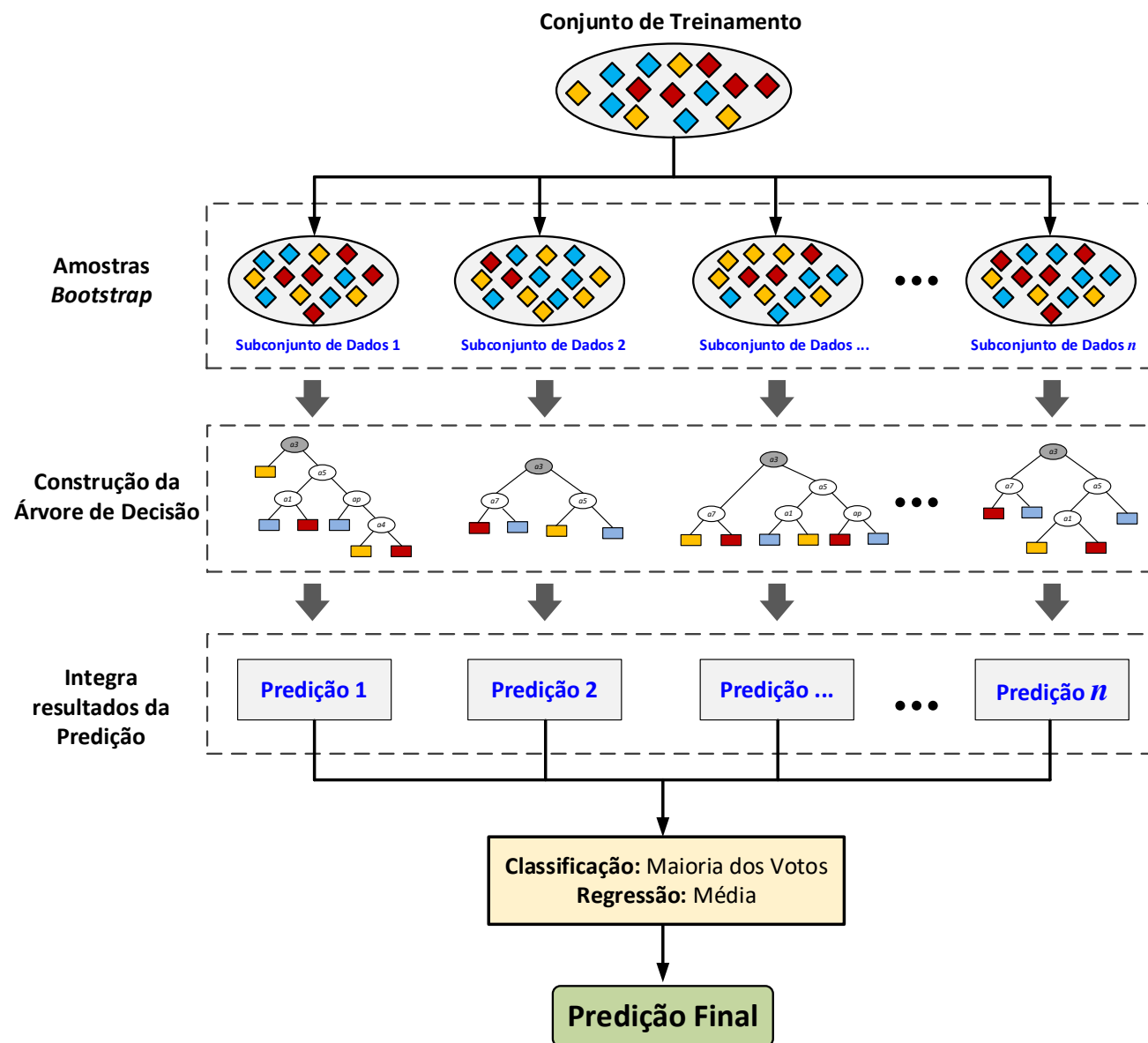
- Baseado na *Random Tree* (classificador e regressor que seleciona aleatoriamente k atributos para construir a árvore de decisão).
- A *Random Forest* **constrói várias árvores de decisão** a partir de uma amostra aleatória de atributos de dados de treinamento.



3. Algoritmos de Aprendizado Supervisionado

RANDOM FOREST

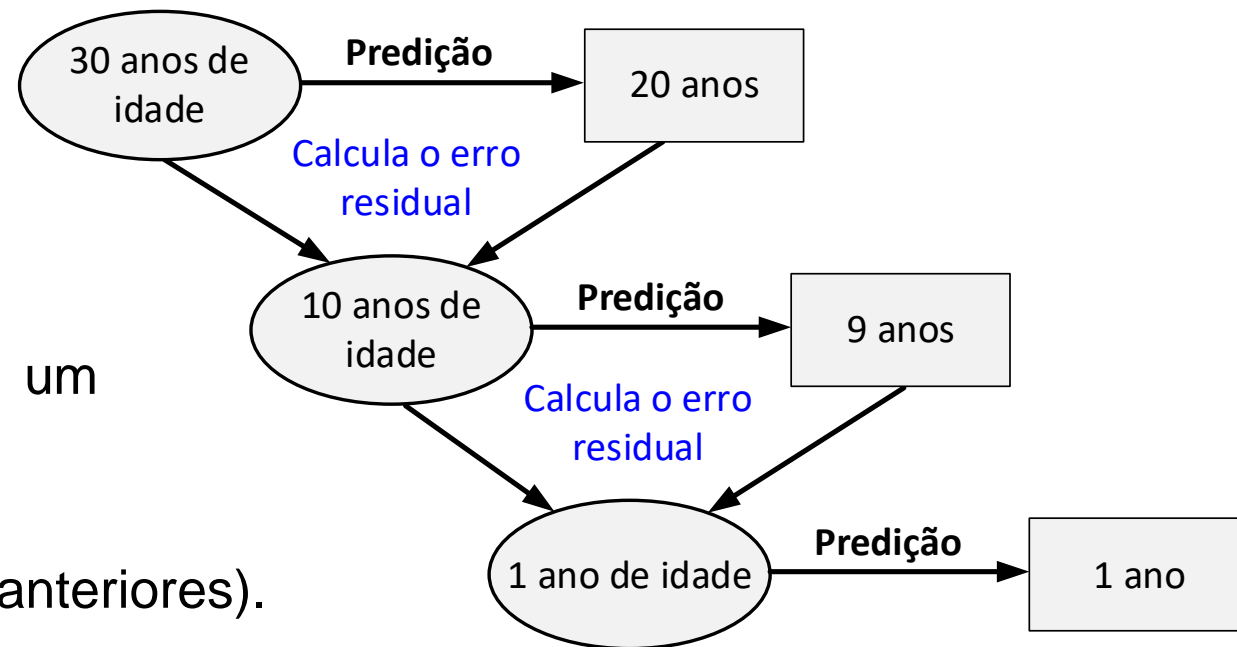
- Baseado na *Random Tree* (classificador e regressor que seleciona aleatoriamente k atributos para construir a árvore de decisão).
- A *Random Forest* constrói várias árvores de decisão a partir de uma amostra aleatória de atributos de dados de treinamento.
- Em seguida toma a média ou votação do **resultado de cada árvore para prever** a saída de um **novo exemplo**.
- **Reduz a variância e o overfitting** dos modelos tradicionais.



3. Algoritmos de Aprendizado Supervisionado

GBDT (Gradient Boosting Decision Tree)

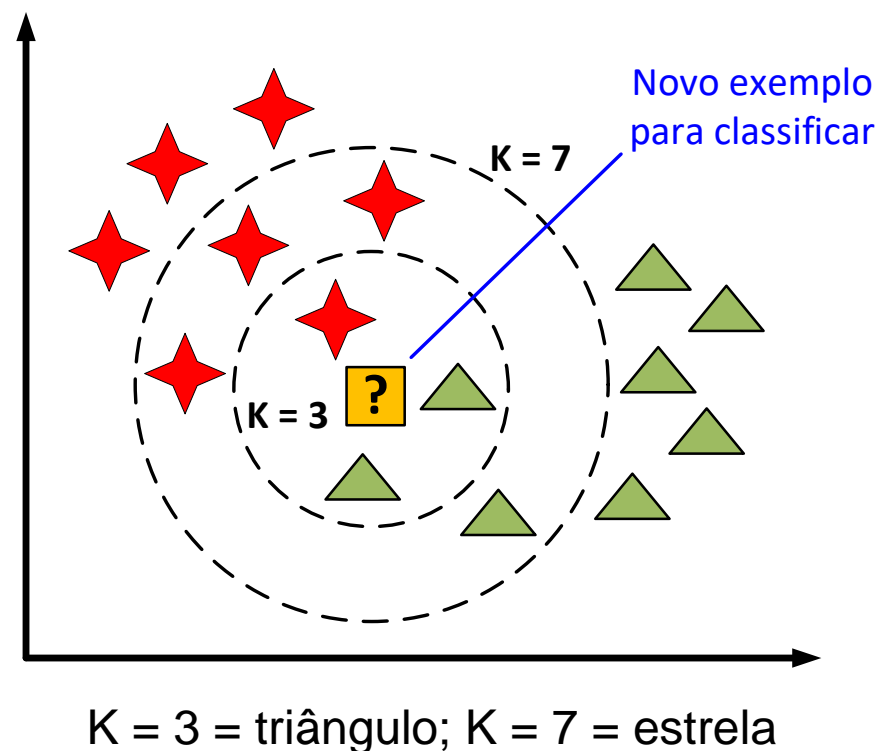
- Método tipo *Boosting*, constrói continuamente novas árvores de decisão para **aprender os resíduos de árvores anteriores** (erro entre o valor previsto e o verdadeiro).
- O valor previsto corresponde a soma dos resultados de todas as árvores de decisão.
- Exemplo: a cada passo é **calculado o erro residual entre a idade real e a prevista**. São geradas novas árvores para aprender o erro da anterior até a convergência do método.
- Corrige o desvio da árvore de decisão **obtendo um certo grau de *underfitting***.
- **Treino não paralelizável** (depende de resultados anteriores).



3. Algoritmos de Aprendizado Supervisionado

KNN (K-Nearest Neighbors)

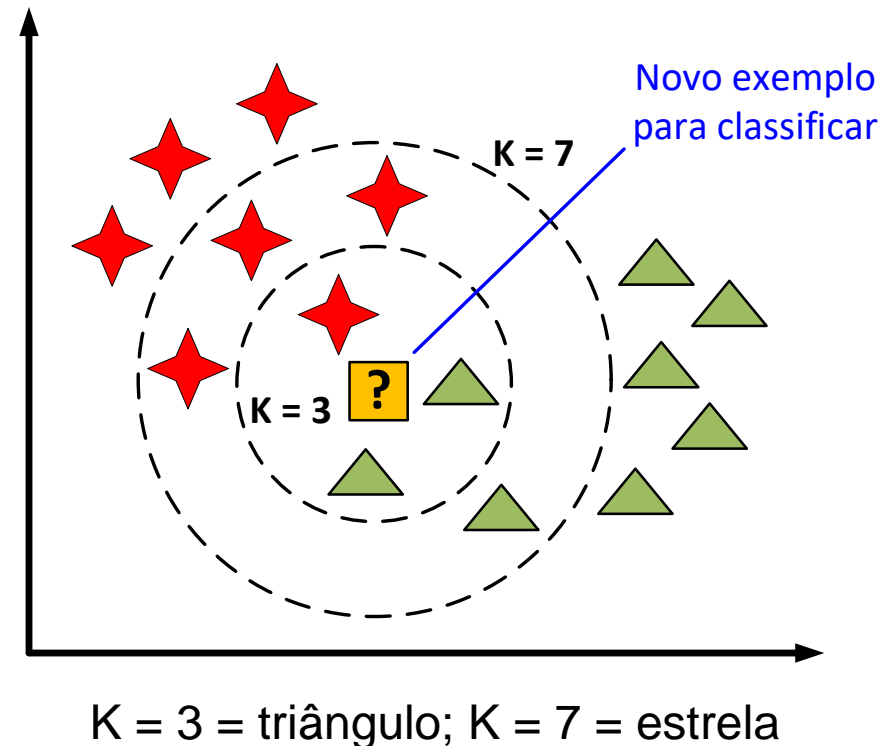
- É um método supervisionado não-paramétrico, usado em tarefas de classificação e regressão. Tende a ter melhor desempenho em conjuntos de dados com limites de decisão irregulares.
- Princípio:** Se a **maioria** dos K vizinhos mais próximos de uma amostra desconhecida pertencem a uma categoria, então a amostra desconhecida também pertence.



3. Algoritmos de Aprendizado Supervisionado

KNN (K-Nearest Neighbors)

- É um método supervisionado não-paramétrico, usado em tarefas de classificação e regressão. Tende a ter melhor desempenho em conjuntos de dados com limites de decisão irregulares.
- Princípio:** Se a maioria dos K vizinhos mais próximos de uma amostra desconhecida pertencem a uma categoria, então a amostra desconhecida também pertence.
- Funcionamento:**
 - Definir o número K de vizinhos mais próximos.
 - Calcular a distância do ponto desconhecido para todos os demais pontos na base de dados de treinamento.
 - Encontre os K pontos mais próximos.
 - Atribuir a classe mais frequente ao objeto desconhecido.



3. Algoritmos de Aprendizado Supervisionado

NÄIVE BAYES

- Algoritmo de classificação baseado no teorema de Bayes. Considera que todos os atributos são independentes entre si, suposição conhecida como *näive* (ingênua).
- **Princípio:** O método divide o problema de classificação em subproblemas mais simples. Calcula a probabilidade condicional de cada atributo para cada classe, combinando-as para obter a probabilidade geral de um objeto pertencer a uma classe (*a posteriori*).

3. Algoritmos de Aprendizado Supervisionado

NÄIVE BAYES

- Algoritmo de classificação baseado no teorema de Bayes. Considera que todos os atributos são independentes entre si, suposição conhecida como *näive* (ingênuo).
- **Princípio:** O método divide o problema de classificação em subproblemas mais simples. Calcula a probabilidade condicional de cada atributo para cada classe, combinando-as para obter a probabilidade geral de um objeto pertencer a uma classe (*a posteriori*).
- **Teorema de Bayes:** Permite calcular a probabilidade a posteriori $P(H|X)$.

X = evidência (tupla de dados)

H = hipótese (ex. tupla de dados X pertencer a classe C)

probabilidade de H ocorrer dado
que X ocorreu (*a posteriori*) ←
$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

3. Algoritmos de Aprendizado Supervisionado

NÄIVE BAYES

- Algoritmo de classificação baseado no teorema de Bayes. Considera que todos os atributos são independentes entre si, suposição conhecida como *näive* (ingênua).
- Princípio:** O método divide o problema de classificação em subproblemas mais simples. Calcula a probabilidade condicional de cada atributo para cada classe, combinando-as para obter a probabilidade geral de um objeto pertencer a uma classe (*a posteriori*).
- Teorema de Bayes:** Permite calcular a probabilidade a posteriori $P(H|X)$.

X = evidência (tupla de dados)

H = hipótese (ex. tupla de dados X pertencer a classe C)

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Diagram illustrating the components of Bayes' Theorem:

- $P(X|H)$: Probabilidade condicional da evidência X dada a hipótese H (indicated by a blue arrow from the text above).
- $P(H)$: probabilidade de H ocorrer (*a priori*) (indicated by a blue arrow from the text above).
- $P(X)$: probabilidade de X ocorrer (indicated by a blue arrow from the text below).
- $P(H|X)$: probabilidade de H ocorrer dado que X ocorreu (*a posteriori*) (indicated by a grey arrow from the text to the left).

3. Algoritmos de Aprendizado Supervisionado

NÄIVE BAYES

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Sobre as Probabilidades do Teorema de Bayes:

- $P(H|X)$ = probabilidade *a posteriori*. **Quão provável é nossa hipótese dada a evidência observada?**
(este é o valor que será calculado)
- $P(X|H)$ = Probabilidade. **Quão provável é a evidência dada que nossa hipótese é verdadeira?**
- $P(H)$ = Probabilidade *a priori*. **Quão provável era nossa hipótese antes de observar a evidência?**
- $P(X)$ = Probabilidade Marginal. **Quão provável é a nova evidência ser verdadeira?**

3. Algoritmos de Aprendizado Supervisionado

NÄIVE BAYES

Sobre as Probabilidades do Teorema de Bayes:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

- $P(H|X)$ = probabilidade *a posteriori*. Quão provável é nossa hipótese dada a evidência observada? (este é o valor que será calculado)
- $P(X|H)$ = Probabilidade. Quão provável é a evidência dada que nossa hipótese é verdadeira?
- $P(H)$ = Probabilidade *a priori*. Quão provável era nossa hipótese antes de observar a evidência?
- $P(X)$ = Probabilidade Marginal. Quão provável é a nova evidência ser verdadeira?

Näive Bayes: Supondo um vetor de atributos $X = (X_1, X_2, \dots, X_n)$, combinando a suposição de independência de cada atributo, a probabilidade *a posteriori* é obtida por:

$$P(X|H) = \prod_{i=1}^n P(X_i|H)$$

3. Algoritmos de Aprendizado Supervisionado

NÄIVE BAYES – Exemplo

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Aplicação no problema de decisão de jogar tênis considerando as condições climáticas.

DIA	AMBIENTE	DECISÃO
1	ensolarado	não
2	ensolarado	não
3	nublado	sim
4	chuva	sim
5	chuva	sim
6	chuva	não
7	nublado	sim
8	ensolarado	não
9	ensolarado	sim
10	chuva	sim
11	ensolarado	sim
12	nublado	sim
13	nublado	sim
14	chuva	não

AMBIENTE (Nominal)

VALOR	SIM	NÃO	
ensolarado	2	3	= 5/14 = 0,36
nublado	4	0	= 4/14 = 0,28
chuva	3	2	= 5/14 = 0,36
	= 9/14 = 0,64	= 5/14 = 0,36	

3. Algoritmos de Aprendizado Supervisionado

NÄIVE BAYES – Exemplo

Aplicação no problema de decisão de jogar tênis considerando as condições climáticas.

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

DIA	AMBIENTE	DECISÃO
1	ensolarado	não
2	ensolarado	não
3	nublado	sim
4	chuva	sim
5	chuva	sim
6	chuva	não
7	nublado	sim
8	ensolarado	não
9	ensolarado	sim
10	chuva	sim
11	ensolarado	sim
12	nublado	sim
13	nublado	sim
14	chuva	não

AMBIENTE (Nominal)

VALOR	SIM	NÃO	
ensolarado	2	3	= 5/14 = 0,36
nublado	4	0	= 4/14 = 0,28
chuva	3	2	= 5/14 = 0,36
	= 9/14 = 0,64	= 5/14 = 0,36	

$P(\text{ensolarado} | \text{sim}) = 2/9 = 0,22$

$P(\text{ensolarado}) = 5/14 = 0,36$

$P(\text{sim}) = 0,64$

$$P(\text{sim} | \text{ensolarado}) = \frac{P(\text{ensolarado} | \text{sim}) \times P(\text{sim})}{P(\text{ensolarado})} = \frac{0,22 \times 0,64}{0,36} = 0,39$$

3. Algoritmos de Aprendizado Supervisionado

NÄIVE BAYES – Exemplo

Aplicação no problema de decisão de jogar tênis considerando as condições climáticas.

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

DIA	AMBIENTE	DECISÃO
1	ensolarado	não
2	ensolarado	não
3	nublado	sim
4	chuva	sim
5	chuva	sim
6	chuva	não
7	nublado	sim
8	ensolarado	não
9	ensolarado	sim
10	chuva	sim
11	ensolarado	sim
12	nublado	sim
13	nublado	sim
14	chuva	não

AMBIENTE (Nominal)		
VALOR	SIM	NÃO
ensolarado	2	3
nublado	4	0
chuva	3	2
	= 9/14 = 0,64	= 5/14 = 0,36

$P(\text{ensolarado} | \text{sim}) = 2/9 = 0,22$

$P(\text{ensolarado} | \text{não}) = 3/5 = 0,60$

$P(\text{ensolarado}) = 5/14 = 0,36$

$P(\text{não}) = 0,36$

$P(\text{sim}) = 0,64$

$$P(\text{sim} | \text{ensolarado}) = \frac{P(\text{ensolarado} | \text{sim}) \times P(\text{sim})}{P(\text{ensolarado})} = \frac{0,22 \times 0,64}{0,36} = 0,39$$

$$P(\text{não} | \text{ensolarado}) = \frac{P(\text{ensolarado} | \text{não}) \times P(\text{não})}{P(\text{ensolarado})} = \frac{0,60 \times 0,36}{0,36} = 0,60$$

Maior probabilidade

3. Algoritmos de Aprendizado Supervisionado

NÄIVE BAYES – Algoritmos

Cada **variante** **Näive Bayes** utiliza uma distribuição diferente para modelar a probabilidade condicional dos atributos para cada classe. O desempenho do modelo depende do grau em que a hipótese de independência de atributos é satisfeita.

Naive Bayes Gaussiano: adequado para conjuntos de **dados contínuos**, assume que os **valores dos atributos** seguem uma **distribuição normal**. Usa a média e o desvio padrão dos atributos em cada classe para calcular a probabilidade condicional de uma determinada instância pertencer a cada classe.

3. Algoritmos de Aprendizado Supervisionado

NÄIVE BAYES – Algoritmos

Cada **variante** **Näive Bayes** utiliza uma **distribuição diferente para modelar a probabilidade** condicional dos atributos para cada classe. O desempenho do modelo depende do grau em que a hipótese de independência de atributos é satisfeita.

Naive Bayes Gaussiano: adequado para conjuntos de dados contínuos, assume que os valores dos atributos seguem uma distribuição normal. Usa a média e o desvio padrão dos atributos em cada classe para calcular a probabilidade condicional de uma determinada instância pertencer a cada classe.

Naive Bayes Multinomial: usado para **dados discretos** (ex. contagem de palavras em um documento). Modela a **frequência com que cada valor de atributo** ocorre no conjunto de dados e a utiliza para calcular a provável classe de um objeto.

3. Algoritmos de Aprendizado Supervisionado

NÄIVE BAYES – Algoritmos

Cada **variante** **Näive Bayes** utiliza uma **distribuição diferente para modelar a probabilidade** condicional dos atributos para cada classe. O desempenho do modelo depende do grau em que a hipótese de independência de atributos é satisfeita.

Naive Bayes Gaussiano: adequado para conjuntos de dados contínuos, assume que os valores dos atributos seguem uma distribuição normal. Usa a média e o desvio padrão dos atributos em cada classe para calcular a probabilidade condicional de uma determinada instância pertencer a cada classe.

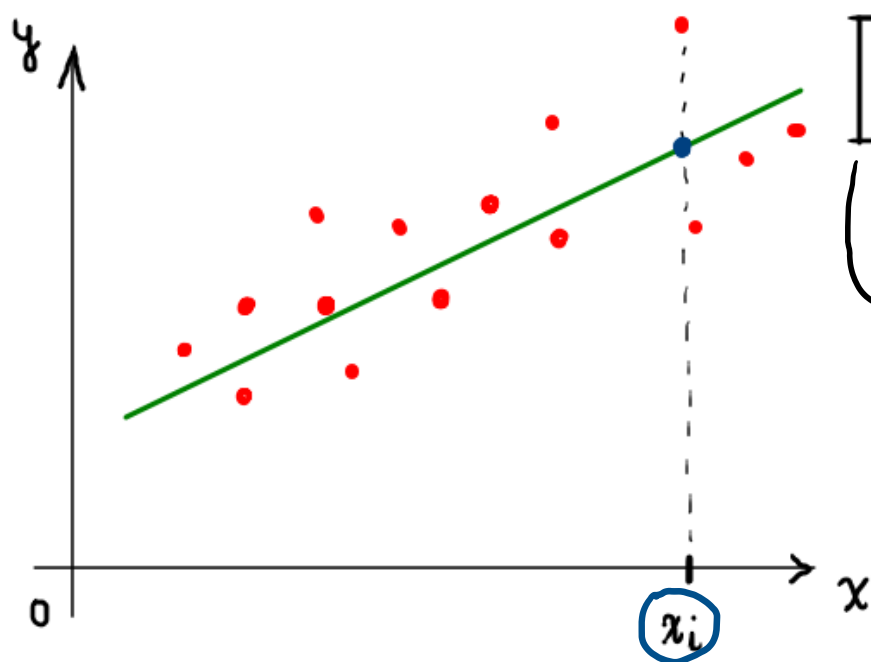
Naive Bayes Multinomial: usado para dados discretos (ex. contagem de palavras em um documento). Modela a frequência com que cada valor de atributo ocorre no conjunto de dados e a utiliza para calcular a provável classe de um objeto.

Naive Bayes Bernoulli: é semelhante ao Multinomial, mas é usado quando os objetos consistem de **atributos binários** (ex. presença ou ausência de um termo em um documento). A **probabilidade** de cada classe **é dada pela ocorrência ou ausência de cada atributo** e não por sua frequência.

3. Algoritmos de Aprendizado Supervisionado

REGRESSÃO LINEAR

- **Quando usar:** estimativa de grandezas.
- **Conjunto de dados:** pares ordenados $(x_i, y_i), i = 1, 2, \dots, m$.



→ modelo: $\phi(x) = \theta_1 + \theta_2 x$

→ erro:

$$e_i = [\phi(x_i) - y_i]^2$$

→ $\theta_1 = ?$, $\theta_2 = ?$

→ minimizar a soma de e_i 's.

3. Algoritmos de Aprendizado Supervisionado

REGRESSÃO LINEAR – Modelo

→ θ_1 e θ_2 são desconhecidos

- Reta: $\phi(x) = \theta_1 + \theta_2 x$

- Erro para cada ponto: $e_i = (y_i - \phi(x_i))^2$

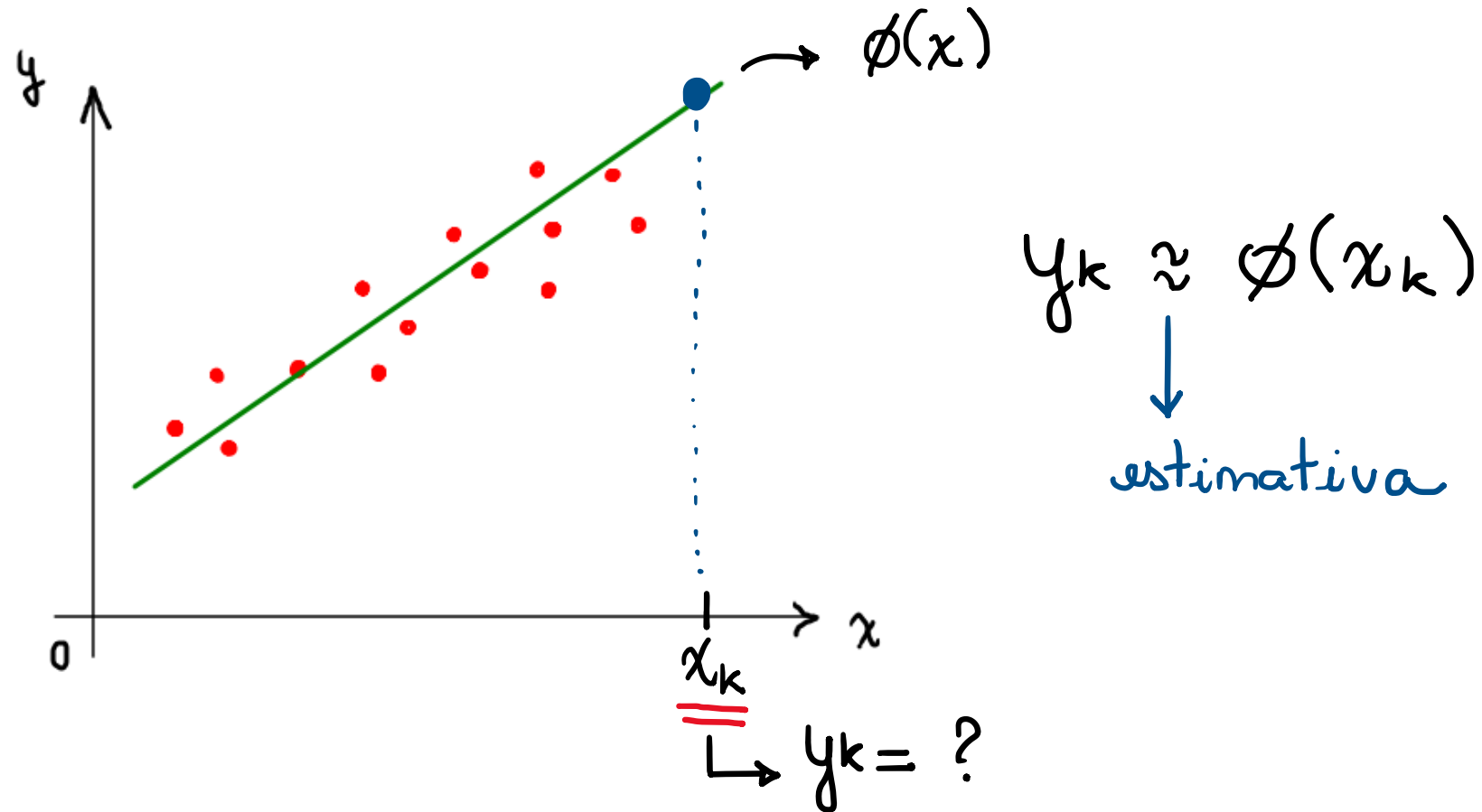
- Problema: minimizar $F(\theta_1, \theta_2) = \sum_{i=1}^m \underbrace{(y_i - \phi(x_i))^2}_{e_i}$ $m \gg 2$

↪ $F(\theta_1, \theta_2) \geq 0$

3. Algoritmos de Aprendizado Supervisionado

REGRESSÃO LINEAR – Modelo

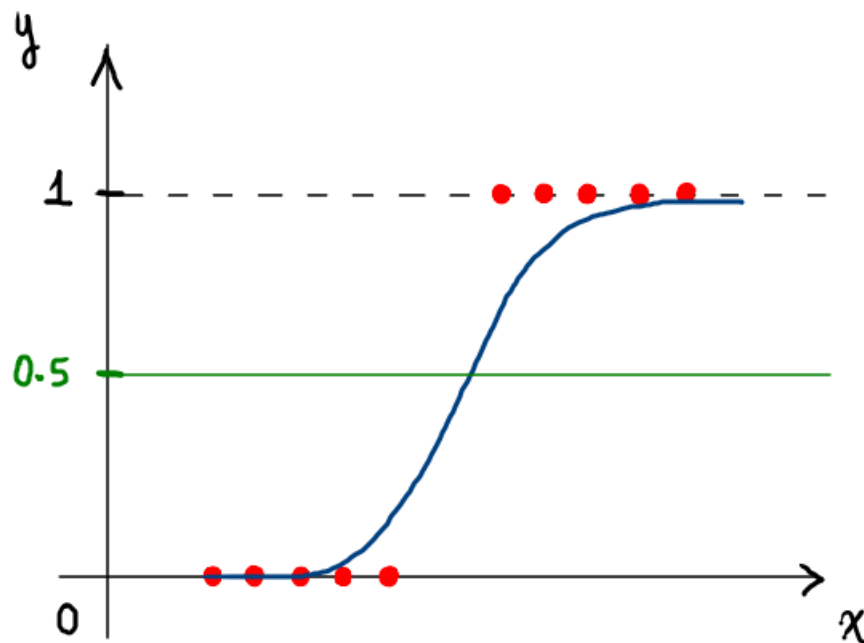
Usando o modelo para estimativas.



3. Algoritmos de Aprendizado Supervisionado

REGRESSÃO LOGÍSTICA

Quando usar: os dados são classificados em duas categorias.



💡 para cada x_i $\rightarrow y_i \in \{0, 1\}$.

$\hookrightarrow \phi(x)$: sigmóide

3. Algoritmos de Aprendizado Supervisionado

REGRESSÃO LOGÍSTICA

Etapa 1: busca dos parâmetros.

- Modelo:**

→ $\phi(t) = \frac{1}{1 + e^{-t}}$

- $t = \theta_0 + \theta_1 x_1 + \theta_2 x_2$

- $\phi(t) \in [0, 1]$ (probabilidade)

- Problema:** minimizar $J(\theta)$

$$J(\theta) = C \sum_{i=1}^n -[\underline{y_i} \log(\underbrace{\phi(\theta, x^{(i)})}_{\text{probabilidade}}) - (1 - \underline{y_i}) \log(1 - \underbrace{\phi(\theta, x^{(i)})}_{\text{probabilidade}})]$$

$(\theta_0, \theta_1, \theta_2)$

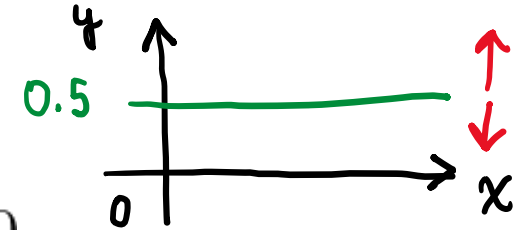
→ os parâmetros são encontrados minimizando $J(\theta)$.

3. Algoritmos de Aprendizado Supervisionado

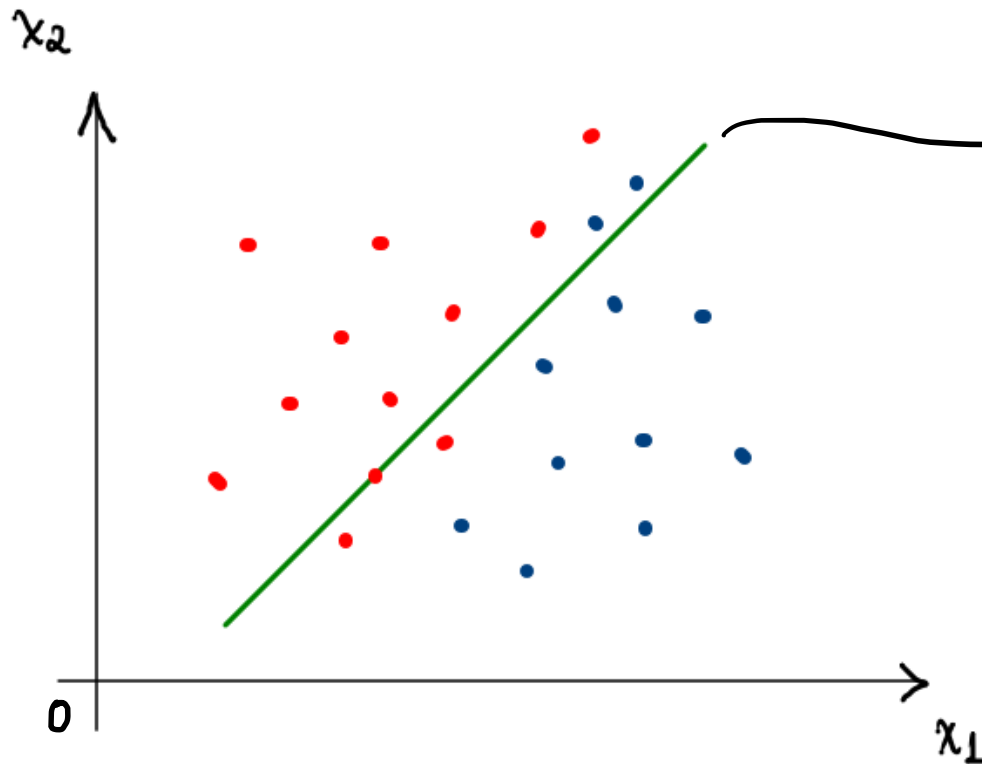
REGRESSÃO LOGÍSTICA

Etapa 2: classificação.

$\phi(t) = 0.5 \Rightarrow \theta_0 + \theta_1 x_1 + \theta_2 x_2 = 0$



reta (fronteira)

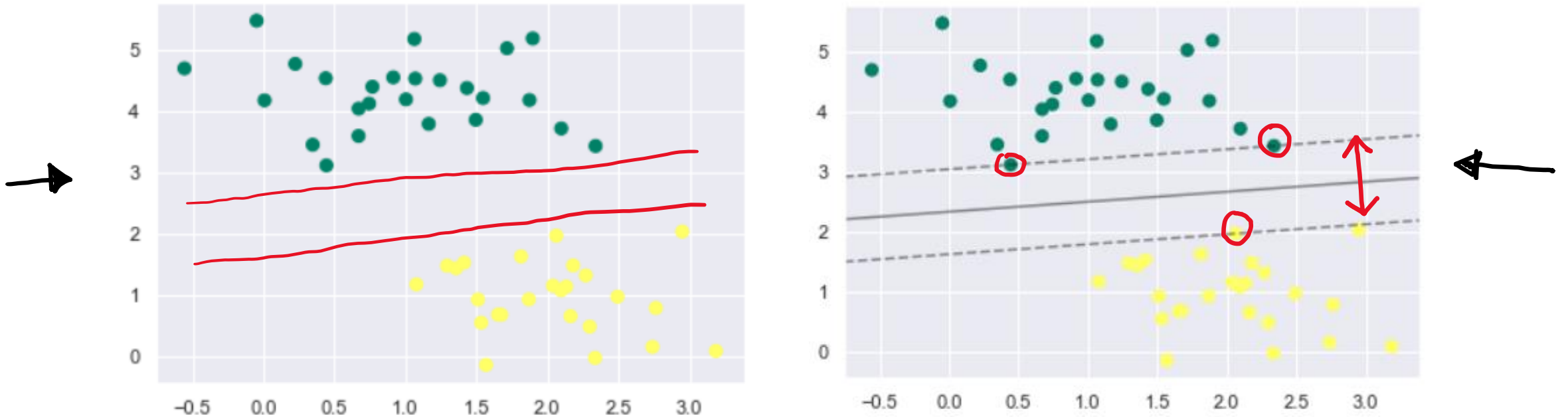


↪ a reta é decisiva para a classificação de novos dados.

3. Algoritmos de Aprendizado Supervisionado

SUPPORT VECTOR MACHINE (SVM)

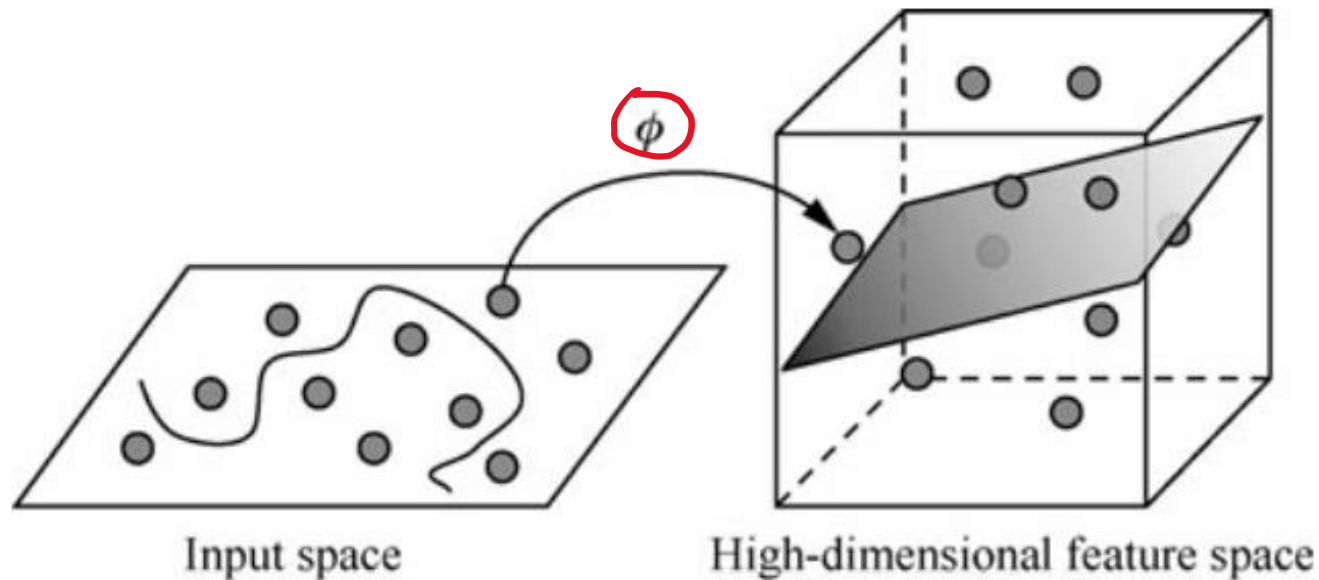
- **Tradução:** Máquina de Vetores de Suporte
- **Ideia:** dados linearmente separáveis.



3. Algoritmos de Aprendizado Supervisionado

SUPPORT VECTOR MACHINE (SVM) – Características

- Se os dados **são** linearmente separáveis: hiperplano ótimo.
- Se os dados **não** são linearmente separáveis: transformação.



Resumo

- Foram revistos os conceitos e as características do aprendizado de máquina Supervisionado para as tarefas de classificação e regressão.
- Foi explicada a importância dos dados como um dos principais componentes do modelo, apresentados os problemas que podem interferir no treinamento e teste do modelo, bem como seus impactos no enviesamento da solução.

Resumo

- Foram revistos os conceitos e as características do aprendizado de máquina Supervisionado para as tarefas de classificação e regressão.
- Foi explicada a importância dos dados como um dos principais componentes do modelo, apresentados os problemas que podem interferir no treinamento e teste do modelo, bem como seus impactos no enviesamento da solução.
- Foram apresentadas técnicas úteis para solucionar problemas nos dados e auxiliar na divisão do conjunto de dados em treinamento e teste, para obtenção de modelos com maior capacidade de generalização.
- Descrito o funcionamento dos algoritmos Supervisionados: *Decision Tree*, *Ensemble Learning* (*Random Forest*, *GBDT*), *K-Nearest Neighbors* (KNN), *Näive Bayes*, *Regressão Linear*, *Regressão Logística*, *Support Vector Machine* (SVM) e apresentados exemplos e aplicação.

Referências Bibliográficas

- [1] Huawei Technologies Co. Artificial Intelligence Technology - Official Textbooks for Huawei ICT Academy. Morgan Kaufmann, 3rd edition, 2023.
- [2] Marc Peter Deisenroth, A Aldo Faisal, and Cheng Soon Ong. Mathematics for Machine Learning. Cambridge University Press, 2020.
- [3] Aurélien Géron. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media, Inc., 3rd edition, 2022.
- [4] Trevor Hastie, Robert Tibshirani, JeromeHFriedman, and JeromeHFriedman. The elements of statistical learning: data mining, inference, and prediction, volume 2. Springer, 2009.
- [5] Han Jiawei, Kamber Micheline, and Pei Jian. Data Mining: Concepts and Techniques. Morgan Kaufmann, 3rd edition, 2012.
- [6] Andreas C Müller and Sarah Guido. Introduction to Machine Learning with Python: a guide for data scientists. O'Reilly Media, Inc., 2016.
- [7] Shai Shalev-Shwartz and Shai Ben-David. Understanding Machine Learning: From theory to algorithms. Cambridge university press, 3rd edition, 2014.

Apoio

Projeto Residência 8 (TPA N° 068/SOFTEX/UNIFEI)

Este projeto é apoiado pelo Ministério da Ciência, Tecnologia e Inovações, com recursos da Lei nº 8.248, de 23 de outubro de 1991, no âmbito do [PPI-Softex| PNM-Design], coordenado pela Softex.





Rafael de Magalhães Dias Frinhani
Universidade Federal de Itajubá
Instituto de Matemática e Computação
frinhani@unifei.edu.br



Rodrigo Silva Lima
Universidade Federal de Itajubá
Instituto de Matemática e Computação
rodlima@unifei.edu.br