

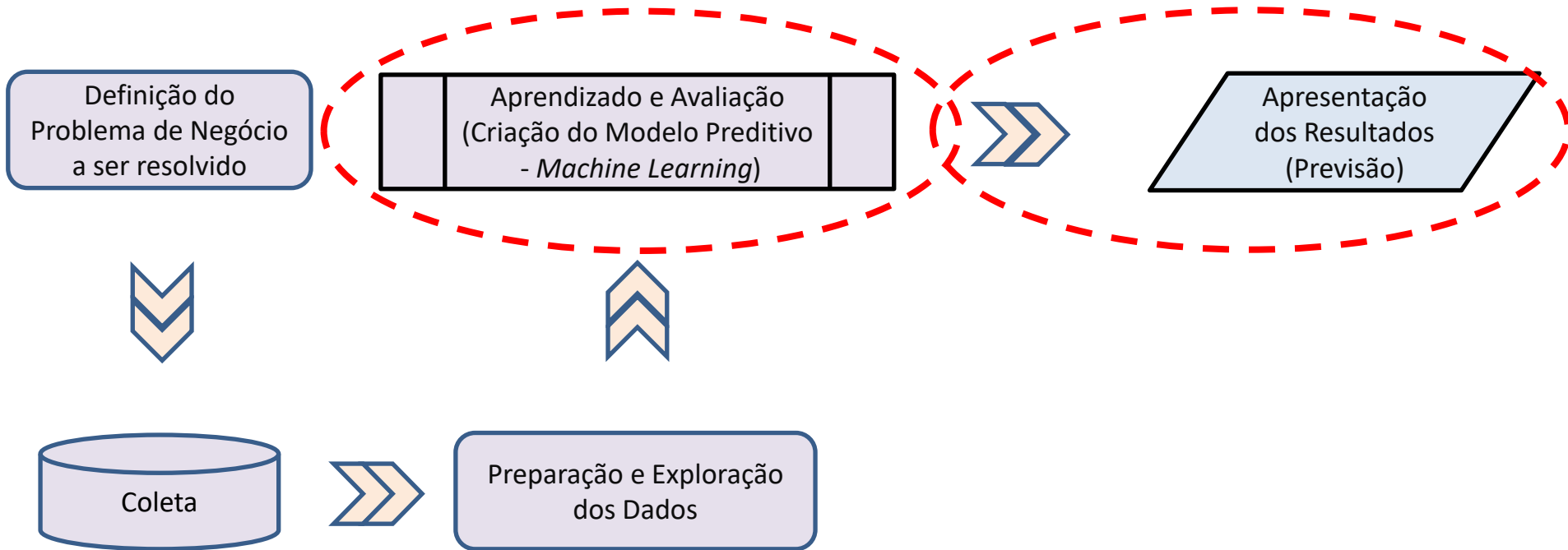
Inteligência Artificial

Aprendizado Supervisionado de Máquina – Classificação:
Outros Algoritmos

Prof. Dr. Ivan Carlos Alcântara de Oliveira

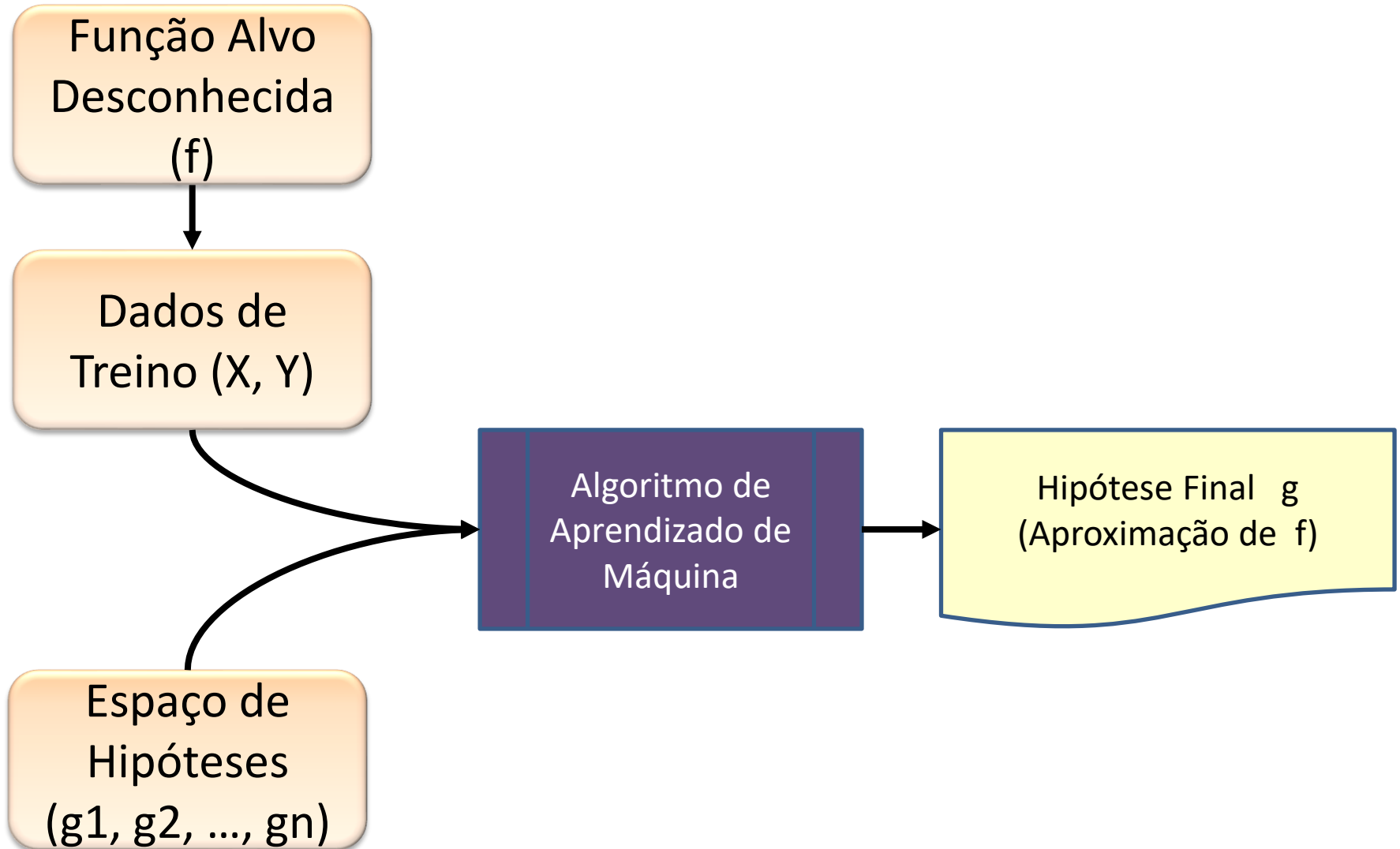
<https://orcid.org/0000-0002-6020-7535>

Processo de Análise de Dados



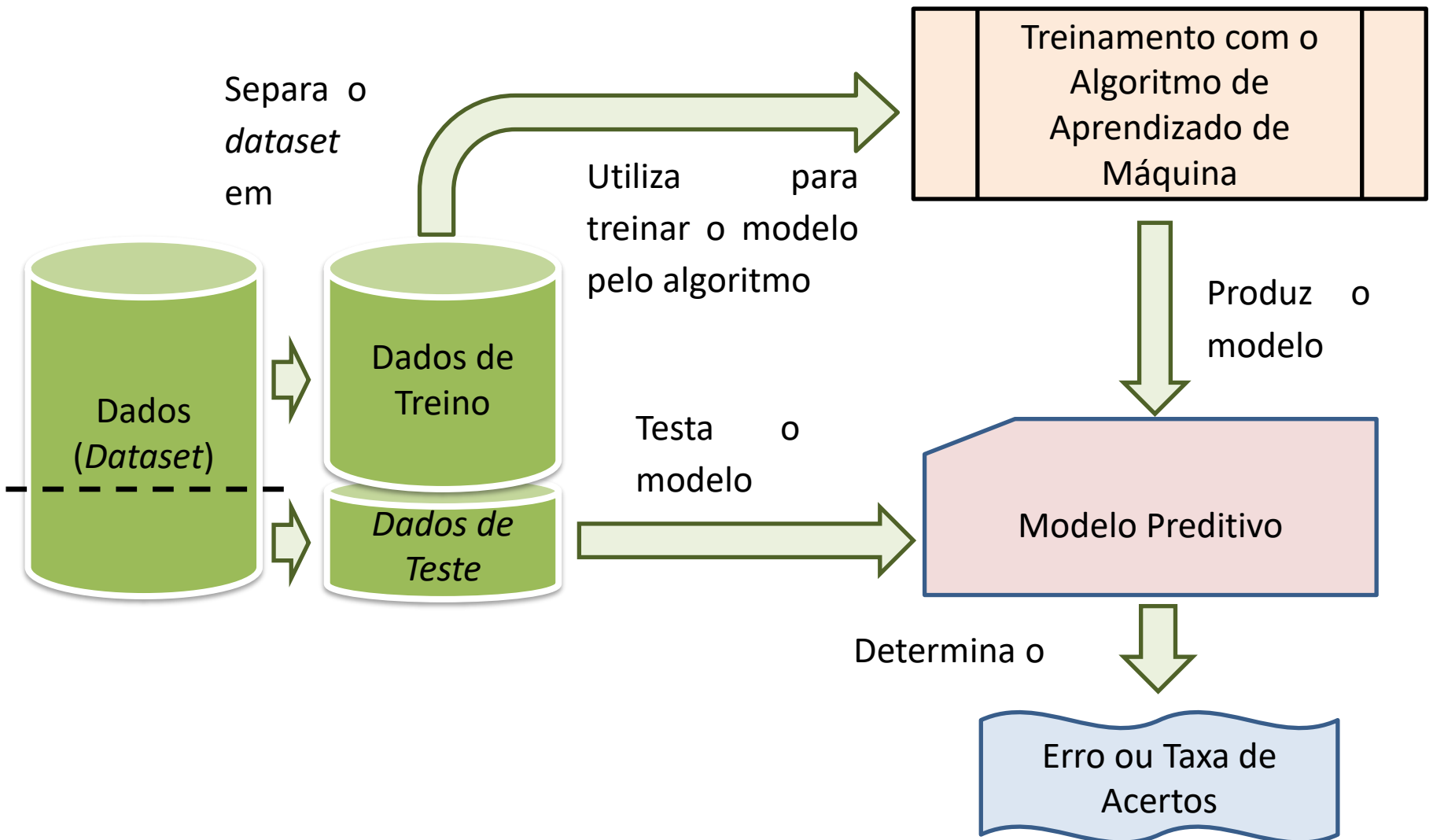
Aprendizado Supervisionado

Elementos do processo - Síntese



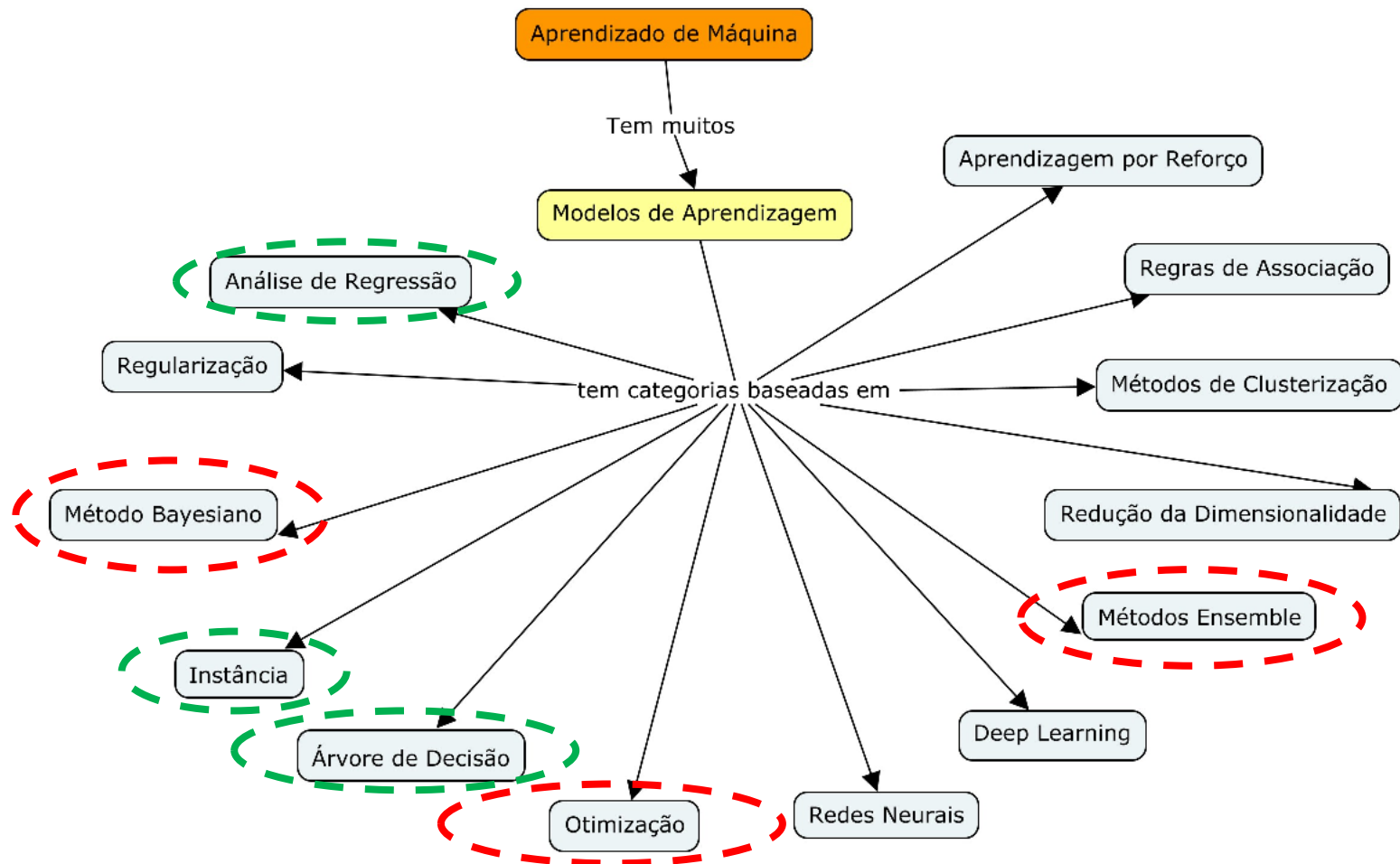
Aprendizado Supervisionado

O processo de Aprendizagem – Treinamento e Teste

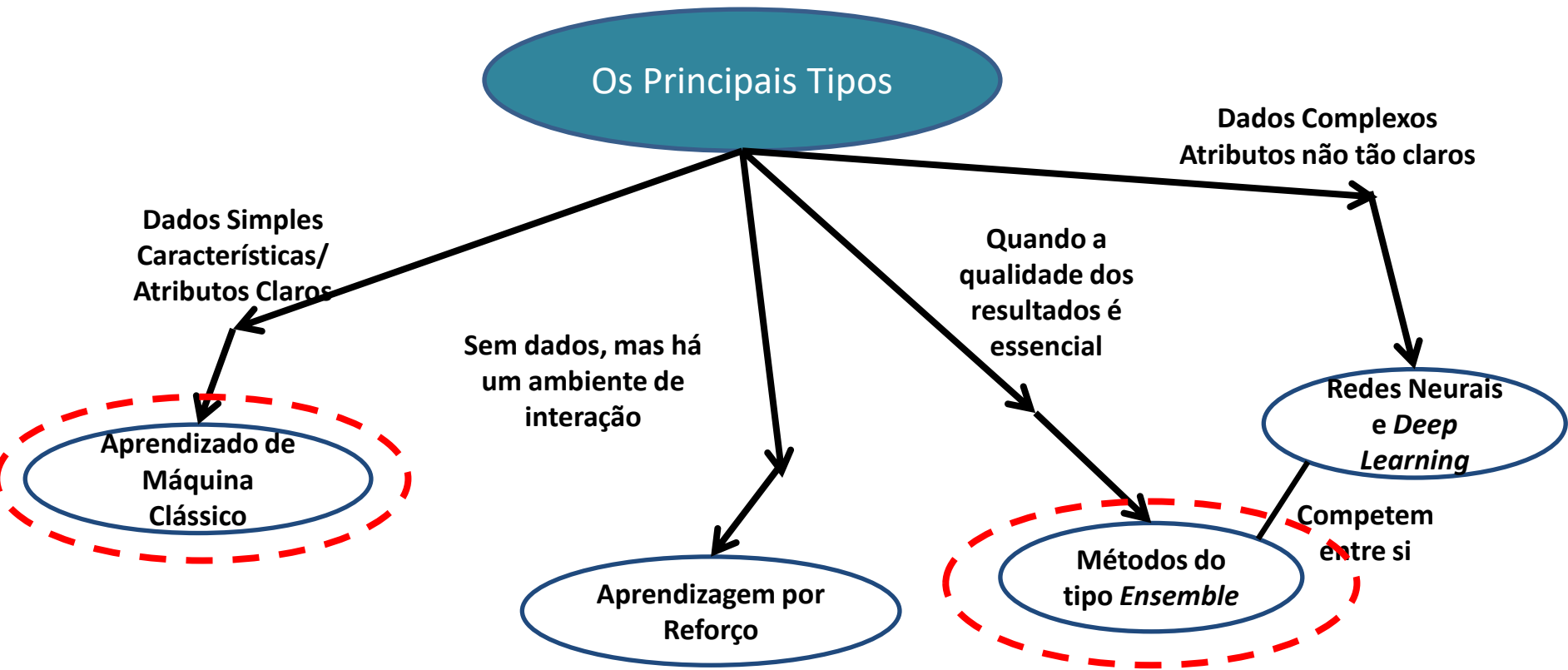


Aprendizado Supervisionado

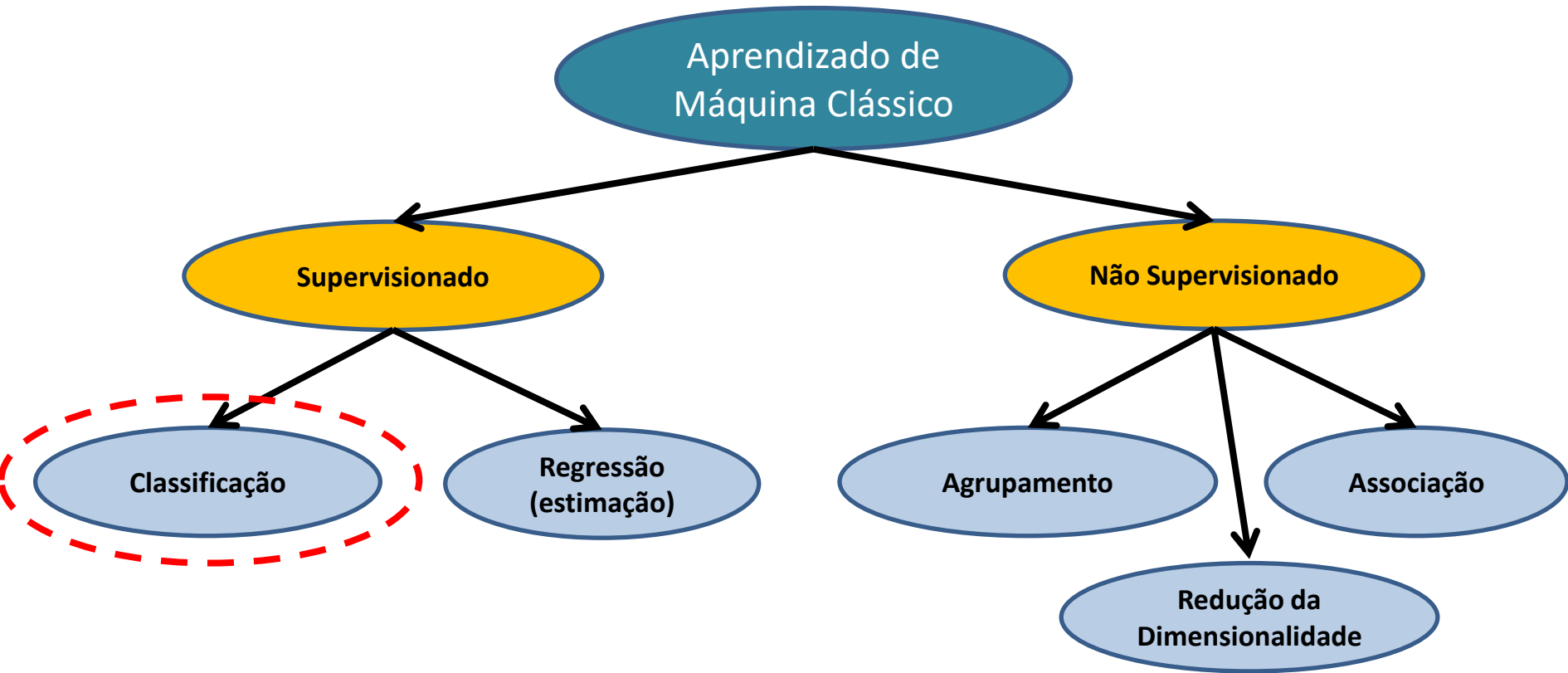
Modelos de Aprendizagem



Os principais tipos de Aprendizado de Máquina



Aprendizado de Máquina Clássico



Aprendizado Supervisionado

Algoritmos de Classificação – Classificador naïve Bayes

- Classificadores bayesianos são **classificadores estatísticos fundamentados no Teorema de Bayes** (e usado para **predizer a probabilidade de pertinência de um objeto a determinada classe**).
- Possui **desempenho comparável a redes neurais artificiais e árvores de decisão para alguns problemas**.
- Apresenta **alta acurácia e velocidade de processamento** quanto aplicados a grandes bases de dados.

Aprendizado Supervisionado

Algoritmo - Naïve Bayes

- É um algoritmo probabilístico **baseado no teorema de Bayes**.

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

Aprendizado Supervisionado

Algoritmos de Classificação – Classificador naïve Bayes

- **Assumem que o efeito do valor de um atributo em uma dada classe é independente dos valores dos outros atributos.**
- Tem como **premissa, independência condicional da classe** (*classe conditional independence*) com o intuito de **simplificar os cálculos**, e por causa dela, o **algoritmo é denominado naïve** (inocente, ingênuo).

Aprendizado Supervisionado

Algoritmos de Classificação – Classificador naïve Bayes - Exemplo

- Utilizando a base de dados Carros:

<https://archive.ics.uci.edu/dataset/19/car+evaluation>

- [**Knowledge acquisition and explanation for multi-attribute decision making**](#) By M. Bohanec, V. Rajkovič. 1988.

Published in 8th Intl Workshop on Expert Systems and their Applications, Avignon, France

- 1728 carros
- 7 atributos, sendo 6 de entrada e 1 alvo (Classe).

Aprendizado Supervisionado

Algoritmos de Classificação – Classificador naïve Bayes - Exemplo

Nome do Atributo	Papel	Tipo	Descrição	Valores
buying	Característica (Entrada)	Categórico	Preço de compra	vhigh, high, med, low
maint	Característica (Entrada)	Categórico	Preço de manutenção	vhigh, high, med, low
doors	Característica (Entrada)	Categórico	Número de Portas	2, 3, 4, 5more
persons	Característica (Entrada)	Categórico	Passageiros	2, 4, more
lug_boot	Característica (Entrada)	Categórico	Bagageiro	small, med, big,
safety	Característica (Entrada)	Categórico	Segurança	low, med, high
class	Alvo	Categórico	Aceitabilidade	unacc, acc, good, vgood

Aprendizado Supervisionado

Algoritmos de Classificação – Classificador naïve Bayes - Exemplo

Nome do Atributo	Papel	Tipo	Descrição	Valores
buying	Característica (Entrada)	Categórico	Preço de compra	vhigh = 1, high = 2, med = 3, low = 4
maint	Característica (Entrada)	Categórico	Preço de manutenção	vhigh = 1, high = 2, med = 3, low = 4
doors	Característica (Entrada)	Categórico	Número de Portas	2, 3, 4, 5more = 5
persons	Característica (Entrada)	Categórico	Passageiros	2, 4, more = 6
lug_boot	Característica (Entrada)	Categórico	Bagageiro	small = 1, med = 2, big = 3,
safety	Característica (Entrada)	Categórico	Segurança	low = 1, med = 2, high = 3
class	Alvo	Categórico	Aceitabilidade	unacc = 1, acc = 2, good = 3, vgood = 4

Aprendizado Supervisionado

Algoritmos de Classificação – Classificador naïve Bayes

- **O conceito de independência condicional de classe (ICC) considerando a base de dados Carro, a aceitabilidade do carro depende:**
 - preço de compra,
 - preço de manutenção,
 - número de portas,
 - quantidade de passageiros,
 - tamanho do bagageiro e
 - segurança.
- **A ICC do naïve Bayes assume que cada um desses fatores contribui de forma independente dos demais para a aceitabilidade final.**

Aprendizado Supervisionado

Algoritmos de Classificação – Classificador naïve Bayes – Teorema de Bayes

- Seja **x um objeto** cuja **classe C é conhecida** e **H uma hipótese** tal que o objeto **x pertença à classe C** .
- Em **problemas de classificação**, o **objetivo é determinar a probabilidade $P(H|x)$** , ou seja, a probabilidade de que a hipótese H seja satisfeita dada o objeto observado x .
- **$P(H|x)$ é a probabilidade a posteriori de H dado x e**
- **$P(H)$ é probabilidade a priori de H .**

Aprendizado Supervisionado

Algoritmos de Classificação – Classificador naïve Bayes – Teorema de Bayes

- **Por exemplo:** suponha uma **bases de dados** sobre **triagem de pacientes com problemas cardíacos**, sendo que a **primeira decisão de triagem é baseada na pressão arterial do paciente**.
- Assuma que **x é um paciente com pressão arterial alta** e que a hipótese **H é de que o atendimento deva ser urgente**.

Aprendizado Supervisionado

Algoritmos de Classificação – Classificador naïve Bayes – Teorema de Bayes

- **$P(H|x)$** indica a **confiança de que** o paciente **x** **seja atendido com urgência (H)**, **dado que a sua pressão é alta**.
- **$P(H)$** é a **probabilidade de que qualquer paciente seja atendido em urgência**.
- **$P(x|H)$** é a **probabilidade de que a pressão do paciente x seja alta dado que o atendimento é urgente (H)**.
- **$P(x)$** é a **probabilidade de que a pressão do paciente X seja alta**.

Aprendizado Supervisionado

Algoritmos de Classificação – Classificador naïve Bayes – Teorema de Bayes

- As probabilidades **$P(x)$, $P(H)$ e $P(x|H)$** são **estimativas** para a base de dados em um estudo.
- A utilidade do **Teorema de Bayes** reside no fato de que **ele fornece uma maneira de calcular a probabilidade a posteriori , $P(H|x)$** , a partir de $P(x)$, $P(H)$ e $P(x|H)$:

$$P(H|x) = \frac{P(x|H)P(H)}{P(x)}$$

Aprendizado Supervisionado

Algoritmos de Classificação – Classificador naïve Bayes – Construção

- O classificador bayesiano, ou **naïve Bayes**, opera da seguinte maneira:
 - Cada objeto é representado por um **vetor de características m-dimensional** $x = (x_1, x_2, \dots, x_m)$, o qual **representa uma medição sobre cada um dos m atributos** A_1, A_2, \dots, A_m .
 - Assuma que a **base de dados possui C classes** C_1, C_2, \dots, C_c .

Aprendizado Supervisionado

Algoritmos de Classificação – Classificador naïve Bayes – Construção

- **Dado um objeto x com classe desconhecida.**
- O classificador deve ser usado para **predizer a classe à qual esse objeto pertence com base na maior probabilidade a posteriori encontrada**, dado x .
- Ou seja, o classificador bayesiano determina uma **classe C_i para o objeto x** se, somente se:
$$P(C_i|x) > P(C_j|x), \forall j \neq i.$$

Aprendizado Supervisionado

Algoritmos de Classificação – Classificador naïve Bayes – Construção

- Como $P(x)$ é constante para todas as classes, somente $P(x | C_i) P(C_i)$ precisa ser maximizado.
- Se as probabilidades a priori não são conhecidas, assume-se que as classes possuem a mesma probabilidade $P(C_1) = P(C_2) = \dots = P(C_c)$, e o objetivo torna-se maximizar $P(x | C_i)$.
- As probabilidades a priori devem ser estimadas por $P(C_i) = s_i/s$, onde s_i é o número de objetos de treinamento da classe C_i e s , o total de objetos.

Aprendizado Supervisionado

Algoritmos de Classificação – Classificador naïve Bayes – Construção

- Para conjuntos de dados com muitos objetos, o **cálculo de $P(x|C_i)$** torna-se computacionalmente caro e, por isso, a premissa da independência condicional de classe é assumida para os atributos, de modo que (lembrando que x é um vetor de características m -dimensional dos m Atributos):

$$P(x|C_i) = \prod_{k=1}^m P(x_k|C_i)$$

Aprendizado Supervisionado

Algoritmos de Classificação – Classificador naïve Bayes – Construção

- As probabilidades $P(x_1 | C_i)$, $P(x_2 | C_i)$, ..., $P(x_m | C_i)$ podem ser estimadas a partir dos objetos de entrada $P(x_i | C_i)$, onde:
 - Se A_k é categórico, então $P(x_k | C_i) = s_{ik}/s_i$, onde s_{ik} é o número de objetos da classe C_i com valor x_k para A_k e s_i é o número de objetos de treinamento pertencentes a classe C_i .

Aprendizado Supervisionado

Algoritmos de Classificação – Classificador naïve Bayes – Construção

- Se **A_k é contínuo**, então, o atributo assume tipicamente uma **distribuição de probabilidade gaussiana**, de forma que:

$$\begin{aligned} P(x_k | C_i) &= g(x_k, \mu_{C_i}, \sigma_{C_i}) \\ &= \frac{1}{\sqrt{2\pi}\sigma_{C_i}} \exp \left(-\frac{(x_k - \mu_{C_i})^2}{2\sigma_{C_i}^2} \right) \end{aligned}$$

- Onde $g(x_k, \mu_{C_i}, \sigma_{C_i})$ é a **função densidade gaussiana** ou normal para o atributo A_k enquanto μ_{C_i} e σ_{C_i} são a média e ao desvio padrão, respectivamente, do atributo A_k para as amostras da classe C_i .

Aprendizado Supervisionado

Algoritmos de Classificação – Classificador naïve Bayes – Construção

- Para **classificar um objeto x de classe desconhecida**, $P(x | C_i)P(C_i)$ é avaliado para cada classe C_i .
- O objeto x é especificado à classe C_i se, e somente se:

$$P(x | C_i) P(C_i) > P(x | C_j) P(C_j)), \forall j \neq i.$$

Aprendizado Supervisionado

Algoritmos de Classificação – Classificador naïve Bayes – Construção

- Um problema pode ocorrer com o naïve Bayes **se determinado atributo não aparecer na base de dados conjuntamente com cada valor de classe.**
- Nesse caso, $P(X | C_j) = 0$, e o resultado da **equação**

$$P(x | C_i) = \prod_{k=1}^m P(x_k | C_i)$$

Também será 0.

Aprendizado Supervisionado

Algoritmos de Classificação – Classificador naïve Bayes – Construção

- Esse **problema pode ser resolvido**, por exemplo, **adicionando-se uma constante μ/m a cada numerador da eq.**
$$P(x|C_i) = \prod_{k=1}^m P(x_k|C_i)$$
 e μ ao denominador, para compensar a constante do numerador.
- Essa **estratégia** de adicionar uma constante a cada numerador é uma técnica padrão **denominada estimador de Laplace.**
- **Independente do objeto a ser classificado, é necessário calcular a matriz de probabilidades para os atributos e seus valores em relação às classes do problema.**

Aprendizado Supervisionado

Algoritmos de Classificação – Classificador naïve Bayes – Exemplo

- Como exemplo, suponha a base de dados Carros.
- Consideremos como entrada a **matriz de probabilidades na tabela a seguir** e as **probabilidades para cada classe** sendo $P(\text{inaceitável}) = 0,700$, $P(\text{aceitável}) = 0,222$, $P(\text{bom}) = 0,040$ e $P(\text{ótimo}) = 0,038$.
- O objetivo é **classificar o objeto x** descrito pelos **atributos: compra = médio, manutenção = médio, portas = 4, passageiros = 4, bagageiro = médio e segurança = média**.

Atributo	Valor	Classe			
		inaceitável	aceitável	bom	ótimo
Compra	muito alto	0,298	0,188	0,000	0,000
Compra	alto	0,268	0,281	0,000	0,000
Compra	médio	0,221	0,299	0,333	0,400
Compra	baixo	0,213	0,232	0,667	0,600
Manutenção	muito alto	0,298	0,188	0,000	0,000
Manutenção	alto	0,260	0,273	0,000	0,200
Manutenção	médio	0,221	0,299	0,333	0,400
Manutenção	baixo	0,221	0,240	0,667	0,400
Portas	2	0,269	0,211	0,217	0,153
Portas	3	0,248	0,258	0,261	0,231
Portas	4	0,241	0,266	0,261	0,308
Portas	5mais	0,241	0,266	0,261	0,308
Passageiros	2	0,476	0,000	0,000	0,000
Passageiros	4	0,258	0,516	0,522	0,462
Passageiros	mais	0,266	0,484	0,478	0,538
Bagageiro	pequeno	0,372	0,273	0,304	0,000
Bagageiro	médio	0,324	0,352	0,348	0,385
Bagageiro	grande	0,304	0,375	0,348	0,615
Segurança	baixa	0,476	0,000	0,000	0,000
Segurança	média	0,295	0,469	0,565	0,000
Segurança	alta	0,229	0,531	0,435	1,000

compra = médio,

Atributo	Valor	Classe			
		inaceitável	aceitável	bom	ótimo
Compra	muito alto	0,298	0,188	0,000	0,000
Compra	alto	0,268	0,281	0,000	0,000
Compra	médio	0,221	0,299	0,333	0,400
Compra	baixo	0,213	0,232	0,667	0,600
Manutenção	muito alto	0,298	0,188	0,000	0,000
Manutenção	alto	0,260	0,273	0,000	0,200
Manutenção	médio	0,221	0,299	0,333	0,400
Manutenção	baixo	0,221	0,240	0,667	0,400
Portas	2	0,269	0,211	0,217	0,153
Portas	3	0,248	0,258	0,261	0,231
Portas	4	0,241	0,266	0,261	0,308
Portas	5mais	0,241	0,266	0,261	0,308
Passageiros	2	0,476	0,000	0,000	0,000
Passageiros	4	0,258	0,516	0,522	0,462
Passageiros	mais	0,266	0,484	0,478	0,538
Bagageiro	pequeno	0,372	0,273	0,304	0,000
Bagageiro	médio	0,324	0,352	0,348	0,385
Bagageiro	grande	0,304	0,375	0,348	0,615
Segurança	baixa	0,476	0,000	0,000	0,000
Segurança	média	0,295	0,469	0,565	0,000
Segurança	alta	0,229	0,531	0,435	1,000

compra = médio,
manutenção = médio,

Atributo	Valor	Classe			
		inaceitável	aceitável	bom	ótimo
Compra	muito alto	0,298	0,188	0,000	0,000
Compra	alto	0,268	0,281	0,000	0,000
Compra	médio	0,221	0,299	0,333	0,400
Compra	baixo	0,213	0,232	0,667	0,600
Manutenção	muito alto	0,298	0,188	0,000	0,000
Manutenção	alto	0,260	0,273	0,000	0,200
Manutenção	médio	0,221	0,299	0,333	0,400
Manutenção	baixo	0,221	0,240	0,667	0,400
Portas	2	0,269	0,211	0,217	0,153
Portas	3	0,248	0,258	0,261	0,231
Portas	4	0,241	0,266	0,261	0,308
Portas	5mais	0,241	0,266	0,261	0,308
Passageiros	2	0,476	0,000	0,000	0,000
Passageiros	4	0,258	0,516	0,522	0,462
Passageiros	mais	0,266	0,484	0,478	0,538
Bagageiro	pequeno	0,372	0,273	0,304	0,000
Bagageiro	médio	0,324	0,352	0,348	0,385
Bagageiro	grande	0,304	0,375	0,348	0,615
Segurança	baixa	0,476	0,000	0,000	0,000
Segurança	média	0,295	0,469	0,565	0,000
Segurança	alta	0,229	0,531	0,435	1,000

compra = médio,
manutenção = médio,
portas = 4,

Atributo	Valor	Classe			
		inaceitável	aceitável	bom	ótimo
Compra	muito alto	0,298	0,188	0,000	0,000
Compra	alto	0,268	0,281	0,000	0,000
Compra	médio	0,221	0,299	0,333	0,400
Compra	baixo	0,213	0,232	0,667	0,600
Manutenção	muito alto	0,298	0,188	0,000	0,000
Manutenção	alto	0,260	0,273	0,000	0,200
Manutenção	médio	0,221	0,299	0,333	0,400
Manutenção	baixo	0,221	0,240	0,667	0,400
Portas	2	0,269	0,211	0,217	0,153
Portas	3	0,248	0,258	0,261	0,231
Portas	4	0,241	0,266	0,261	0,308
Portas	5mais	0,241	0,266	0,261	0,308
Passageiros	2	0,476	0,000	0,000	0,000
Passageiros	4	0,258	0,516	0,522	0,462
Passageiros	mais	0,266	0,484	0,478	0,538
Bagageiro	pequeno	0,372	0,273	0,304	0,000
Bagageiro	médio	0,324	0,352	0,348	0,385
Bagageiro	grande	0,304	0,375	0,348	0,615
Segurança	baixa	0,476	0,000	0,000	0,000
Segurança	média	0,295	0,469	0,565	0,000
Segurança	alta	0,229	0,531	0,435	1,000

compra = médio,
 manutenção = médio,
 portas = 4,
 passageiros = 4,

Atributo	Valor	Classe			
		inaceitável	aceitável	bom	ótimo
Compra	muito alto	0,298	0,188	0,000	0,000
Compra	alto	0,268	0,281	0,000	0,000
Compra	médio	0,221	0,299	0,333	0,400
Compra	baixo	0,213	0,232	0,667	0,600
Manutenção	muito alto	0,298	0,188	0,000	0,000
Manutenção	alto	0,260	0,273	0,000	0,200
Manutenção	médio	0,221	0,299	0,333	0,400
Manutenção	baixo	0,221	0,240	0,667	0,400
Portas	2	0,269	0,211	0,217	0,153
Portas	3	0,248	0,258	0,261	0,231
Portas	4	0,241	0,266	0,261	0,308
Portas	5mais	0,241	0,266	0,261	0,308
Passageiros	2	0,476	0,000	0,000	0,000
Passageiros	4	0,258	0,516	0,522	0,462
Passageiros	mais	0,266	0,484	0,478	0,538
Bagageiro	pequeno	0,372	0,273	0,304	0,000
Bagageiro	médio	0,324	0,352	0,348	0,385
Bagageiro	grande	0,304	0,375	0,348	0,615
Segurança	baixa	0,476	0,000	0,000	0,000
Segurança	média	0,295	0,469	0,565	0,000
Segurança	alta	0,229	0,531	0,435	1,000

compra = médio,
 manutenção = médio,
 portas = 4,
 passageiros = 4,
 bagageiro = médio

Atributo	Valor	Classe			
		inaceitável	aceitável	bom	ótimo
Compra	muito alto	0,298	0,188	0,000	0,000
Compra	alto	0,268	0,281	0,000	0,000
Compra	médio	0,221	0,299	0,333	0,400
Compra	baixo	0,213	0,232	0,667	0,600
Manutenção	muito alto	0,298	0,188	0,000	0,000
Manutenção	alto	0,260	0,273	0,000	0,200
Manutenção	médio	0,221	0,299	0,333	0,400
Manutenção	baixo	0,221	0,240	0,667	0,400
Portas	2	0,269	0,211	0,217	0,153
Portas	3	0,248	0,258	0,261	0,231
Portas	4	0,241	0,266	0,261	0,308
Portas	5mais	0,241	0,266	0,261	0,308
Passageiros	2	0,476	0,000	0,000	0,000
Passageiros	4	0,258	0,516	0,522	0,462
Passageiros	mais	0,266	0,484	0,478	0,538
Bagageiro	pequeno	0,372	0,273	0,304	0,000
Bagageiro	médio	0,324	0,352	0,348	0,385
Bagageiro	grande	0,304	0,375	0,348	0,615
Segurança	baixa	0,476	0,000	0,000	0,000
Segurança	média	0,295	0,469	0,565	0,000
Segurança	alta	0,229	0,531	0,435	1,000

compra = médio,
manutenção = médio,
portas = 4,
passageiros = 4,
bagageiro = médio e
segurança = média.

Atributo	Valor	Classe			
		inaceitável	aceitável	bom	ótimo
Compra	muito alto	0,298	0,188	0,000	0,000
Compra	alto	0,268	0,281	0,000	0,000
Compra	médio	0,221	0,299	0,333	0,400
Compra	baixo	0,213	0,232	0,667	0,600
Manutenção	muito alto	0,298	0,188	0,000	0,000
Manutenção	alto	0,260	0,273	0,000	0,200
Manutenção	médio	0,221	0,299	0,333	0,400
Manutenção	baixo	0,221	0,240	0,667	0,400
Portas	2	0,269	0,211	0,217	0,153
Portas	3	0,248	0,258	0,261	0,231
Portas	4	0,241	0,266	0,261	0,308
Portas	5mais	0,241	0,266	0,261	0,308
Passageiros	2	0,476	0,000	0,000	0,000
Passageiros	4	0,258	0,516	0,522	0,462
Passageiros	mais	0,266	0,484	0,478	0,538
Bagageiro	pequeno	0,372	0,273	0,304	0,000
Bagageiro	médio	0,324	0,352	0,348	0,385
Bagageiro	grande	0,304	0,375	0,348	0,615
Segurança	baixa	0,476	0,000	0,000	0,000
Segurança	média	0,295	0,469	0,565	0,000
Segurança	alta	0,229	0,531	0,435	1,000

compra = médio,
 manutenção = médio,
 portas = 4,
 passageiros = 4,
 bagageiro = médio e
 segurança = média.

Ak é categórico, então
 $P(x_k | C_i) = s_{ik}/s_i$, onde
 s_{ik} é o número de
 objetos da classe C_i com
 valor x_k para A_k e s_i é o
 número de objetos de
 treinamento pertencentes a
 classe C_i.

Atributo	Valor	Classe			
		inaceitável	aceitável	bom	ótimo
Compra	muito alto	0,298	0,188	0,000	0,000
Compra	alto	0,268	0,281	0,000	0,000
Compra	médio	0,221	0,299	0,333	0,400
Compra	baixo	0,213	0,232	0,667	0,600
Manutenção	muito alto	0,298	0,188	0,000	0,000
Manutenção	alto	0,260	0,273	0,000	0,200
Manutenção	médio	0,221	0,299	0,333	0,400
Manutenção	baixo	0,221	0,240	0,667	0,400
Portas	2	0,269	0,211	0,217	0,153
Portas	3	0,248	0,258	0,261	0,231
Portas	4	0,241	0,266	0,261	0,308
Portas	5mais	0,241	0,266	0,261	0,308
Passageiros	2	0,476	0,000	0,000	0,000
Passageiros	4	0,258	0,516	0,522	0,462
Passageiros	mais	0,266	0,484	0,478	0,538
Bagageiro	pequeno	0,372	0,273	0,304	0,000
Bagageiro	médio	0,324	0,352	0,348	0,385
Bagageiro	grande	0,304	0,375	0,348	0,615
Segurança	baixa	0,476	0,000	0,000	0,000
Segurança	média	0,295	0,469	0,565	0,000
Segurança	alta	0,229	0,531	0,435	1,000

compra = médio,
 manutenção = médio,
 portas = 4,
 passageiros = 4,
 bagageiro = médio e
 segurança = média.

Ak é categórico, então
 $P(x_k | C_i) = s_{ik}/s_i$, onde
 sik é o número de
 objetos da classe Ci com
 valor xk para Ak e si é o
 número de objetos de
 treinamento pertencentes a
 classe Ci.

Compra é categórico,
 então $P(\text{médio} | \text{bom}) =$
0,333, ou seja, **sik/si**,
 onde **sik** é o número de
 objetos da **classe bom**
 com valor **médio** para
Compra e **si** é o número
 de objetos de
 treinamento pertencentes a
classe bom.

Aprendizado Supervisionado

Algoritmos de Classificação – Classificador naïve Bayes – Exemplo

- Partindo dos dados de entrada, o algoritmo calcula a probabilidade de o objeto x pertencer a cada classe:
 - $P(\text{inaceitável}/x) = P(x/\text{inaceitável}) \times P(\text{inaceitável})$
 - $P(\text{aceitável}/x) = P(x/\text{aceitável}) \times P(\text{aceitável})$
 - $P(\text{bom}/x) = P(x/\text{bom}) \times P(\text{bom})$
 - $P(\text{ótimo}/x) = P(x/\text{ótimo}) \times P(\text{ótimo})$
- Lembrando: $P(\text{inaceitável}) = 0,700$, $P(\text{aceitável}) = 0,222$, $P(\text{bom}) = 0,040$ e $P(\text{ótimo}) = 0,038$

Atributo	Valor	Classe			
		inaceitável	aceitável	bom	ótimo
Compra	muito alto	0,298	0,188	0,000	0,000
Compra	alto	0,268	0,281	0,000	0,000
Compra	médio	0,221	0,299	0,333	0,400
Compra	baixo	0,213	0,232	0,667	0,600
Manutenção	muito alto	0,298	0,188	0,000	0,000
Manutenção	alto	0,260	0,273	0,000	0,200
Manutenção	médio	0,221	0,299	0,333	0,400
Manutenção	baixo	0,221	0,240	0,667	0,400
Portas	2	0,269	0,211	0,217	0,153
Portas	3	0,248	0,258	0,261	0,231
Portas	4	0,241	0,266	0,261	0,308
Portas	5mais	0,241	0,266	0,261	0,308
Passageiros	2	0,476	0,000	0,000	0,000
Passageiros	4	0,258	0,516	0,522	0,462
Passageiros	mais	0,266	0,484	0,478	0,538
Bagageiro	pequeno	0,372	0,273	0,304	0,000
Bagageiro	médio	0,324	0,352	0,348	0,385
Bagageiro	grande	0,304	0,375	0,348	0,615
Segurança	baixa	0,476	0,000	0,000	0,000
Segurança	média	0,295	0,469	0,565	0,000
Segurança	alta	0,229	0,531	0,435	1,000

$$P(\text{inaceitável}/x) = \\ P(x/\text{inaceitável}) \times \\ P(\text{inaceitável})$$

$$P(\text{inaceitável}/x) = \\ P(x/\text{inaceitável}) \times 0,700$$

$$P(\text{inaceitável}/x) = \\ (0,221 \times 0,221 \times 0,241 \times 0,258 \times \\ 0,324 \times 0,295) \times 0,700$$

$$P(\text{inaceitável}/x) = \\ 2,04 \times 10^{-4}$$

Aprendizado Supervisionado

Algoritmos de Classificação – Classificador naïve Bayes – Exemplo

- Partindo dos dados de entrada, o algoritmo calcula a probabilidade de o objeto x pertencer a cada classe:

$$\begin{aligned} & P(x|inaceitável) \times P(inaceitável) \\ &= \left(\frac{0,221 \times 0,221 \times 0,241 \times 0,258}{0,324 \times 0,295} \right) \times 0,700 \\ &= 2,04 \times 10^{-4} \end{aligned}$$

$$\begin{aligned} & P(x|aceitável) \times P(aceitável) \\ &= \left(\frac{0,299 \times 0,229 \times 0,266 \times 0,516}{0,352 \times 0,469} \right) \times 0,222 \\ &= 4,5 \times 10^{-4} \end{aligned}$$

Aprendizado Supervisionado

Algoritmos de Classificação – Classificador naïve Bayes – Exemplo

- Partindo dos dados de entrada o algoritmo calcula a probabilidade de o objeto x pertencer a cada classe:

$$\begin{aligned} &P(x|bom) \times P(bom) \\ &= \left(\begin{array}{c} 0,333 \times 0,333 \times 0,261 \times 0,522 \times \\ 0,348 \times 0,565 \end{array} \right) \times 0,040 \\ &= 1,19 \times 10^{-4} \end{aligned}$$

$$\begin{aligned} &P(x|ótimo) \times P(ótimo) \\ &= \left(\begin{array}{c} 0,4 \times 0,4 \times 0,308 \times 0,462 \times \\ 0,385 \times 0,0 \end{array} \right) \times 0,038 \\ &= 0,000 \end{aligned}$$

Aprendizado Supervisionado

Algoritmos de Classificação – Classificador naïve Bayes – Exemplo

- Para **determinar a classe do objeto x, basta encontrar a probabilidade máxima dada pelo algoritmo, que é $4,50 \times 10^{-4}$ para a classe aceitável.**

Aprendizado Supervisionado

Algoritmo - Naïve Bayes - Exemplo

- **O método probabilístico é bem indicado quando eu tenho uma informação a priori**, enquanto o baseado em instâncias (por exemplo, KNN) não considera esta informação.

Aprendizado Supervisionado

Algoritmo - Naïve Bayes - Exemplo

- **Quem faz os cálculos é o algoritmo**, o nosso trabalho é identificar que valores passar a ele, ou seja, é muito mais um trabalho de compreensão do problema do que de estatística propriamente dito.

Aprendizado Supervisionado

Algoritmo - Naïve Bayes

- Existem três **tipos de modelos Naive Bayes** na biblioteca ***scikit-learn***:
 - **Gaussiano**: é usado na classificação e assume que as características seguem uma distribuição normal de probabilidade.
 - **Multinomial**: É usado para contagens discretas.
 - **Bernoulli**: o modelo binomial é útil se seus vetores de características são binários (ou seja, zeros e uns).

Aprendizado Supervisionado

Algoritmo - Naïve Bayes - Aplicações

- **Previsão em tempo real:** Naive Bayes é um classificador de aprendizagem rápido. Assim, ele pode ser usado para fazer previsões em tempo real.
- **Predição multiclasse:** Este algoritmo também é conhecido pelo recurso de predição multiclasse. Ou seja, pode-se prever a probabilidade de várias classes da variável alvo.

Aprendizado Supervisionado

Algoritmo - Naïve Bayes - Aplicações

- **Classificação de texto:** muito usados para a classificação de texto (devido ao melhor resultado em problemas de multiclasse e regra de independência) têm maior taxa de sucesso em comparação com outros algoritmos.
- **Sistema de recomendação:** o classificador Naive Bayes e a filtragem colaborativa juntos criam um sistema de recomendação que usa técnicas de aprendizado de máquina e mineração de dados para filtrar informações invisíveis e prever se um usuário gostaria de um determinado recurso ou não.

Aprendizado Supervisionado

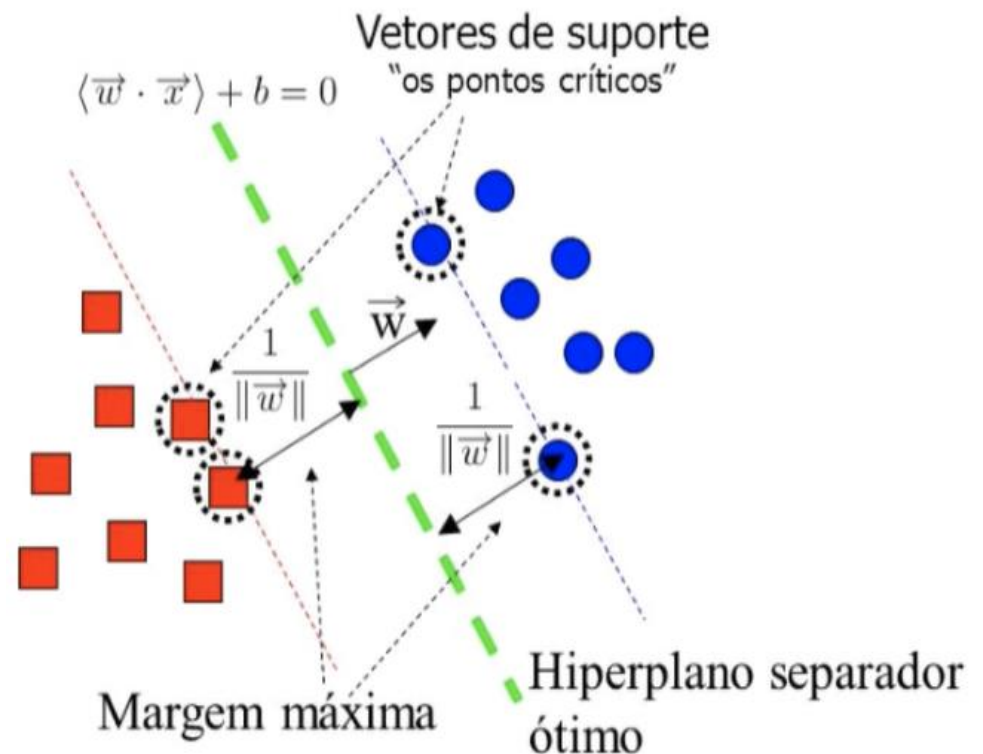
Algoritmo - Máquinas de Vetores de Suporte (*Support Vector Machines* - SVMs)

- As **SVMs** pertence a categoria de algoritmos baseada em **otimização**.
- A SVM é uma **técnica de aprendizado estatístico**, baseada no princípio da Minimização do Risco Estrutural (SEM) e **pode ser usada para resolver problemas de classificação e regressão**.
- Tem sua origem na aplicação de conceitos da teoria do aprendizado estatístico.
- O treinamento das SVMs **envolve a solução de um problema de otimização quadrática formulada com o objetivo de maximizar a margem de separação entre os objetos de diferentes classes**.

Aprendizado Supervisionado

Algoritmo - Máquinas de Vetores de Suporte (SVMs) - Funcionamento

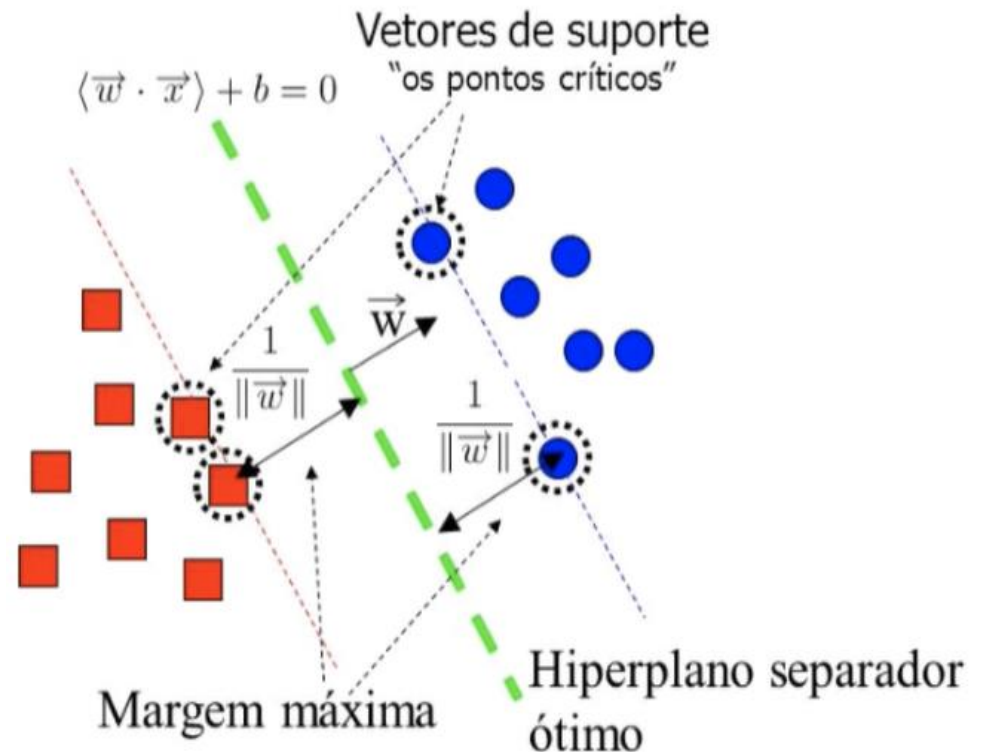
- Na figura temos a linha verde tracejada (representa o hiperplano) – uma linha que separa as duas classes.



Aprendizado Supervisionado

Algoritmo - Máquinas de Vetores de Suporte (SVMs) - Funcionamento

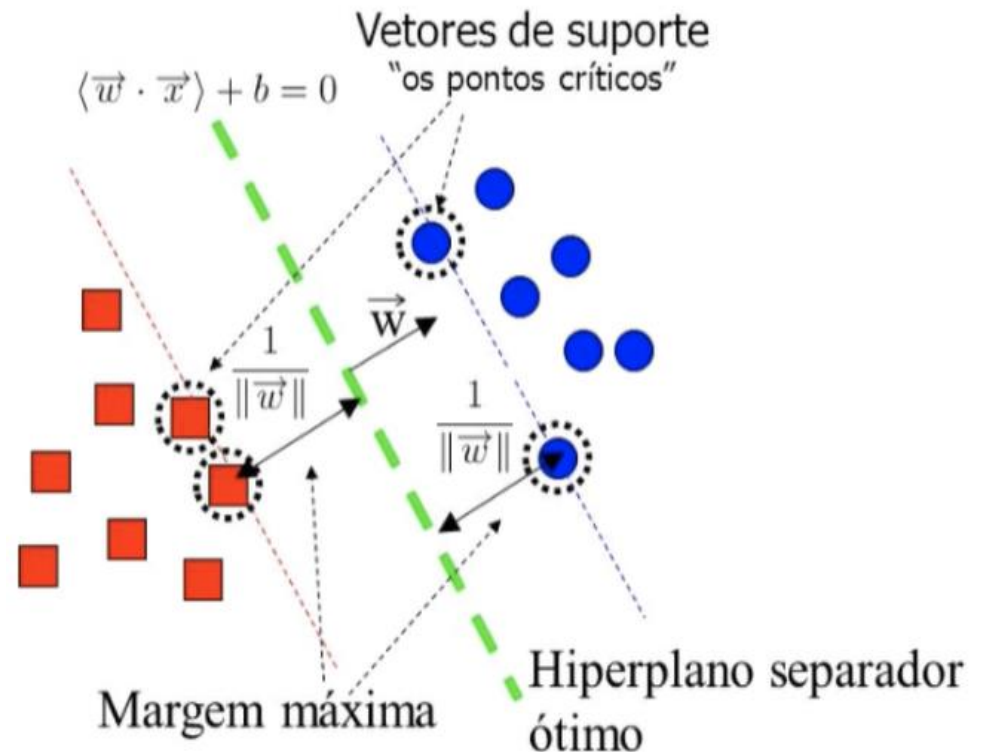
- Os pontos de dados que representam a classe vermelha e os pontos de dados que representam a classe azul.



Aprendizado Supervisionado

Algoritmo - Máquinas de Vetores de Suporte (SVMs) - Funcionamento

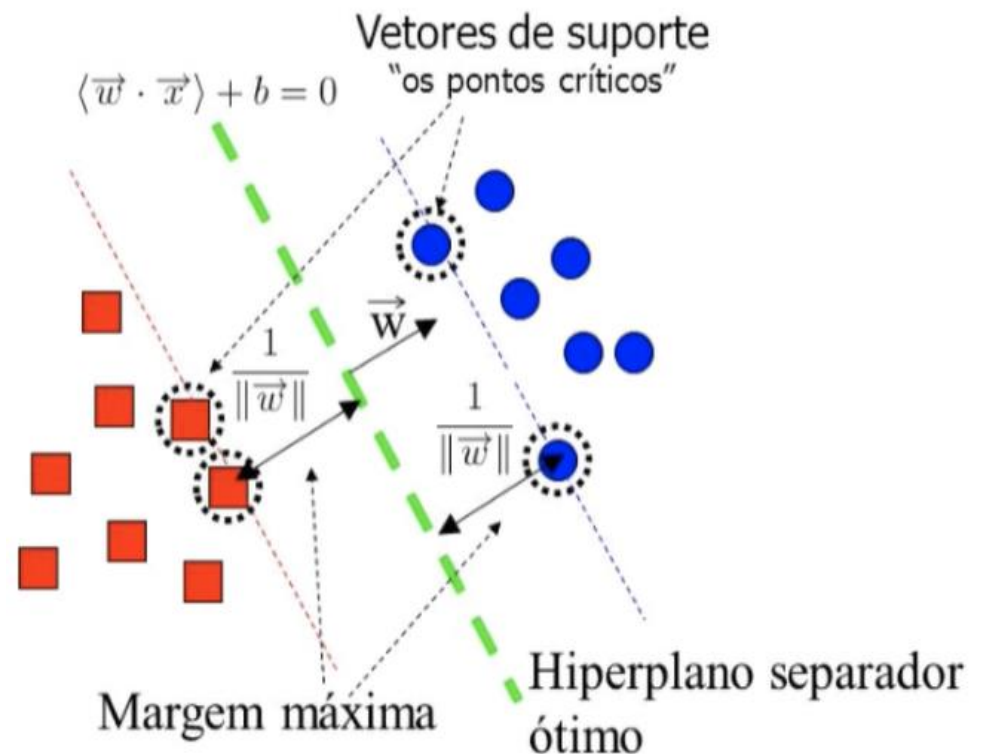
- Para **encontrar o hiperplano é necessário calcular as margens de cada classe (margem máxima).**



Aprendizado Supervisionado

Algoritmo - Máquinas de Vetores de Suporte (SVMs) - Funcionamento

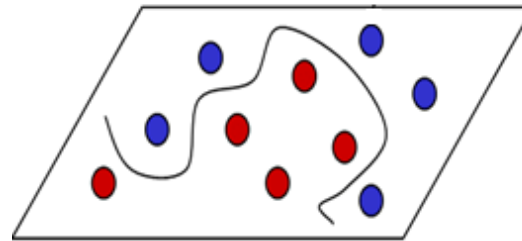
- Esses **elementos que estão na margem** são chamados de **vetores de suporte** ou apenas pontos críticos.
- Os dados devem estar normalizados para obter melhores resultados.



Aprendizado Supervisionado

Algoritmo - Máquinas de Vetores de Suporte (SVMs) - Funcionamento

- O algoritmo faz:
 - 1) Verifica se os dados são linearmente separáveis;

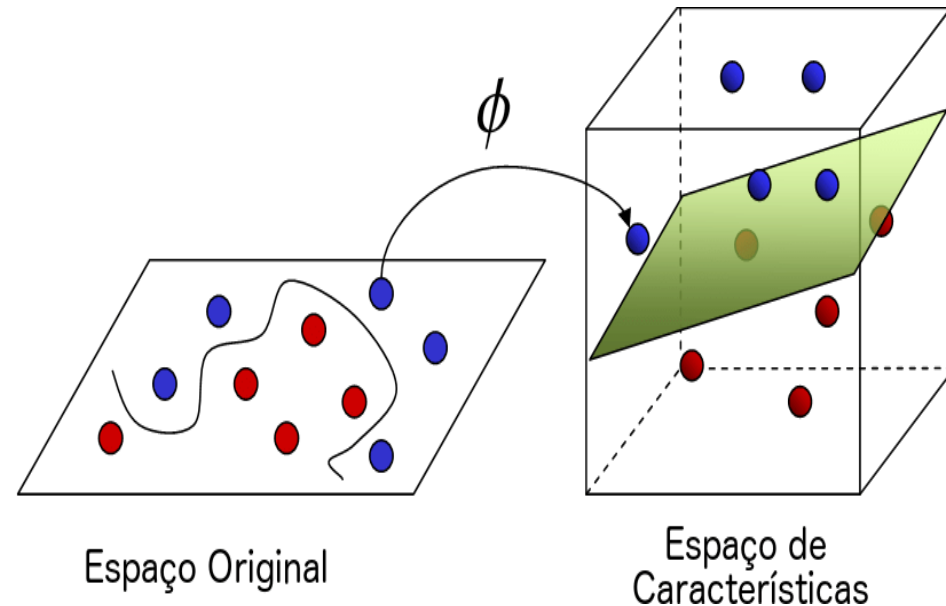


Espaço Original

Aprendizado Supervisionado

Algoritmo - Máquinas de Vetores de Suporte (SVMs) - Funcionamento

- O algoritmo faz:
 - 1) Verifica se os dados são linearmente separáveis;
 - 2) Caso não haja uma separação linear, então, aumento a dimensão (crio uma nova dimensão).
 - 3) E então, passo a ter os dados linearmente separáveis.
 - 4) A partir daí, os vetores de suporte são encontrados.
 - 5) Então, conseguimos encontrar o hiperplano (linha exatamente no meio dos dados).



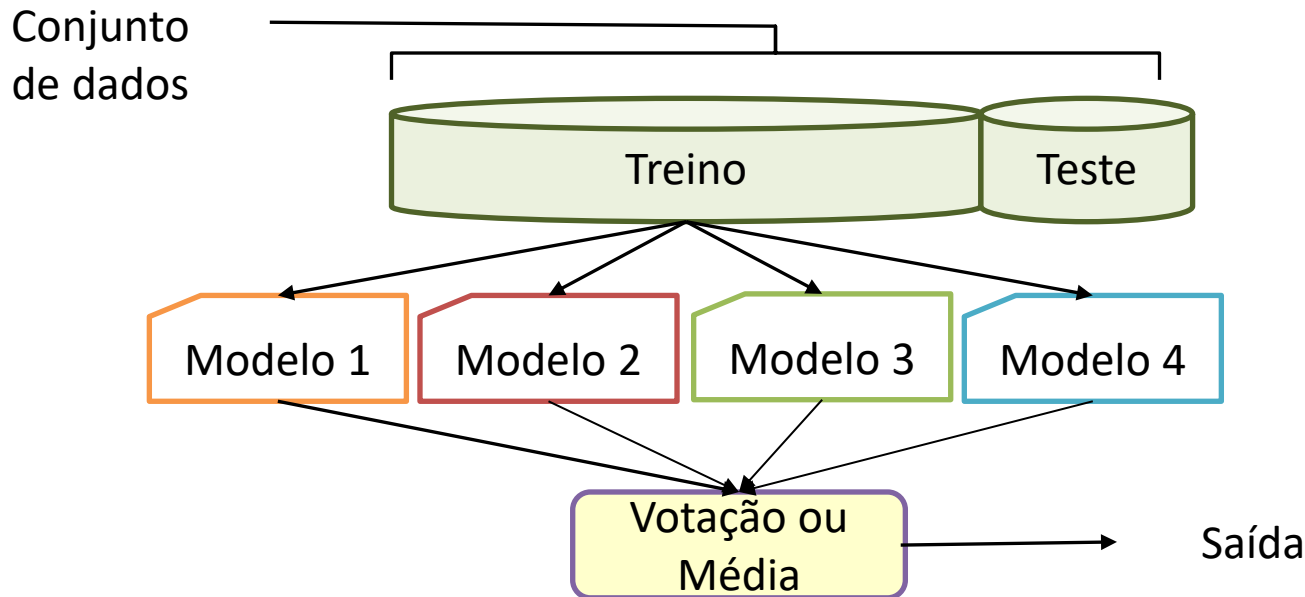
Aprendizado Supervisionado

Algoritmo - Métodos do tipo ensemble

- É um **grupo de algoritmos de aprendizado de máquina treinados em paralelo, unindo ao final os resultados obtidos.**
- **Ensemble:** palavra francesa que significa **conjunto - grupo.**
- Em geral, requer um pouco mais de capacidade computacional, mas **consegue obter uma precisão muito alta.**

Aprendizado Supervisionado

Algoritmo - Métodos do tipo ensemble



- Os dados de treino são utilizados para alimentar quatro modelos de maneira simultânea.
- Então, no final, é possível calcular: a média, votação, fazer com que um modelo corrija o erro do outro, são várias técnicas dentro dos métodos ensemble.

Aprendizado Supervisionado

Algoritmo - Métodos do tipo ensemble

- Considere as opções:
 - 1) Utilizar o mesmo *dataset* de treino para os quatro algoritmos diferentes.
 - 2) Utilizar o *dataset* de treino e coletar quatro amostras diferentes com técnicas estatísticas de amostragem aleatória.
- Alimentar os algoritmos de aprendizado, fazer o treinamento e criar quatro classificadores.
- Por exemplo, atribuir pesos a cada um desses classificadores.
- No final, combinar as capacidades de classificação desses modelos, alimentando com novos dados e fazer as previsões.

Aprendizado Supervisionado

Algoritmo - Métodos do tipo ensemble

- Como serão gerados (treinados) mais modelos de Aprendizado de Máquina, tem-se um **maior custo (tempo)** e **necessidade de maior capacidade computacional**.
- Então, se for necessário uma **acurácia (precisão) maior**, **método ensemble** são indicados, porém será necessário **maior tempo para treinamento** e conseqüentemente de uma **máquina mais poderosa**.

Aprendizado Supervisionado

Algoritmo - Métodos do tipo ensemble - Exemplo

- Considere **os três algoritmos: KNN, Árvore de Decisão e SVM**. Além disso, **uma amostra do conjunto de dados de treino diferente ou os mesmos dados de treino**.
- **Cria-se o modelo preditivo de cada um deles.**
- No final, utiliza-se **uma técnica para calcular a média dos resultados, para fazer um sistema de votação, ou qualquer outra estratégia.**

Aprendizado Supervisionado

Algoritmo - Métodos do tipo ensemble - Exemplo

- Abordagens principais:
- ***Bootstrap Aggregation* ou *Bagging***
- ***Boosting***
- ***Votação***

Aprendizado Supervisionado

Algoritmo - Métodos do tipo ensemble – **Abordagens Principais - *Bagging***

- ***Bootstrap Aggregation* ou *Bagging***: é uma técnica que faz uso de algoritmos do mesmo tipo em paralelo para criar modelos diferentes. Por exemplo, árvores de decisão em paralelo.
- **Algoritmos exemplo:**
 - Bagged CART (*Classification and Regression Trees*);
 - Random Forest.

Aprendizado Supervisionado

Algoritmo - Métodos do tipo ensemble - ***Random Forest*** (Floresta Aleatória)

- Usa **vários algoritmos de árvore de decisão considerados fracos que quando combinados torna-se poderoso.**
- Os vários algoritmos **trabalham entre si para encontrar o melhor caminho para a saída e no final é realizada uma votação (classificação) ou média (regressão) para obter o melhor valor entre os algoritmos que atuaram em paralelo.**
- **É um dos mais precisos em aprendizado de máquina.**
- **É utilizado para a seleção de variáveis e criação do modelo preditivo.**
- **Pode ser utilizado tanto para classificação (atributo alvo categórico) como para regressão (atributo alvo contínuo).**

Aprendizado Supervisionado

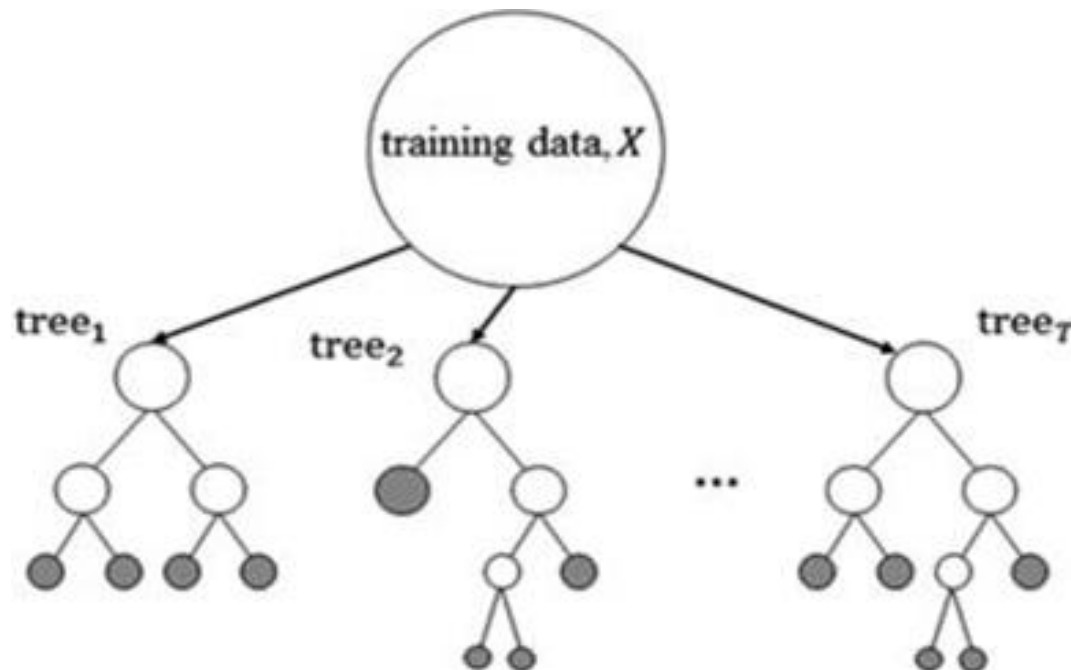
Algoritmo - Métodos do tipo ensemble - ***Random Forest*** (Floresta Aleatória)

- Para melhor compreensão do algoritmo Random Forest, temos as **etapas (simplificadas)**:
 1. **Escolha k pontos de dados aleatórios** (amostra com K valores) **do conjunto de treinamento.**
 2. **Construa uma árvore de decisão associada a esses k pontos de dados.**
 3. **Escolha o número N de árvores que deseja construir e repita as etapas 1 e 2.**
 4. **Para um novo ponto de dados, faça com que cada uma de suas árvores N -tree prevejam o valor (classe/valor contínuo) para o ponto de dados em questão e atribua o novo ponto de dados à média de todos os valores de y previstos (regressão) ou encontre a classe que obteve maior incidência (classificação).**

Aprendizado Supervisionado

Algoritmo - Métodos do tipo ensemble - ***Random Forest*** (Floresta Aleatória)

- Diagrama conceitual.

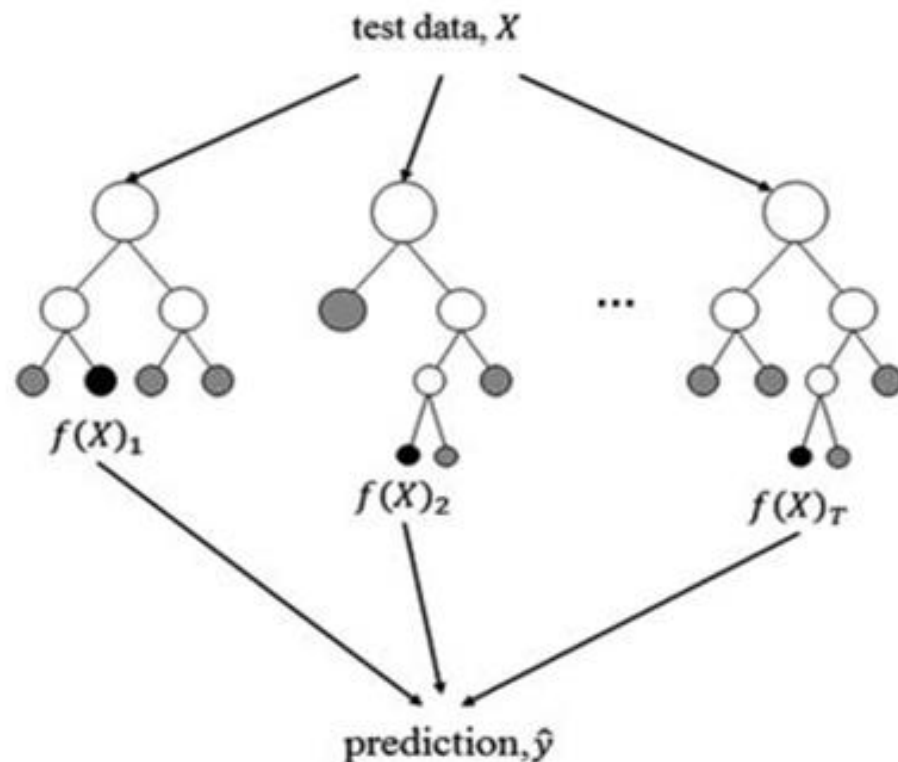


- **As árvores são treinadas de forma independente** por particionamento binário recursivo de uma amostra dos dados de entrada, X .

Aprendizado Supervisionado

Algoritmo - Métodos do tipo ensemble - **Random Forest (Floresta Aleatória)**

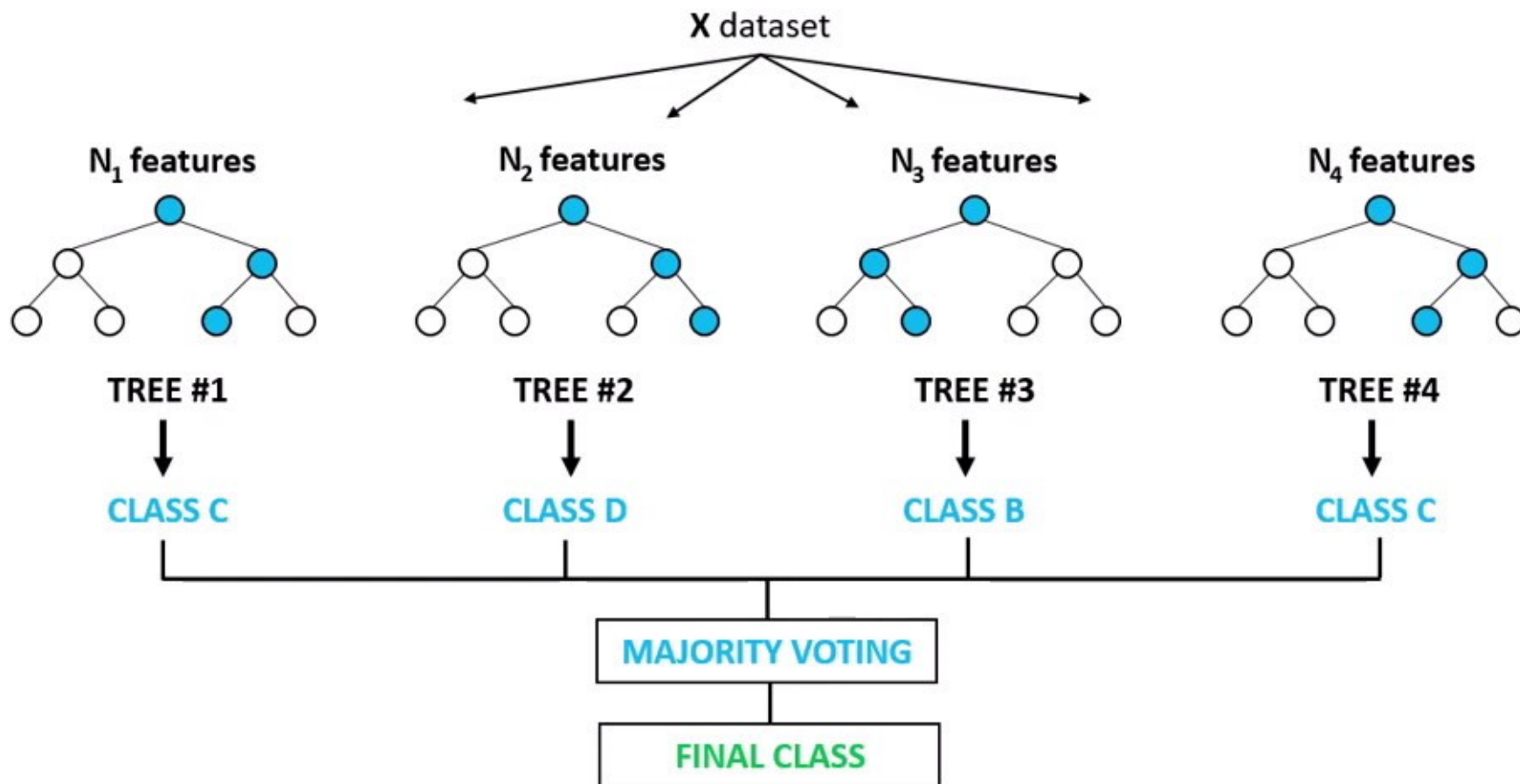
- Diagrama conceitual.



- Os dados de teste são inseridos em cada árvore e a estimativa de resposta é a média de todas as previsões individuais ou uma votação das previsões na floresta.

Aprendizado Supervisionado

Algoritmo - Métodos do tipo ensemble - **Random Forest** (Floresta Aleatória) – Classificação – Exemplo



Aprendizado Supervisionado

Algoritmo - Métodos do tipo ensemble - *Random Forest* (Floresta Aleatória) – Prós e Contras (1)

- **Prós:**
 - É poderoso e preciso.
 - Tem um ótimo desempenho em muitos problemas, incluindo atributos (variáveis) com relacionamentos não lineares.
 - Hiperparâmetros fáceis de entender e não são muitos.
- **Contras:**
 - Não muito fáceis de interpretar estatisticamente.
 - *Overfitting* pode facilmente ocorrer.
 - Deve-se escolher o número de árvores a incluir no modelo.

Aprendizado Supervisionado

Algoritmo - Métodos do tipo ensemble – **Abordagens Principais - *Boosting***

- ***Boosting***: é uma técnica que faz uso de algoritmos do mesmo tipo, porém cada modelo tenta corrigir os erros do modelo anterior.
- É mais ou menos como se fosse uma linha de produção (modelos sequenciais), ou seja, um **algoritmo é treinado e gera o modelo**.
- **Esse modelo vai ter uma taxa de erros. Essa taxa de erros vai ser corrigida pelo próximo algoritmo.**
- **Que então, vai ser treinado, criar um outro modelo e então obter outra taxa de erros, na qual espera-se seja menor.**
- **Isso vai alimentar um outro algoritmo que segue o mesmo procedimento descrito anteriormente.**

Aprendizado Supervisionado

Algoritmo - Métodos do tipo ensemble – **Abordagens Principais - *Boosting***

- Assim, no final, **o modelo obtido é aquele que foi corrigindo os erros dos modelos anteriores.**
- Para isso funcionar todos os algoritmos tem que ser do mesmo tipo.
- O tempo de treinamento é bem maior e em geral a capacidade computacional também.
- Exemplos:
 - C5.0
 - *Stochastic Gradient Boosting* (GBM)
 - *eXtreme Gradient Boosting* (XGBoost)
 - **AdaBoost.**

Aprendizado Supervisionado

Algoritmo - Métodos do tipo ensemble - **AdaBoost**

- Utiliza a **abordagem do tipo boosting**, ou seja, faz uso de **algoritmos do mesmo tipo (por exemplo, árvores de decisão do tipo CART)**, porém cada modelo gerado tenta corrigir os erros do modelo anterior.
- Em geral, **requer um pouco mais de capacidade computacional**, mas consegue obter uma **precisão muito alta**.
- **Um dos algoritmos mais utilizados pelos participantes do Kaggle (competição de Ciência de Dados)**, pois **apresenta uma alta precisão**, porém é mais difícil de treinar.
- **Pode ser utilizado para classificação ou regressão.**

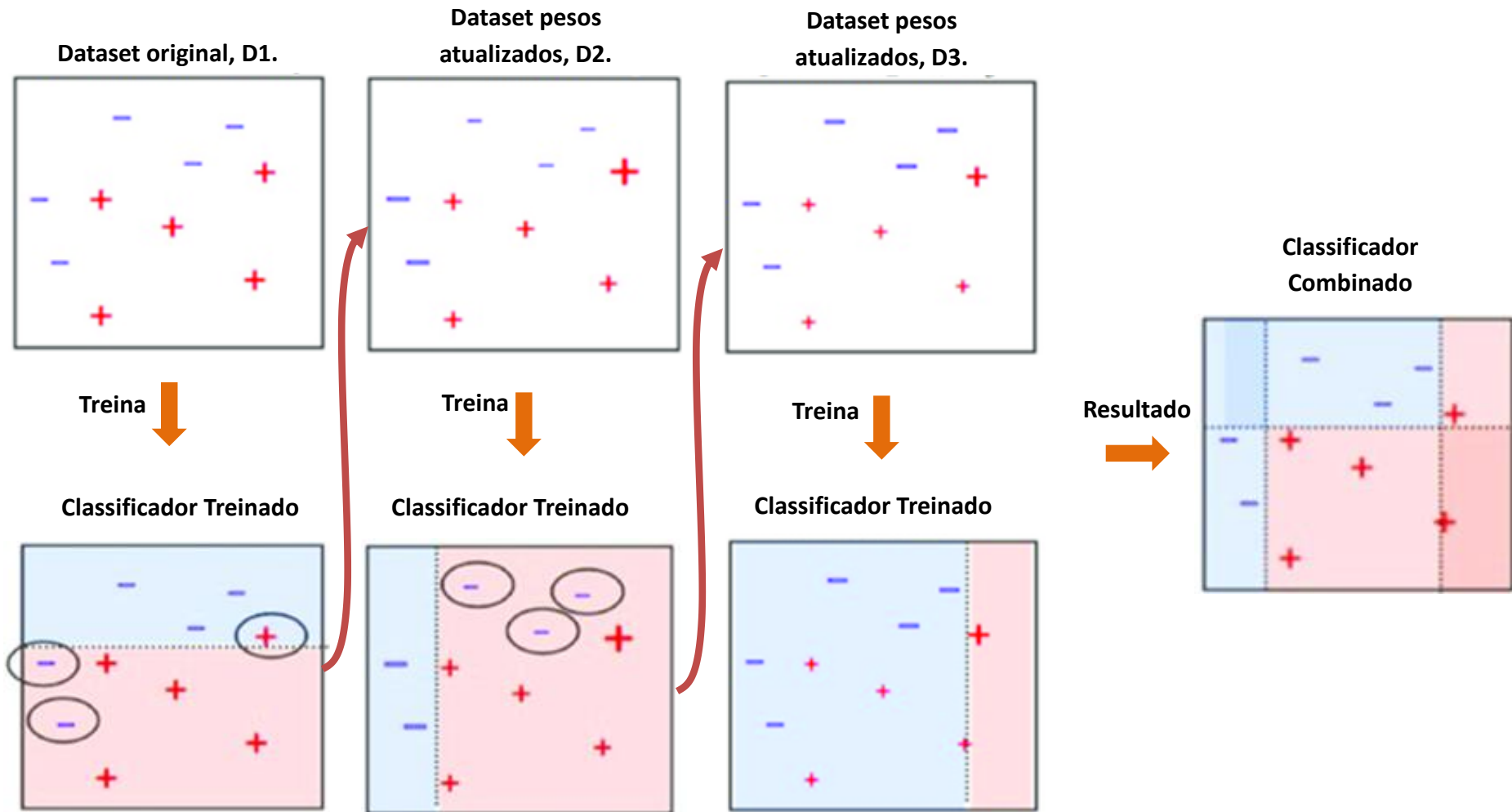
Aprendizado Supervisionado

Algoritmo - Métodos do tipo ensemble - **AdaBoost** – Funcionamento (1)

1. **Seleciona um subconjunto de treinamento aleatoriamente.**
2. **Ele treina iterativamente o modelo selecionando o conjunto de treinamento com base na taxa de acertos do último treinamento.**
3. **Ele atribui o peso mais alto às observações realizadas incorretamente, de modo que na próxima iteração essas observações obtenham a alta probabilidade de acertos.**
4. **Além disso, atribui o peso ao estimador treinado em cada iteração de acordo com a taxa de acertos do modelo. O estimador mais preciso terá alto peso.**
5. **Este processo itera até que os dados de treinamento completos se ajustem sem nenhum erro ou até que seja alcançado o número máximo especificado de estimadores (modelos preditivos).**
6. **Para encontrar o novo valor, execute uma “Combinação” dentre todos os modelos construídos.**

Aprendizado Supervisionado

Algoritmo - Métodos do tipo ensemble - **AdaBoost** – Funcionamento (1)



Aprendizado Supervisionado

Algoritmo - Métodos do tipo ensemble - **AdaBoost – Prós (1)**

1. É fácil de implementar.
2. Corrige iterativamente os erros do classificador fraco e melhora a precisão combinando estimadores fracos.
3. Pode usar muitos classificadores básicos (por exemplo, CART).
4. Não é propenso a *overfitting*.

Aprendizado Supervisionado

Algoritmo - Métodos do tipo ensemble - AdaBoost –Contras (1)

1. É sensível a dados de ruído.
2. É altamente afetado por *outliers* (valores extremos) porque tenta ajustar cada ponto perfeitamente.
3. AdaBoost é mais lento em comparação com XGBoost.

Aprendizado Supervisionado

Algoritmo - Métodos do tipo ensemble – **Abordagens Principais - Votação**

- **Votação:** quando é feito o uso de **algoritmos diferentes treinados em paralelo e ao final faz-se uma votação para escolher qual daqueles algoritmos apresentou um modelo com melhor desempenho.**
- **Exemplos:**
 - *Coombs' method*
 - *Boyer–Moore majority vote algorithm*
 - *Nanson's method*

Aprendizado Supervisionado

Algoritmo - Métodos do tipo ensemble - **Discussão**

- **Método do tipo ensemble** faz a **união das saídas de diferentes modelos para encontrar a melhor resposta para o problema.**
- Métodos do tipo ensemble são **utilizados quando precisamos de uma acurácia (taxa de acertos) muito alta!**
- O que é **comum em competições!**
- Porém, **dependendo do problema de negócio, não é somente os acertos (acurácia) que devem ser avaliados**, há outras questões que também devem ser consideradas.
- Cuidado! Isso não é a solução para todos os problemas! **Em geral, será necessário muito mais tempo de treinamento e capacidade computacional.**

Aprendizado Supervisionado

Algoritmo - **Fatores a considerar na escolha do algoritmo ideal**



Precisão

Tipo de Problema a
ser Resolvido



Tempo de Treinamento

Linearidade



Número de Recursos

Número de Parâmetros



Aprendizado Supervisionado

Algoritmo - Fatores a considerar na escolha do algoritmo ideal

- **Precisão:** é desejável um algoritmo que acerte na maioria das vezes, ou seja, maior precisão. Mas, melhor precisão, pode gerar *overffiting* (aprende demais - sobreajustamento), por exemplo, se o conjunto de dados for muito pequeno.
- **Tipo de Problema:** Classificação , Regressão, Clusterização, etc.
- **Tempo de treinamento:** deve ser definido o tempo disponível para gerar um resultado aceitável.
Por exemplo, dependendo do conjunto de dados pode levar duas semanas, algoritmos complexos.
Google realizou um treinamento de PLN (Processamento de Linguagem Natural) com 1 trilhão de registros e levou 3 semanas.
Neste caso, podem ser necessárias milhares de máquinas para processar e armazenar os dados.
- **Linearidade:** os dados podem ser linearmente separáveis ou não.







Aprendizado Supervisionado

Algoritmo - Fatores a considerar na escolha do algoritmo ideal

- **Números de parâmetros:** cada algoritmo tem uma série de parâmetros (hiperparâmetros) a sua disposição. É necessário verificar até que ponto isso pode complicar ou aumentar o tempo de treinamento do modelo, consequentemente o tempo do projeto.
- **Número de recursos:** número de variáveis ou atributos no seu conjunto de dados (ex.: colunas numa planilha do Excel).
Por exemplo, *dataset* com 400 variáveis (atributos preditores), talvez seja necessário fazer uma redução da dimensionalidade, utilizando, por exemplo, o algoritmo PCA (*Principal Components Analysis*).
Pode-se usar um algoritmo para seleção de variáveis, depois um algoritmo para treinar o modelo final. Ou, pode-se usar um algoritmo para redução da dimensionalidade, depois um algoritmo para treinar o modelo final, e assim por diante.
Então, o número de variáveis também impacta na escolha do melhor algoritmo.

Aprendizado Supervisionado

Algoritmo - Comparação entre os principais algoritmos - Classificação binária (2 classes possíveis).

Algoritmo	Tempo de Treinamento	Precisão	Linearidade
Regressão Logística			
Árvore de Decisão			N/A
Random Forest			N/A
Redes Neurais			N/A
SVM			
Métodos Bayesianos			



Alto



Moderado

Aprendizado Supervisionado

Algoritmo - Comparação entre os principais algoritmos - Classificação multiclasse (mais de 2 classes possíveis).



Algoritmo	Tempo de Treinamento	Precisão	Linearidade
Regressão Logística			
Árvore de Decisão			N/A
Random Forest			N/A
Redes Neurais			N/A
SVM			

Bibliografia

BÁSICA:

- AGGARWAL, Charu C. **Artificial Intelligence: A Textbook**. New York: Springer: 2021.
- CHOLLET, François. **Deep Learning with Python, 2ed**. Shelter Island: Manning, 2021.
- GÉRON, Aurélien. **Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems**, 2 ed. Sebastopol: O'Reilly, 2019.

COMPLEMENTAR:

- GOODFELLOW, Ian; BENGIO, Yoshua, COURVILLE, Aaron. **Deep Learning**. Cambridge: MIT Press, 2016.
- RASCHKA, Sebastian; MIRJALILI, Vahid. **Python Machine Learning**. 3 ed. Birmingham: Packt, 2017.
- RUSSEL, Stuart; NORVIG, Peter. **Artificial Intelligence: A Modern Approach**. 3 ed. Upper Saddle River: Pearson, 2010.
- TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introduction to Data Mining**. 2 ed. Upper Saddle River: Pearson, 2018.
- VANDERPLAS, Jake. **Python Data Science Handbook**. Sebastopol: O'Reilly, 2017.

ADICIONAIS:

- FACELI, Katti et al. **Inteligência artificial: uma abordagem de aprendizado de máquina**. 2ª Ed. Rio de Janeiro: LTC- Livros Técnicos e Científicos, 2021.
- LUGER, George F. **Inteligência Artificial** - 6ª ed. São Paulo: Pearson Education do Brasil, 2015.
- RUSSELL, Stuart; NORVIG, Peter. **Inteligência artificial: Uma Abordagem Moderna** - 4ª. Ed. GEN LTC, 2022.