

1 – Conceitos Básicos da IA do Microsoft Azure

1.2 – Princípios Base da Machine Learning

PREPARATÓRIO AI 900 DE A-Z

O QUE É APRENDIZADO DE MÁQUINA?

- ✖ Origem na Estatística e Matemática, ideia de **utilizar dados históricos** para **prever resultados** ou **valores desconhecidos**.

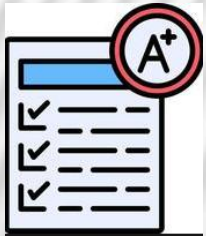


Prever diabetes

peso

glicose

...



Prever nota

faltas

qt_horas_estudos

pratica_atividade_fisica

qt_horas_sono



Prever qtde sorvetes

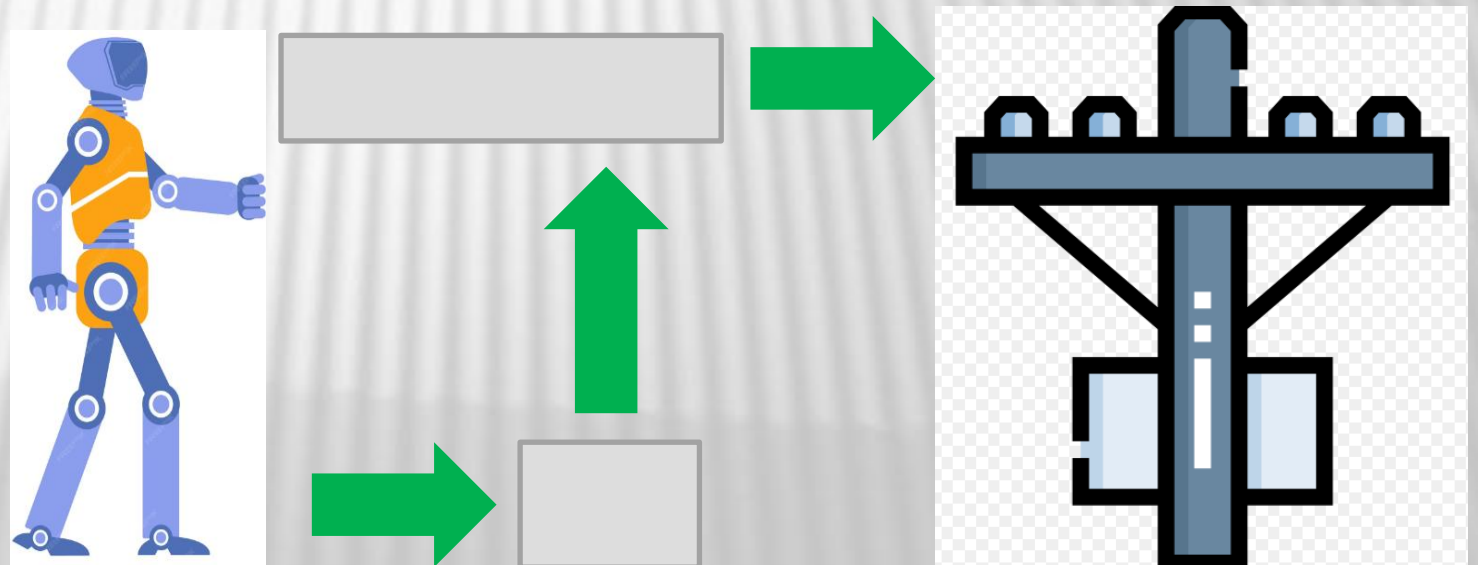
tempC

dia_semana

e_feriado

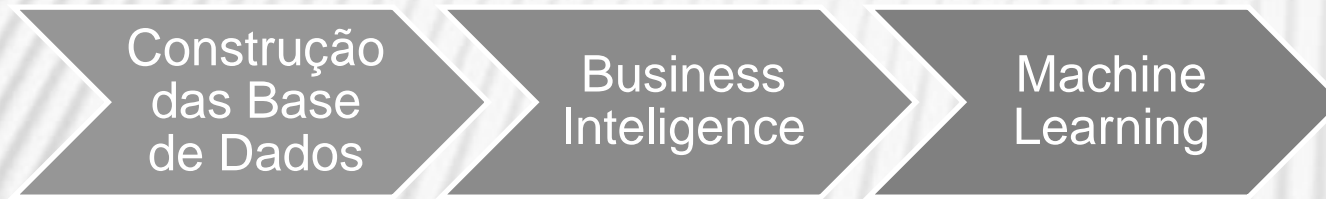
O QUE É APRENDIZADO DE MÁQUINA?

- ✖ Observar os **dados**, descobrir **padrões**, computador **aprender sem intervenção humana**.
- ✖ Não é oráculo (mágica).
- ✖ Ocorre um aprendizado de fato, sem explicitamente ser programado.
- ✖ Exemplo:



O QUE É APRENDIZADO DE MÁQUINA?

- ✖ E quando não tem a matéria-prima?
- ✖ Etapas:



O QUE É APRENDIZADO DE MÁQUINA?

- ✖ Um modelo preditivo é uma função matemática, que recebe valores de entrada (x) e retorna uma saída (y)

$$y = f(x)$$

- ✖ Exemplo:

Treinamento

Idade	Renda	Quantidade de produtos	Possui imóvel próprio?	Cartão liberado
30	6.000	4	sim	sim
40	10.000	2	sim	sim
50	1.800	2	não	não

X (recursos, features, variáveis preditoras)
[x1, x2, x3, x4 ...]

Y (rótulo, target, alvo)

Previsões(inferência)

Idade	Renda	Quantidade de produtos	Possui imóvel próprio?	Cartão liberado
40	1.000	1	não	?
38	1.400	1	não	?
18	2.000	2	não	?

X recursos

Y rótulo

O QUE É APRENDIZADO DE MÁQUINA?

- ✖ A complexidade da fórmula depende do algoritmo utilizado. Existem desde mais simples até mais avançados. Regressão linear é um exemplo de algoritmo.
- ✖ Exemplo de regressão linear múltipla.
 - ✖ $y = F(x)$
 - ✖ $F(x) = a * X1 + b * X2 + c$
- ✖ Exemplo de regressão linear simples.
 - ✖ $y = F2(x)$
 - ✖ $F2(x) = a * X1 + c$

O QUE É APRENDIZADO DE MÁQUINA?

- ❖ Exemplo de regressão linear múltipla. Prever a idade em meses de um bebê usando o peso e altura.

- ❖ $y = F2(x)$

- ❖ $F2(x) = a * X1 + b * X2 + c$

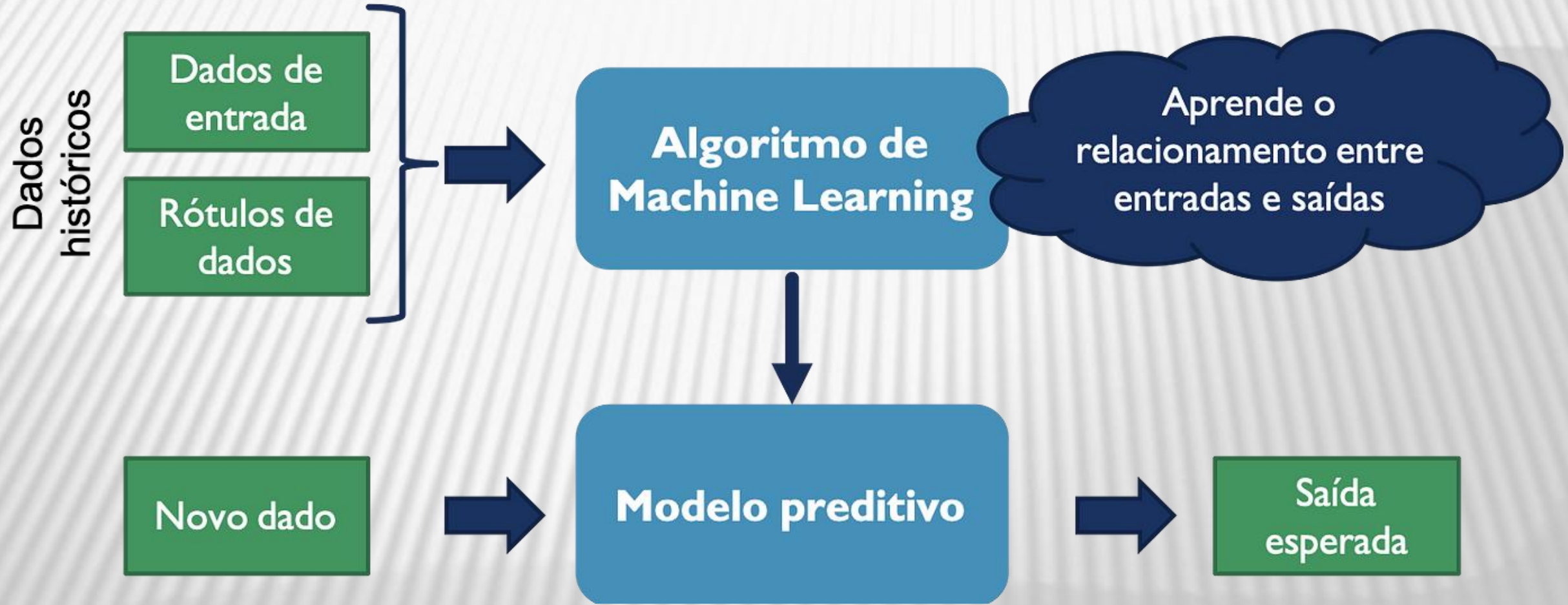
Peso	Altura	Idade
23,17 kg	130 cm	9
24 kg	128 cm	10
24 kg	126 cm	11

Treinamento

Peso	Altura	Idade
23 kg	122 cm	?
22 kg	120 cm	?
21 kg	90 cm	?

Inferências

O QUE É APRENDIZADO DE MÁQUINA?



ALGORITMOS DE MACHINE LEARNING

- ✖ Cada um tem prós x contras. Precisão x Tempo de Treinamento.

- ✖ Exemplos:

Decision Tree

Random Forest

SVM

Regressão Linear

Rede Neural

Regressão Logística

Rede Neural
Profunda (Deep
Learning)

- ✖ Pacotes para ML:



TIPOS DE APRENDIZADO DE MÁQUINA



TIPOS DE APRENDIZADO DE MÁQUINA

- × Supervisionado. Classificação binária. Ex 1.

Nome	Idade	Renda	Quantidade de produtos	Possui imóvel próprio?	Cartão liberado
Carlos	30	6.000	4	sim	sim
Ana	40	10.000	2	sim	sim
José	50	1.800	2	não	não

Variáveis
preditoras

Label/target/rótulo



TIPOS DE APRENDIZADO DE MÁQUINA

- × Supervisionado. Classificação binária. Ex 1.

Nome	Idade	Renda	Quantidade de produtos	Possui imóvel próprio?	Cartão liberado
Carlos	30	6.000	4	sim	1
Ana	40	10.000	2	sim	1
José	50	1.800	2	não	0

Variáveis
preditoras

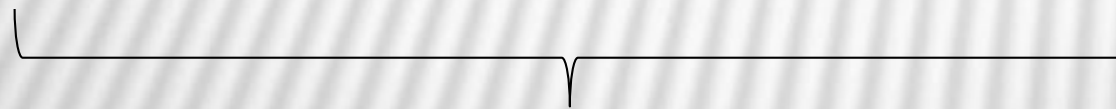
Label/target/rótulo



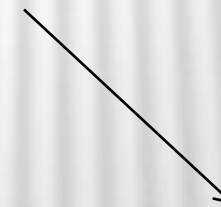
TIPOS DE APRENDIZADO DE MÁQUINA

- ✖ Supervisionado. Classificação multiclasse. Ex 2.

nome	HP	Ataque	Defesa	tipo1
Bulbasaur	318	45	49	Grama
Charmander	309	39	52	Fogo
Squirtle	314	44	48	Água



Variáveis
preditoras



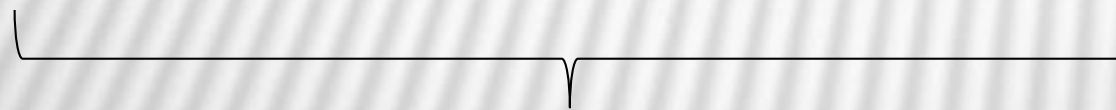
Label/target/rótulo

18 tipos, exemplo: grama, normal, fogo, água, fantasma ...

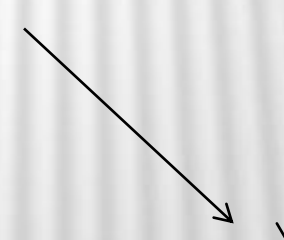
TIPOS DE APRENDIZADO DE MÁQUINA

- ✖ Supervisionado. Classificação multilabel. Ex 2.

nome	HP	Ataque	Defesa	tipo1	tipo2
Bulbasaur	318	45	49	Grama	Venenoso
Charmander	309	39	52	Fogo	-
Squirtle	314	44	48	Água	



Variáveis
preditoras



Label/target/rótulo

18 tipos, exemplo: grama, normal, fogo, água, fantasma ...

TIPOS DE APRENDIZADO DE MÁQUINA

× Supervisionado. Regressão Ex 3

Local	Qtde quartos	Área	Tem garagem?	Quitada ?	Preço
Taguatinga	4	100 x 100	sim	sim	500.000
Taguatinga	3	80 x 60	sim	sim	350.000
Asa Sul	2	30 x 40	não	não	600.000

Variáveis
preditoras

Label/target



TIPOS DE APRENDIZADO DE MÁQUINA

× Exemplo 3. Não supervisionado. Clusterização

Nome	Idade	Renda	Quantidade de produtos	Possui imóvel próprio?
Carlos	30	6.000	4	sim
Ana	40	10.000	2	sim
José	50	1.800	2	não

Variáveis
preditoras

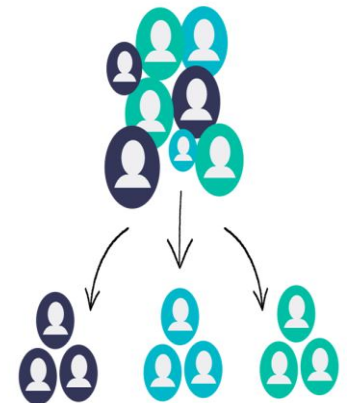
Não há TARGET/ ALVO/RÓTULO



TIPOS DE APRENDIZADO DE MÁQUINA

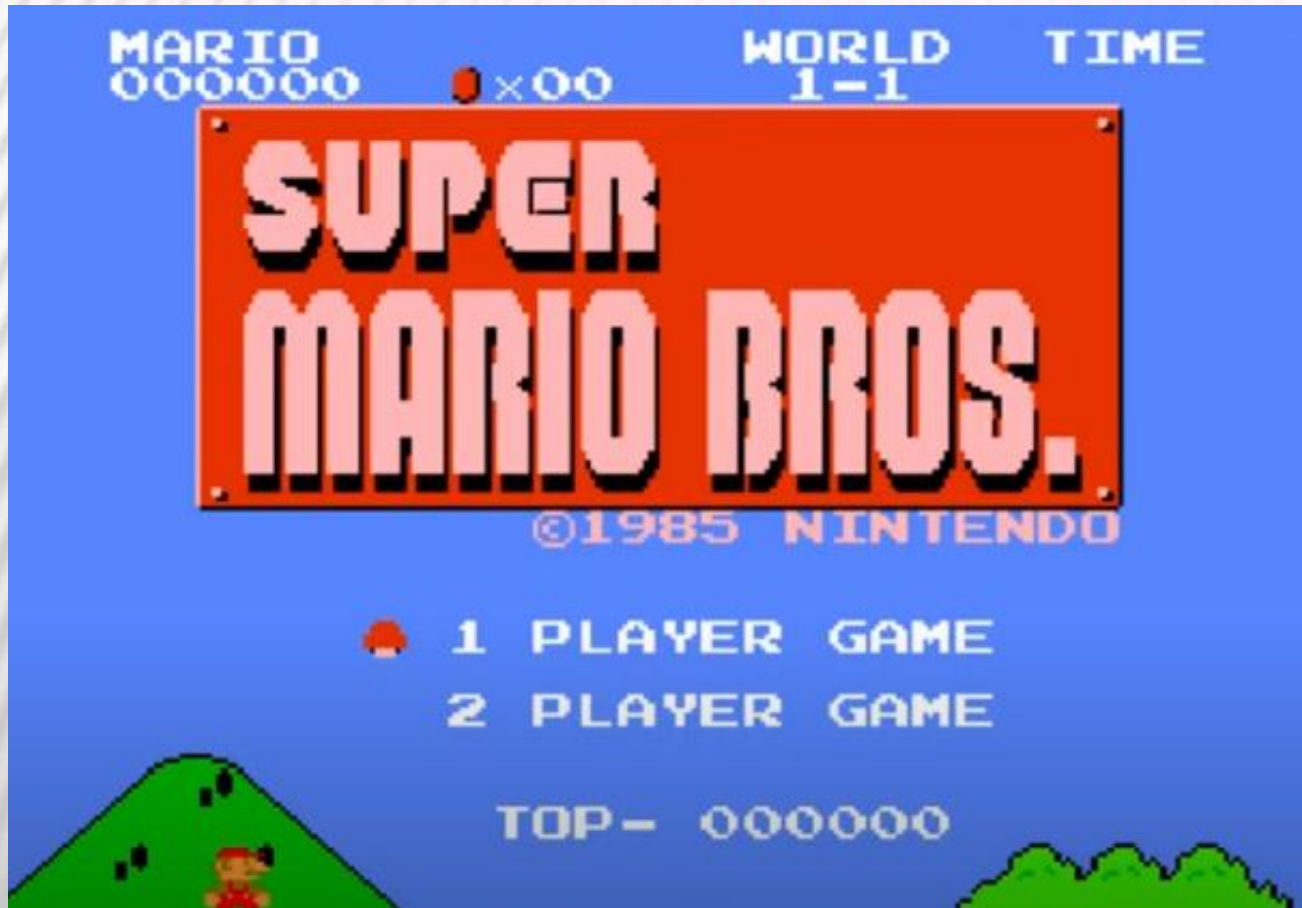
- ✖ Exemplo 3. Não supervisionado. Clusterização.
- ✖ Desafio: como avaliar os dados, se não há definição explícita do ‘certo e errado’?

Nome	Idade	Renda	Quantidade de produtos	Possui imóvel próprio?	Grupo (Cluster)
Carlos	30	6.000	4	sim	Grupo 2 (alta renda)
Ana	40	10.000	2	sim	Grupo 3 (renda média)
José	50	1.800	2	não	Grupo 1 (baixa renda)

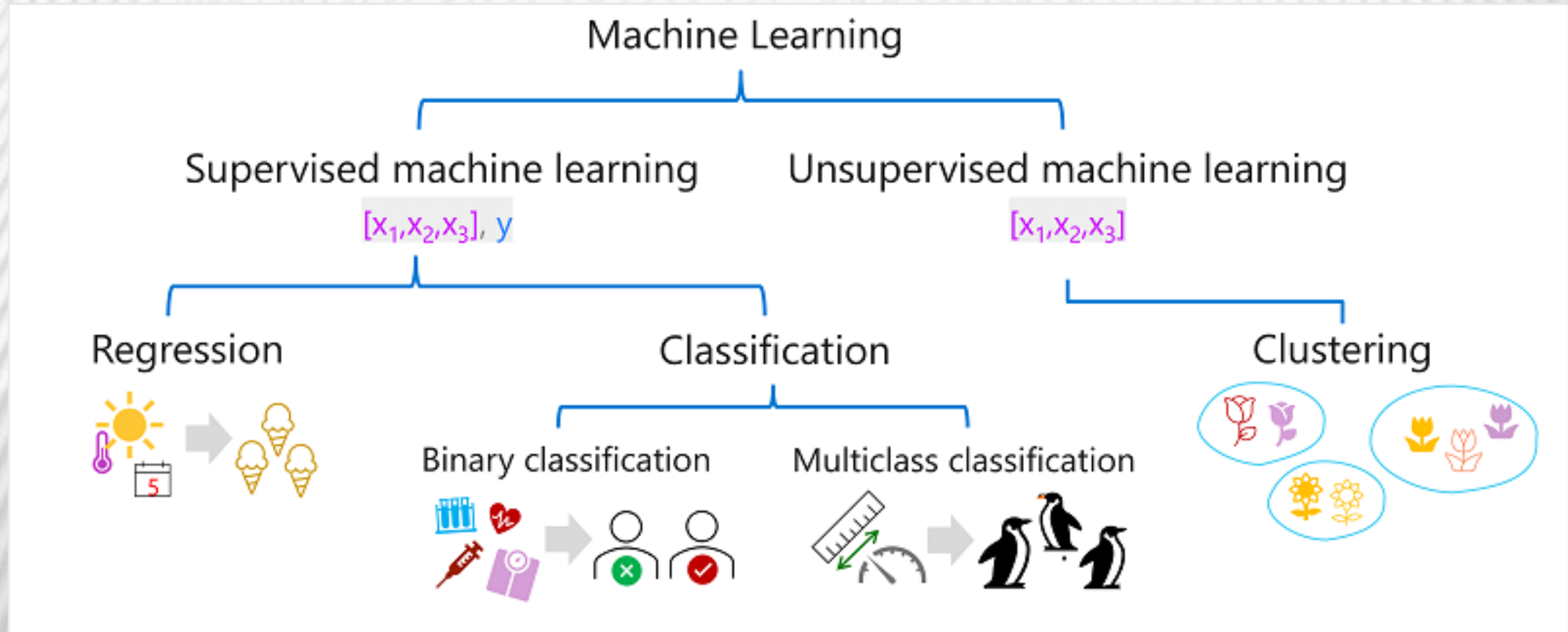


TIPOS DE APRENDIZADO DE MÁQUINA

- × Exemplo 4. Sistema de Recompensas, Tentativa e Erro.



TIPOS DE APRENDIZADO DE MÁQUINA



TIPOS DE APRENDIZADO DE MÁQUINA

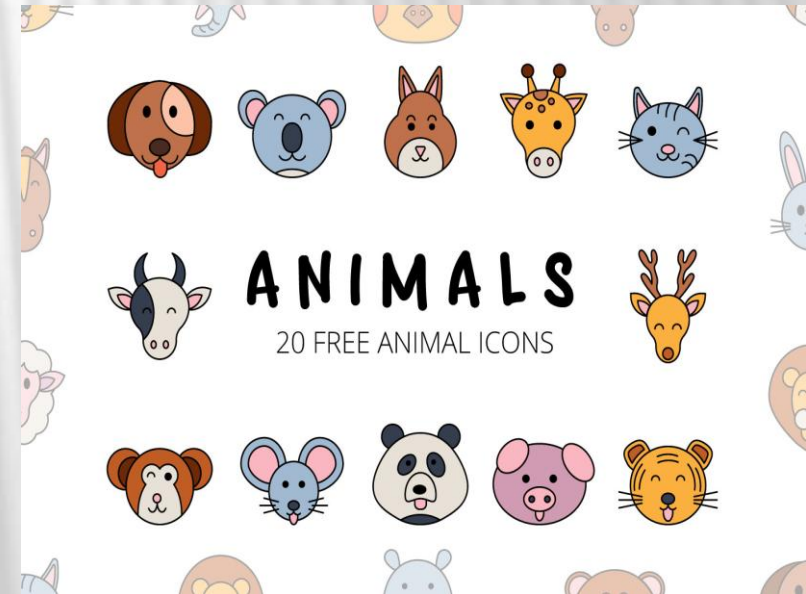
✖ Mais Exemplos.



Prever nota final de um aluno na matéria de Matemática do Professor Victor.



Prever a probabilidade de câncer um paciente.

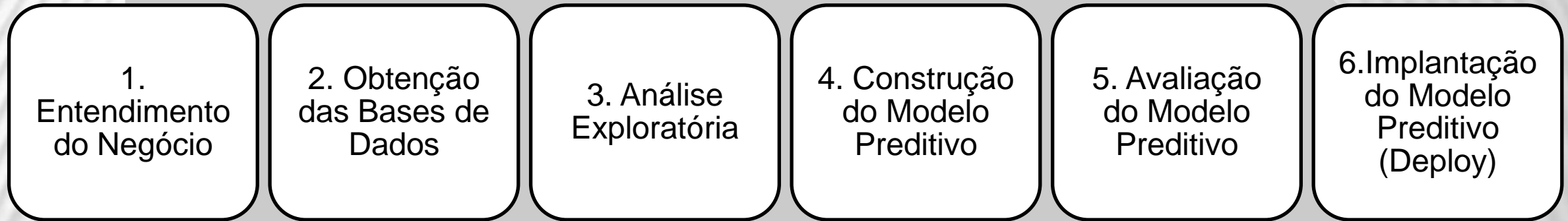


Prever qual é o animal através de reconhecimento por imagem.

PROCESSO DE CIÊNCIA DE DADOS

- Pontos Importantes.

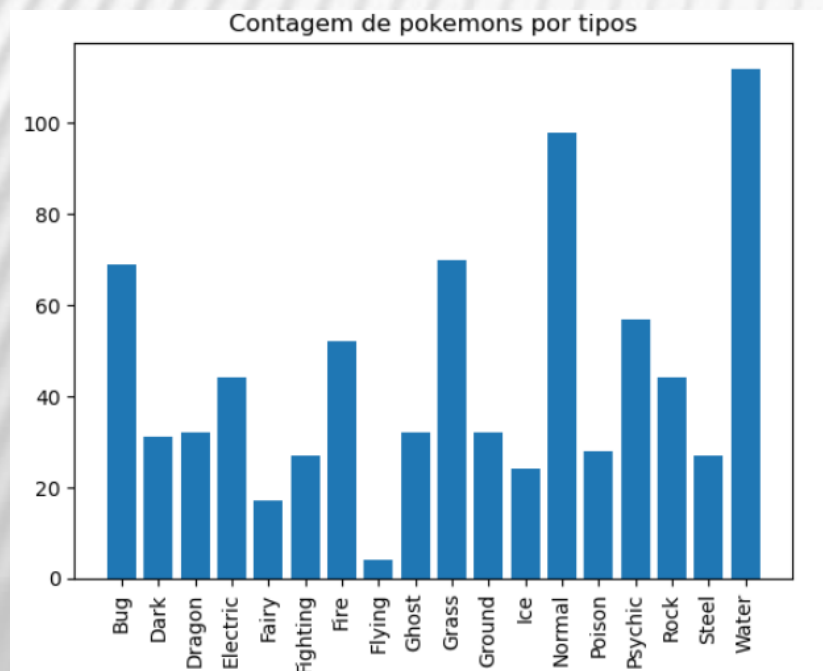
- 1 – Cerca de 80% ou mais é na Análise Exploratória.
- 2 – A construção do modelo é a mais ‘fácil’.
- 3 – Avaliação do modelo por meio de métricas (medidas).



PROCESSO DE CIÊNCIA DE DADOS

- Análise Exploratória de Dados

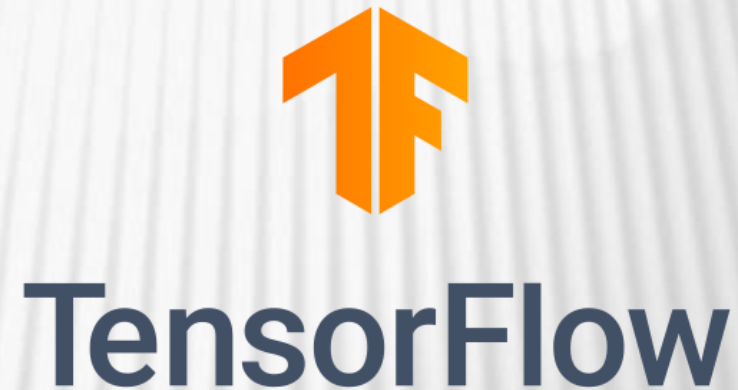
Nome	Tipo	Total	HP	Ataque	Defesa
Bulbasaur	Grass	318	45	49	49
Ivysaur	Grass	405	60	62	--
Venusaur	Grass	25	--	--	83



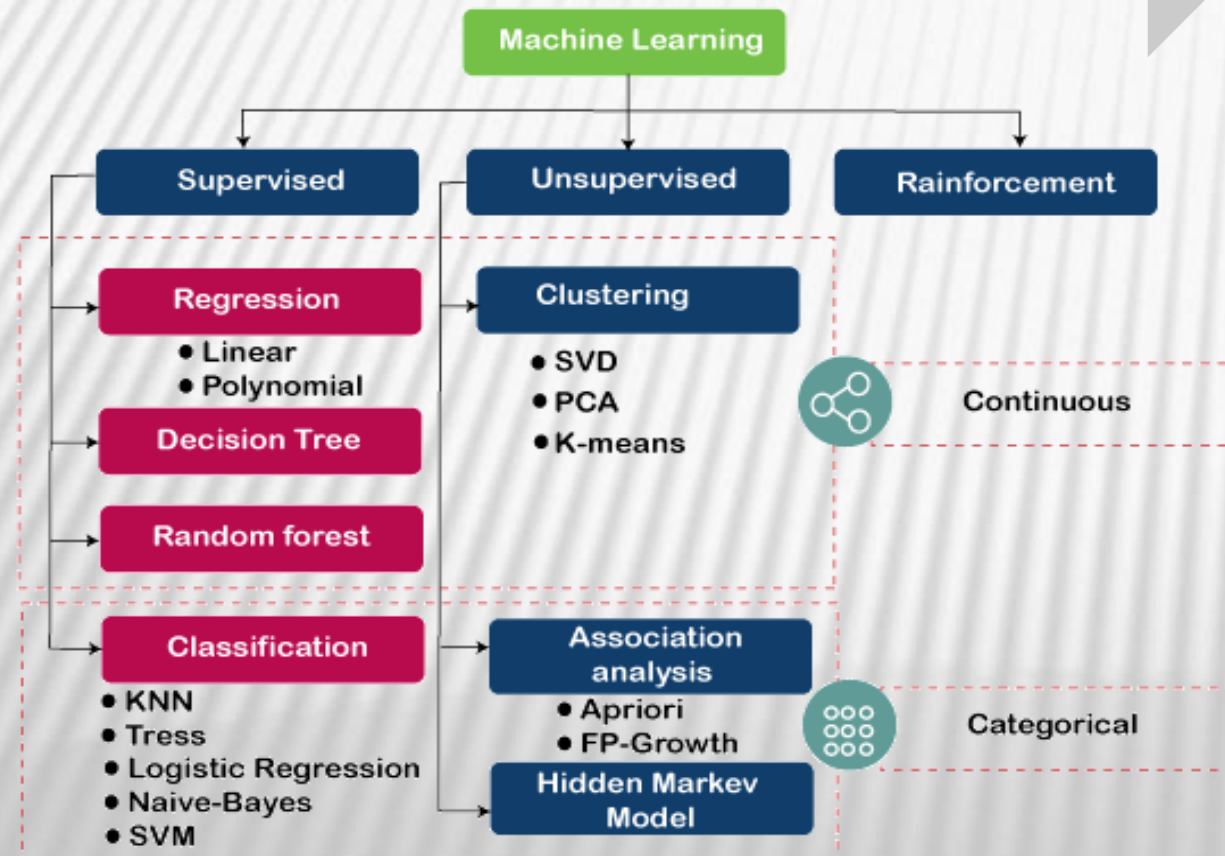
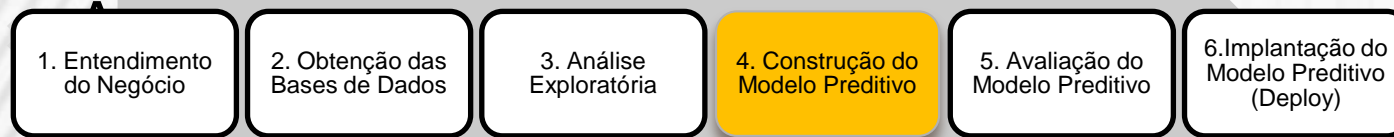
O que devemos fazer?
Como devemos tratar esse campo com o valor desconhecido, ausente (NULO)?

SOLUÇÕES DE MACHINE LEARNING

- × Soluções para Machine Learning



SOLUÇÕES DE MACHINE LEARNING

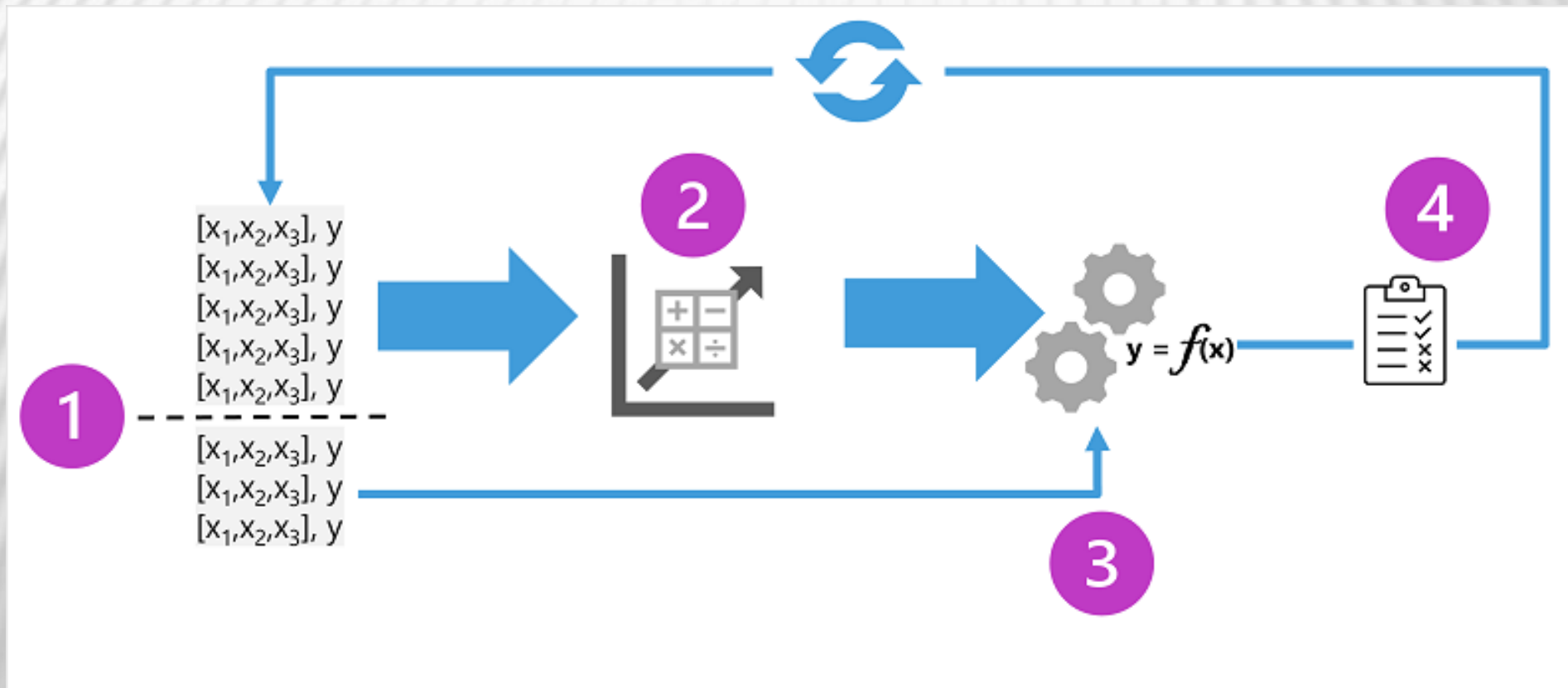


Exemplos:

1. Árvore de Decisão (Tree Decision)
2. Regressão Linear
3. Floresta Aleatória (Random Forest)
4. K-means
5. Rede Neural
6. Rede Neural Profunda

REGRESSÃO

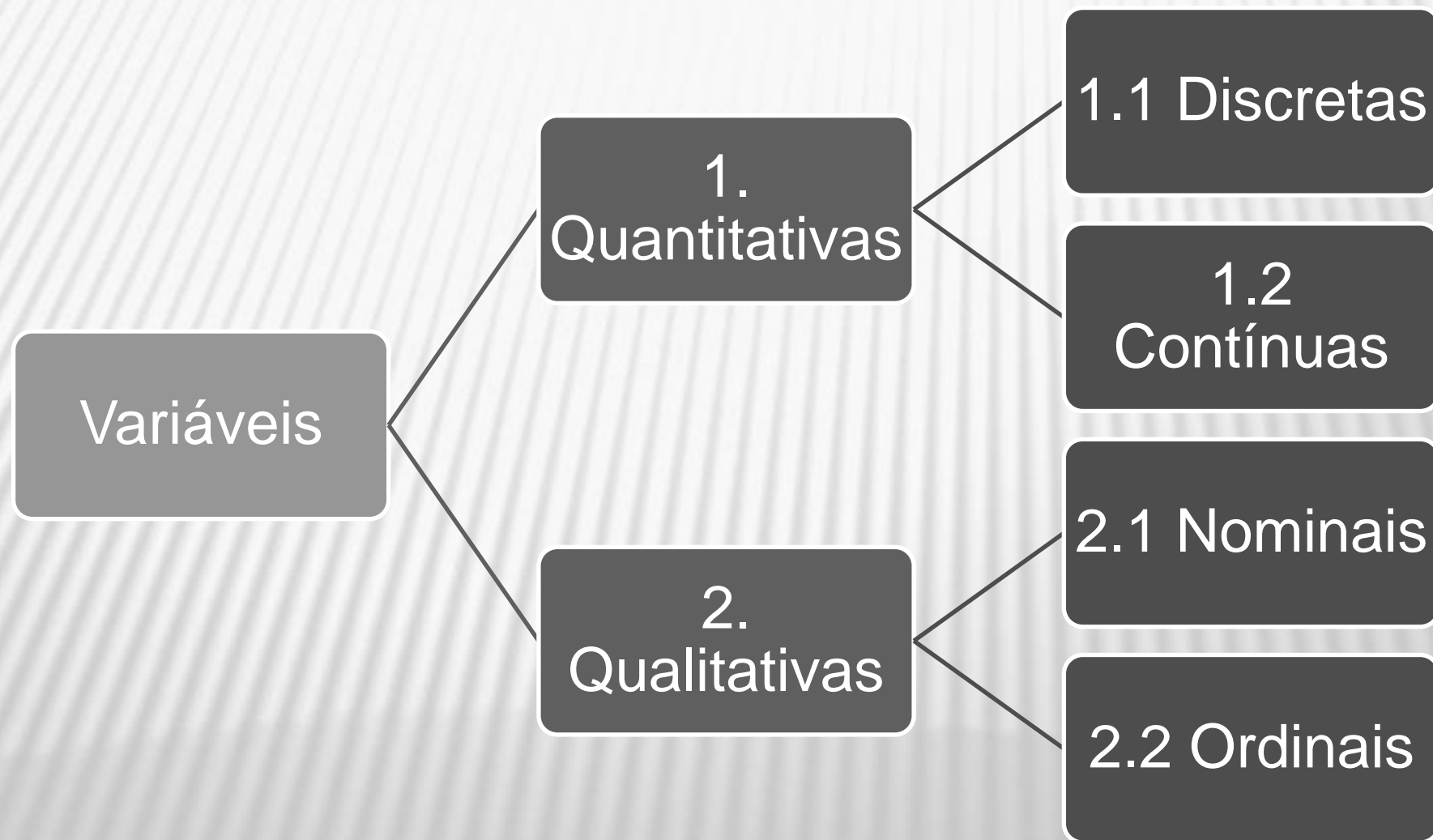
- ✖ Prever rótulos de valores numéricos. Na Estatística seriam as variáveis quantitativas (discretas ou contínuas)
- ✖ Processo de Treinamento.



Explicações.



1. Conjunto de treino/validação(Teste)
2. Aplicação de um algoritmo.
3. Com os dados, previsão de novos dados.
4. Comparação rótulos reais e previstos. Métrica que indica performance.

ESTATÍSTICA – TIPOS DE VARIÁVEIS.



REGRESSÃO

- ✖ Exemplo. Vendas de sorvetes.
- ✖ Recursos: temperatura (x). Rótulo: quantidade de sorvetes (y).

 Temperatura (x)	 Vendas de sorvete (y)
51	1
52	0
67	14
65	14
70	23
69	20
72	23
75	26
73	22
81	30
78	26
83	36

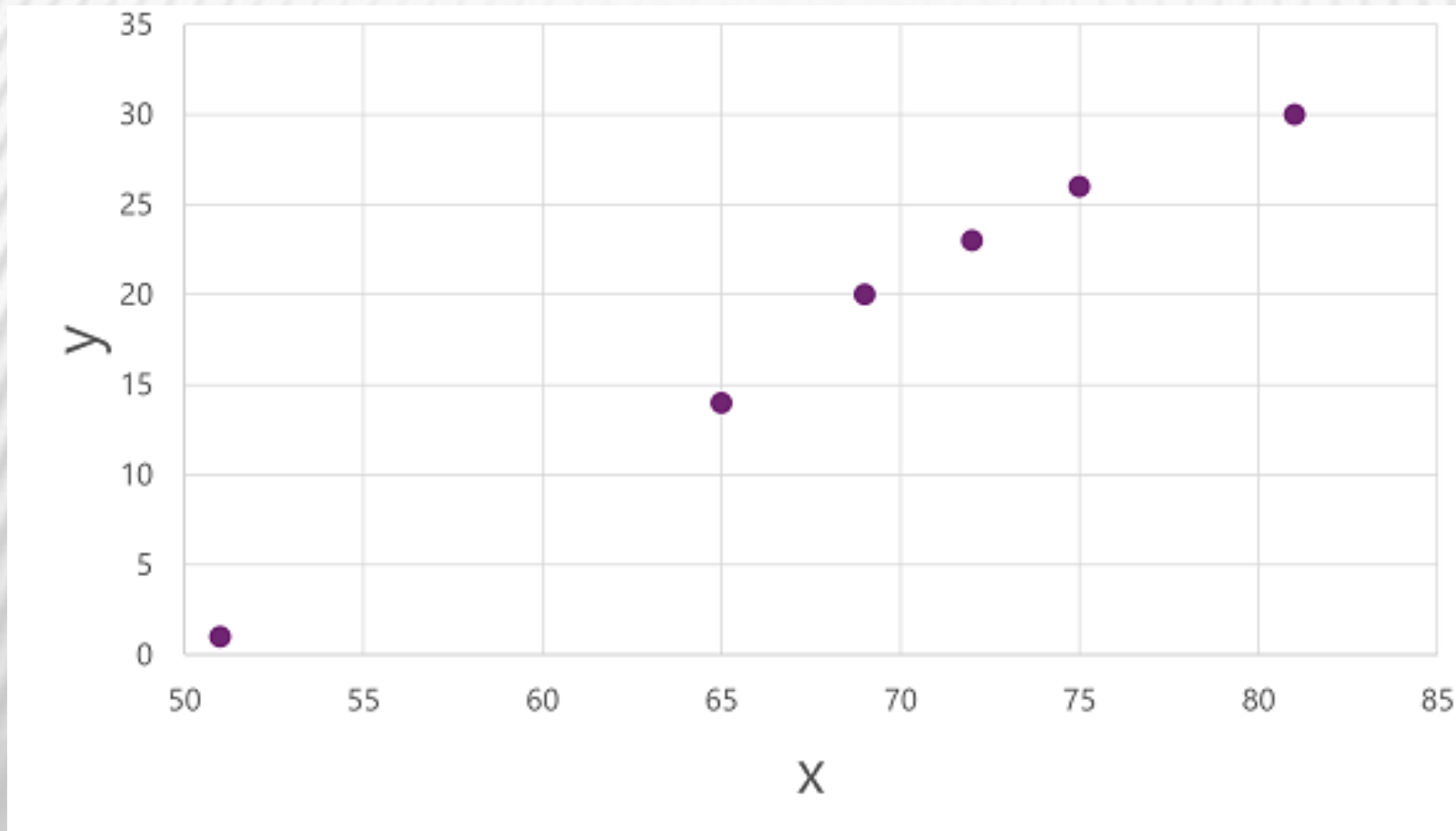
REGRESSÃO

✖ Conjunto de Treinamento.

Temperatura (x)	Vendas de sorvete (y)
51	1
65	14
69	20
72	23
75	26
81	30

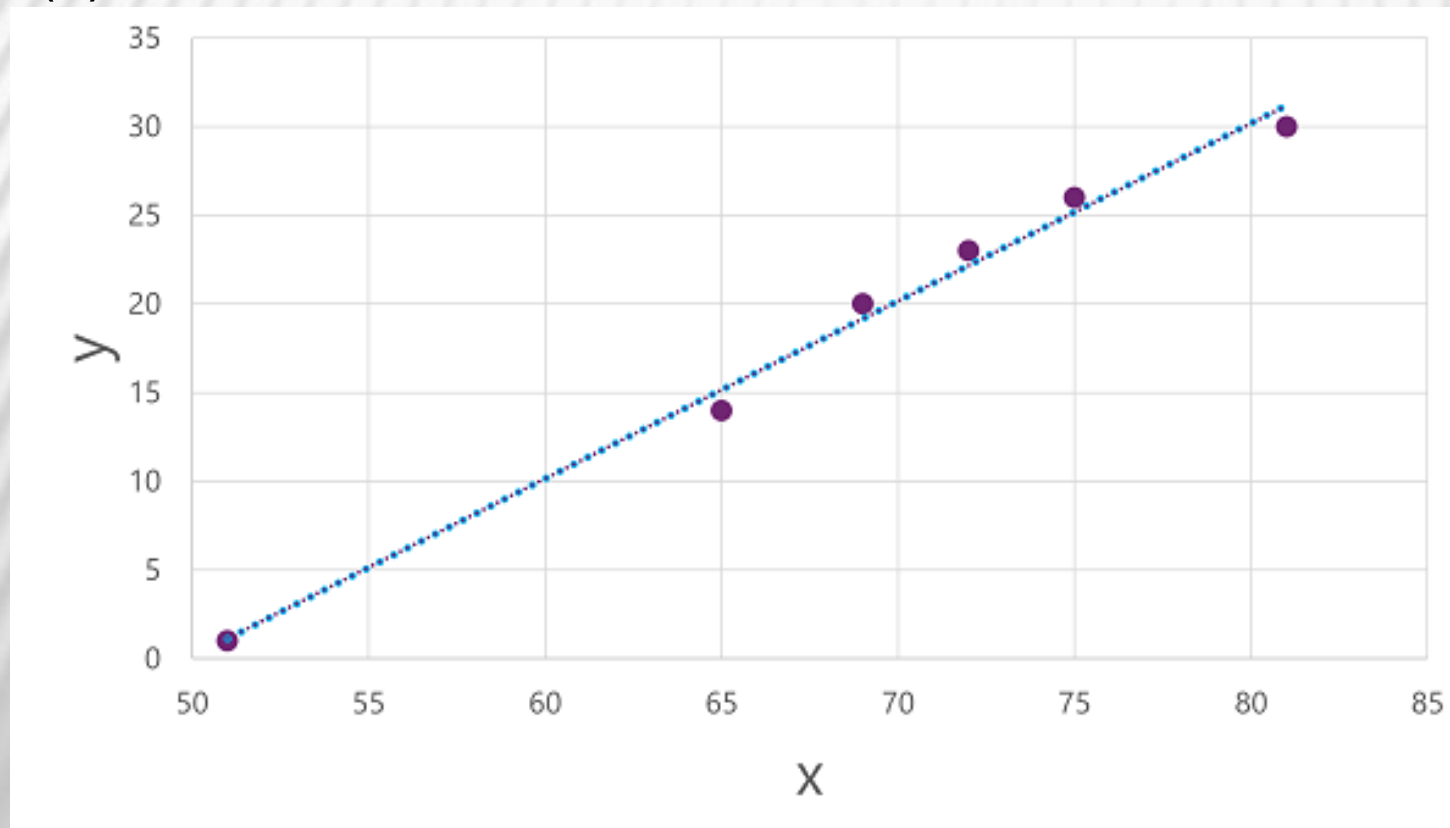
REGRESSÃO

- × Quantidade de Sorvetes Vendidas (y) e Temperatura (x).



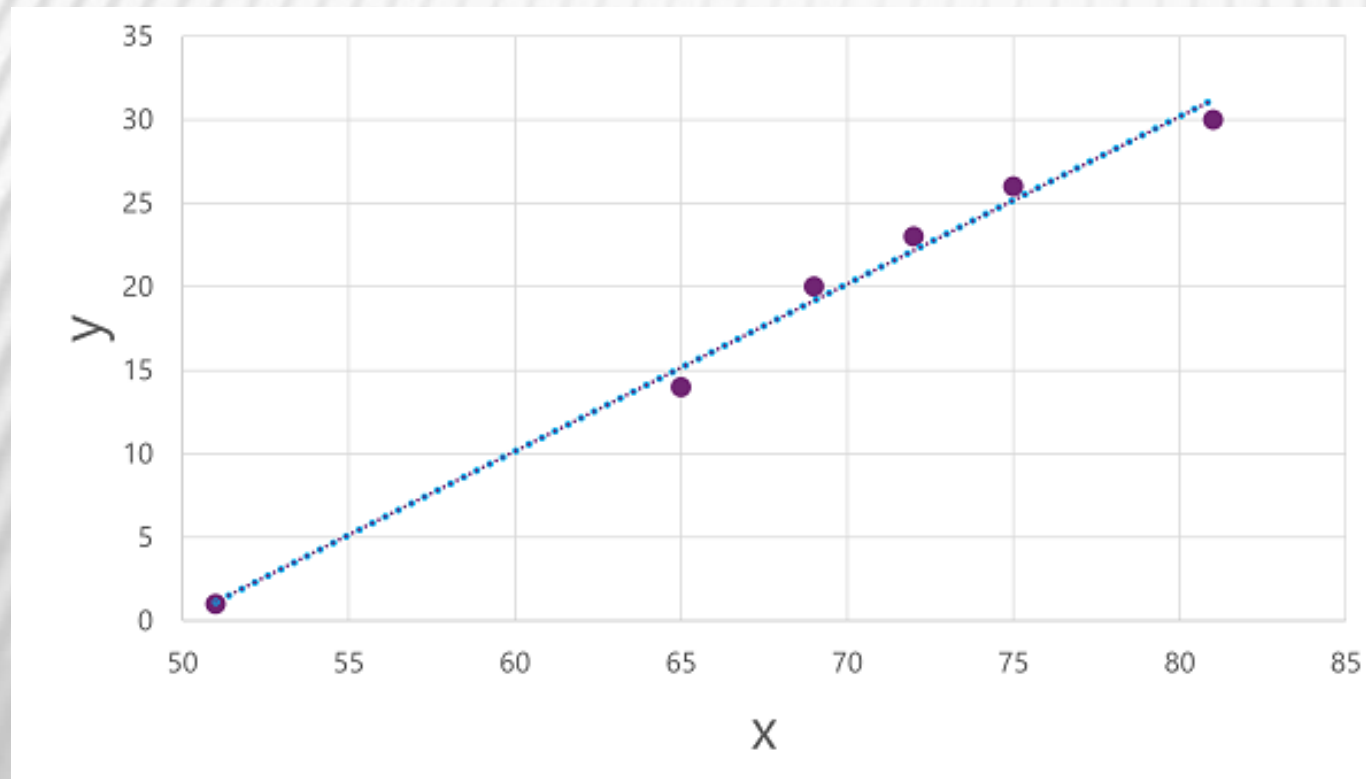
REGRESSÃO

- × Regressão Linear Simples. Função representada por uma reta. A linha intercepta o eixo em x em 50, quando $x = 50$, $y = 0$. Os demais pontos aumentam em 5.
- × $f(x) = x - 50$



REGRESSÃO

- ✖ $f(x) = x - 50$
- ✖ $x = 72?$
- ✖ $F(72) = 72 - 50$
- ✖ $F(72) = 22$



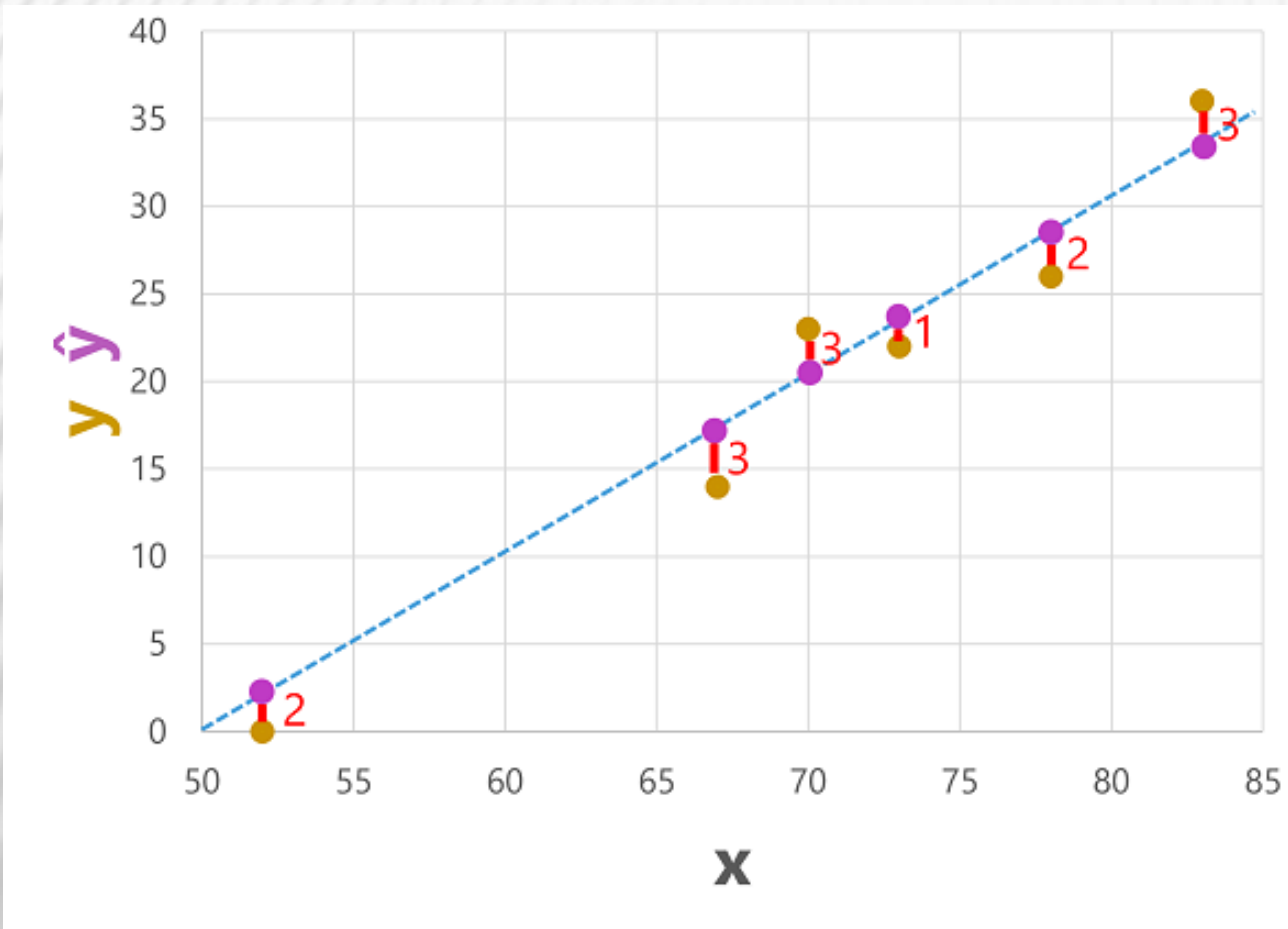
REGRESSÃO

- × Avaliação do modelo por meio de métricas.
- × Separação do conjunto para a avaliação.

Temperatura (x)	Vendas reais (y)	Vendas previstas (\hat{y})
52	0	2
67	14	17
70	23	20
73	22	23
78	26	28
83	36	33

REGRESSÃO

- × Visualização dos valores previstos, reais, reta.



REGRESSÃO

× 1. MAE (Erro Médio Absoluto)

- × Olhamos todos os erros, independentemente se foi para mais ou menos. Ao final obtemos a média.
- × No nosso exemplo (2, 3, 3, 1, 2 e 3). Note que enquanto menor o MAE, melhor.
- × $MAE = (2 + 3 + 3 + 1 + 2 + 3) / 6$
- × $MAE = (2 + 7 + 5) / 6$
- × $MAE = (9 + 5) / 6$
- × $MAE = 14/6$
- × $MAE = 2,33$

Temperatura (x)	Vendas reais (y)	Vendas previstas (\hat{y})
52	0	2
67	14	17
70	23	20

REGRESSÃO

× 2. EQR (Erro Quadrático Médio)

- × Melhor ter um modelo que erre uma quantidade do que um que produz poucos erros, mais maiores. Note que enquanto menor o EQR, melhor
- × Exemplo $\ddot{y} = (50, 100, 200) \dots$
- × Temperatura 50 – 0 = 50
- × Temperatura 100 – 14 = 86
- × Amplificar os erros maiores é elevando-se ao quadrado todos os erros individuais e calculando-se a média.
- × $\ddot{Y} = (2, 3, 3, 1, 2 \text{ e } 3).$
- × $\text{EQR} = (4, 9, 9, 1, 4 \text{ e } 9)/6$
- × $\text{EQR} = 36/6$
- × $\text{EQR} = 6$

Temperatura (x)	Vendas reais (y)	Vendas previstas (\hat{y})
52	0	2
67	14	17
70	23	20

REGRESSÃO

- ✖ **3. REQM (Raiz do Erro Quadrático Médio)**
- ✖ EQR não mede a precisão em termos de número de sorvetes que foram previstos erradamente. Indica apenas o nível de erro nas previsões.
- ✖ E como medir o erro em termos do número de sorvetes? Basta $\sqrt{6} = 2.45$ sorvetes.
- ✖ Exemplo. Modelo com RMSE de R\$ 50.000 em preços de casas.

Temperatura (x)	Vendas reais (y)	Vendas previstas (\hat{y})
52	0	2
67	14	17
70	23	20

REGRESSÃO

- ✖ **4. R^2 (Coeficiente de Determinação)**
- ✖ Varia entre 0 – 1. Quanto maior, melhor. Expressa a quantidade de variância dos dados que é explicada pelo modelo.
- ✖ Exemplo R^2 82,0 = modelo explica que 82,0% da variável dependente é explicada pelos regressões.



REGRESSÃO

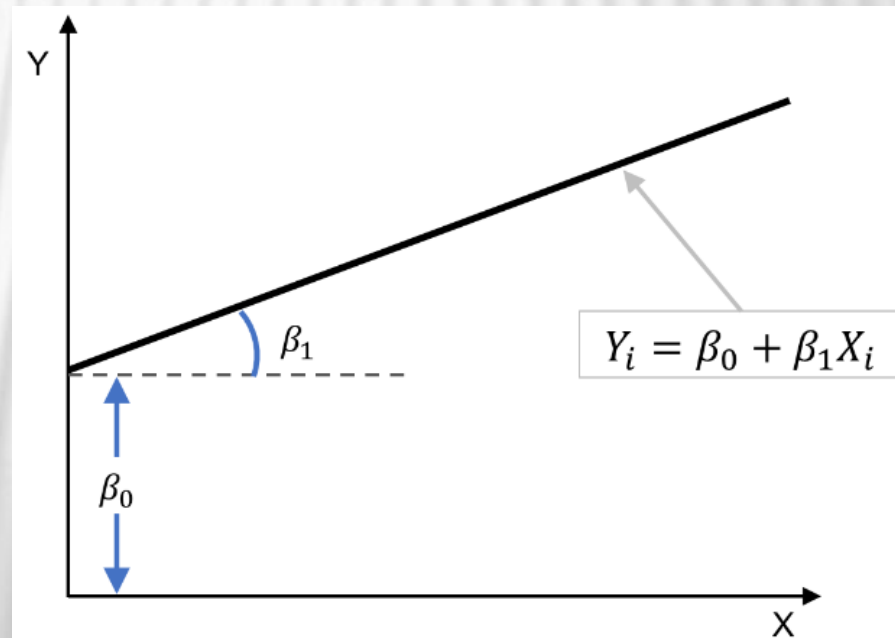
- × Após o treinamento saberemos os valores dos **coeficientes**. (β_0 e β_1).
- × **Y** - variável dependente.
- × β_0 – intercepto. Valor de Y quando X é 0.
- × β_1 – slope (inclinação). Taxa de variação, representando o quão inclinada será a reta.
- × **e** – erro. Diferença entre o valor observado e previsto.
- × **X** – variável independente

$$Y = \beta_0 + \beta_1 X + e$$

Variável
reposta Intercepto Coeficiente
angular Variável
explicativa Erro

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

Regressão Linear Múltipla





REGRESSÃO

- ✖ Na prática avaliamos elas em conjunto com a área de negócio.
- ✖ E quando o modelo não tem um bom desempenho? O que fazer?
- ✖ Exercício prático.



CLASSIFICAÇÃO BINÁRIA

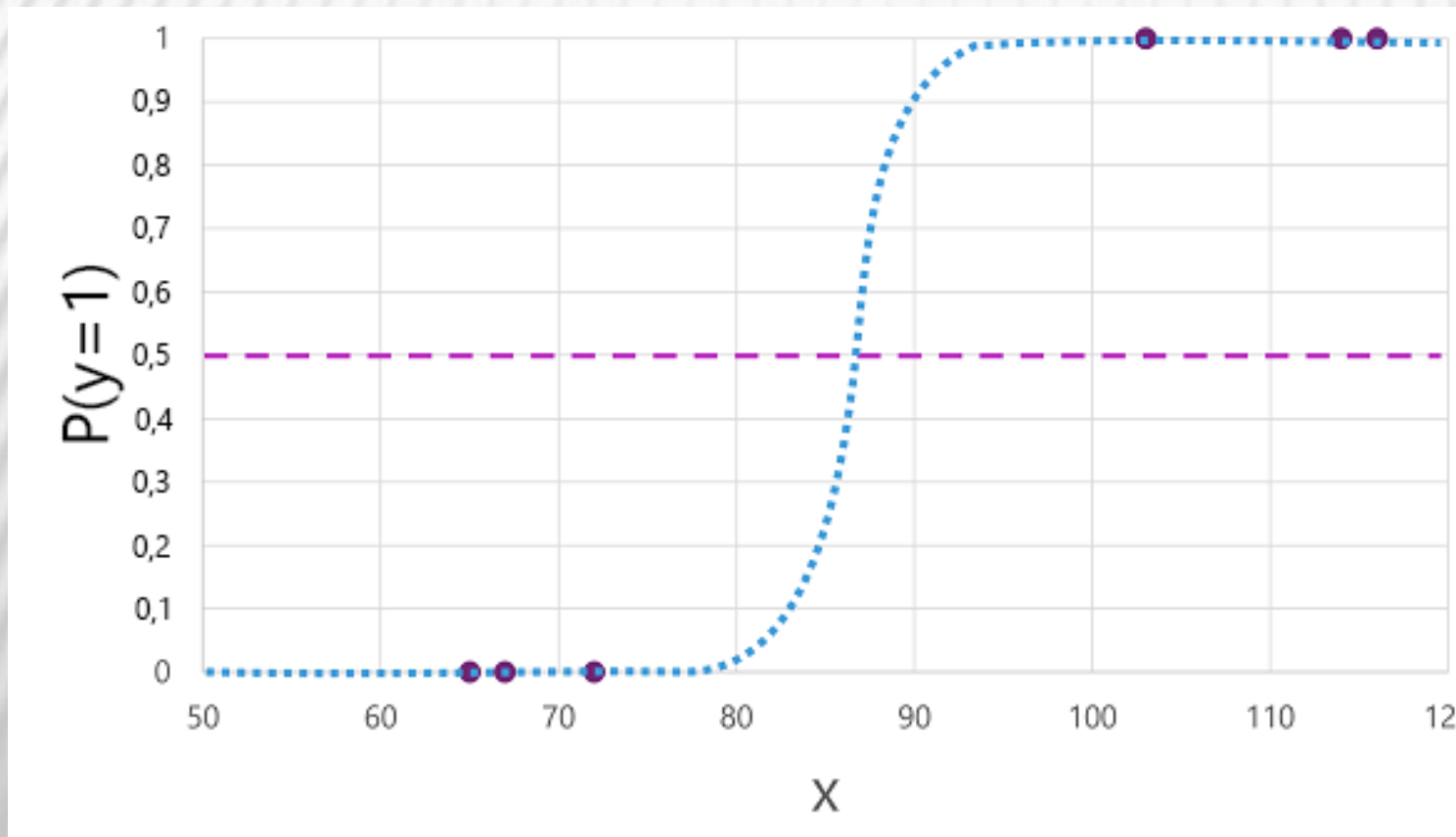
- ✖ Prever um **rótulo** que representa uma **categoria binária** (2 opções). Variável qualitativa seja nominal ou ordinal. Resultado será uma probabilidade 0 – 1 para cada categoria.
- ✖ Exemplo: prever diabetes baseado da glicose no sangue.
- ✖ **Y** = tem diabete ou não.
- ✖ **X** = nível de glicose.
- ✖ Exemplos de dados de **treinamento**.



Glicose no sangue (x)	Diabético? (y)
67	0
103	1
114	1
72	0
116	1
65	0

CLASSIFICAÇÃO BINÁRIA

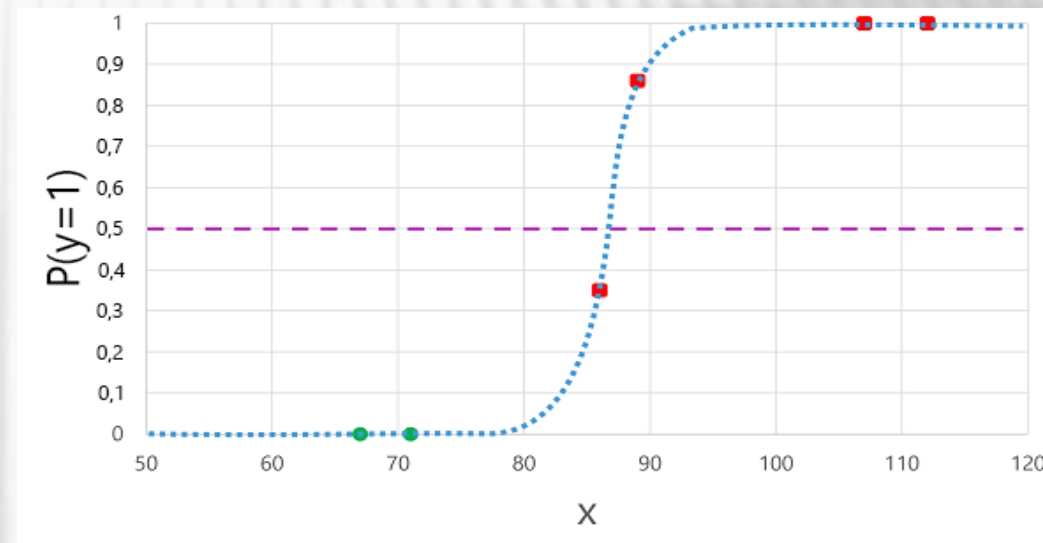
- ✖ O algoritmo utilizado irá tentar gerar uma função aproximada que calcula a probabilidade do rótulo ser verdadeira. Existem vários: Decision Tree, Regressão Logística (curva em formato de S).
- ✖ **$f(x) = P(y = 1 | x)$. Probabilidade de y ser verdadeiro(igual a 1) dado um valor de x . Threshold (limite)**



CLASSIFICAÇÃO BINÁRIA

- ✖ Dados utilizados no **teste**. Contabilizaremos os **erros** e **acertos**.

Glicose no sangue (x)	Diagnóstico real de diabetes (y)	Diagnóstico previsto de diabetes (\hat{y})
66	0	0
107	1	1
112	1	1
71	0	0
87	1	0
89	1	1



CLASSIFICAÇÃO BINÁRIA

- × Métricas.
- × 1. Matriz de Confusão.

REAL			

acertos

erros

VP - Verdadeiros Positivos

VN - Verdadeiros Negativos

FP - Falsos Positivos

FN - Falsos Negativos

- × **VP**. O original era VERDADEIRO (1) e o modelo previu VERDADEIRO (1).
- × **VN**. O original era FALSO(0) e o modelo previu FALSO(0).
- × **FP**. O original era FALSO(0) e o modelo previu POSITIVO(1).
- × **FN**. O original era VERDEIRO(1) e o modelo previu FALSO(0).

CLASSIFICAÇÃO BINÁRIA

- ✖ Dados utilizados no **teste**. Contabilizaremos os **erros** e **acertos**.

Glicose no sangue (x)	Diagnóstico real de diabetes (y)	Diagnóstico previsto de diabetes (\hat{y})
66	0	0
107	1	1
112	1	1
71	0	0
87	1	0
89	1	1

	VALOR PREVISTO	
VALOR ORIGINAL	0 NEGATIVO	1 POSITIVO
0 NEGATIVO	2 (VN)	0 (FP)
1 POSITIVO	1 (FN)	3 (VP)

CLASSIFICAÇÃO BINÁRIA

2. Acurácia.

$$\text{Acurácia} = \frac{\text{total de previsões corretas}}{\text{total de previsões}}$$

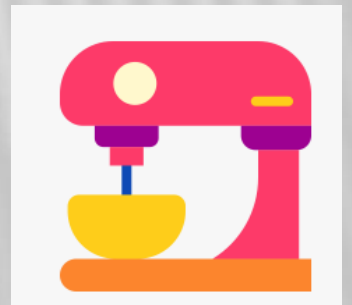
- É a mais famosa, mais fácil de interpretar. Conhecida com taxa de acerto. Detalhando.

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN}$$

- Dentro **todas** as classificações, **quantas** o **modelo classificou** corretamente.
- Exercício. Calcular a acurácia para nosso exemplo. Cuidado que apenas essa métrica pode enganar.

$$\text{Acurácia} = 0,83$$

	VALOR PREVISTO	
VALOR ORIGINAL	0 NEGATIVO	1 POSITIVO
0 NEGATIVO	2 (VN)	0 (FP)
1 POSITIVO	1 (FN)	3 (VP)



CLASSIFICAÇÃO BINÁRIA

3. Precisão.

$$\text{Precisão} = \frac{\text{VP}}{\text{VP} + \text{FP}}$$

- Dentro **todas** as classificações verdadeiras (positiva) que o **modelo fez**, **quantas estão corretas**. Ou seja, o modelo disse que era positiva e de fato era.
- Exercício. Calcular a precisão para nosso exemplo.

$$\text{Precisão} = 1,0$$

VALOR ORIGINAL	VALOR PREVISTO	
	0 NEGATIVO	1 POSITIVO
0 NEGATIVO	2 (VN)	0 (FP)
1 POSITIVO	1 (FN)	3 (VP)

CLASSIFICAÇÃO BINÁRIA

3. Recall (Revocação).

$$\text{Recall} = \frac{\text{VP}}{\text{VP} + \text{FN}}$$

- ✖ Dentro **todas** as situações de classe Positivo como valor esperado, quantas estão corretas. Exemplo de utilidade: previsão de câncer.
- ✖ **Falso positivo** – o original era falso e o modelo previu positivo.
- ✖ **Falso negativo** - o original era positivo e o modelo previu negativo.
- ✖ Falso negativos mais graves que falso positivo.
- ✖ Exercício. Calcular o recall para nosso exemplo.

$$\text{Recall} = 0,75$$

VALOR ORIGINAL	VALOR PREVISTO	
	0 NEGATIVO	1 POSITIVO
0 NEGATIVO	2 (VN)	0 (FP)
1 POSITIVO	1 (FN)	3 (VP)

CLASSIFICAÇÃO BINÁRIA

* 4. F1-Score.

$$\text{Recall} = \frac{2 * \text{Precisão} * \text{Revocação}}{\text{Precisão} + \text{Revocação}}$$

- * Média harmônica entre precisão x recall.
- * Exercício. Calcular o F1 para nosso exemplo.

$$\text{F1-Score} = 0,86$$

	VALOR PREVISTO	
VALOR ORIGINAL	0 NEGATIVO	1 POSITIVO
0 NEGATIVO	2 (VN)	0 (FP)
1 POSITIVO	1 (FN)	3 (VP)

CLASSIFICAÇÃO BINÁRIA

✖ 5. Curvas AUC e ROC

- ✖ **ROC** significa Características operacionais do receptor, *Receiver Operating Characteristic*. Essa curva mostra o quão bom o modelo pode distinguir entre duas classes (0 negativa ou 1 positiva). Uma curva **ROC** traça “*True Positive Rate* vs. *False Positive Rate*” em diferentes limiares de classificação.

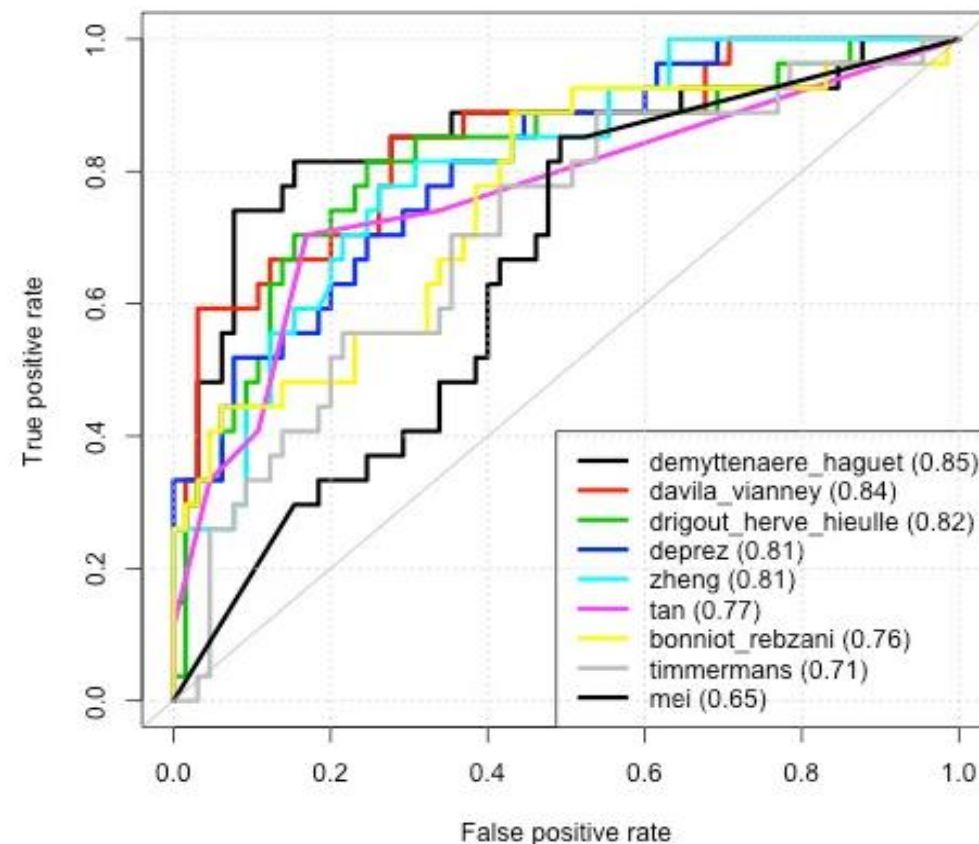
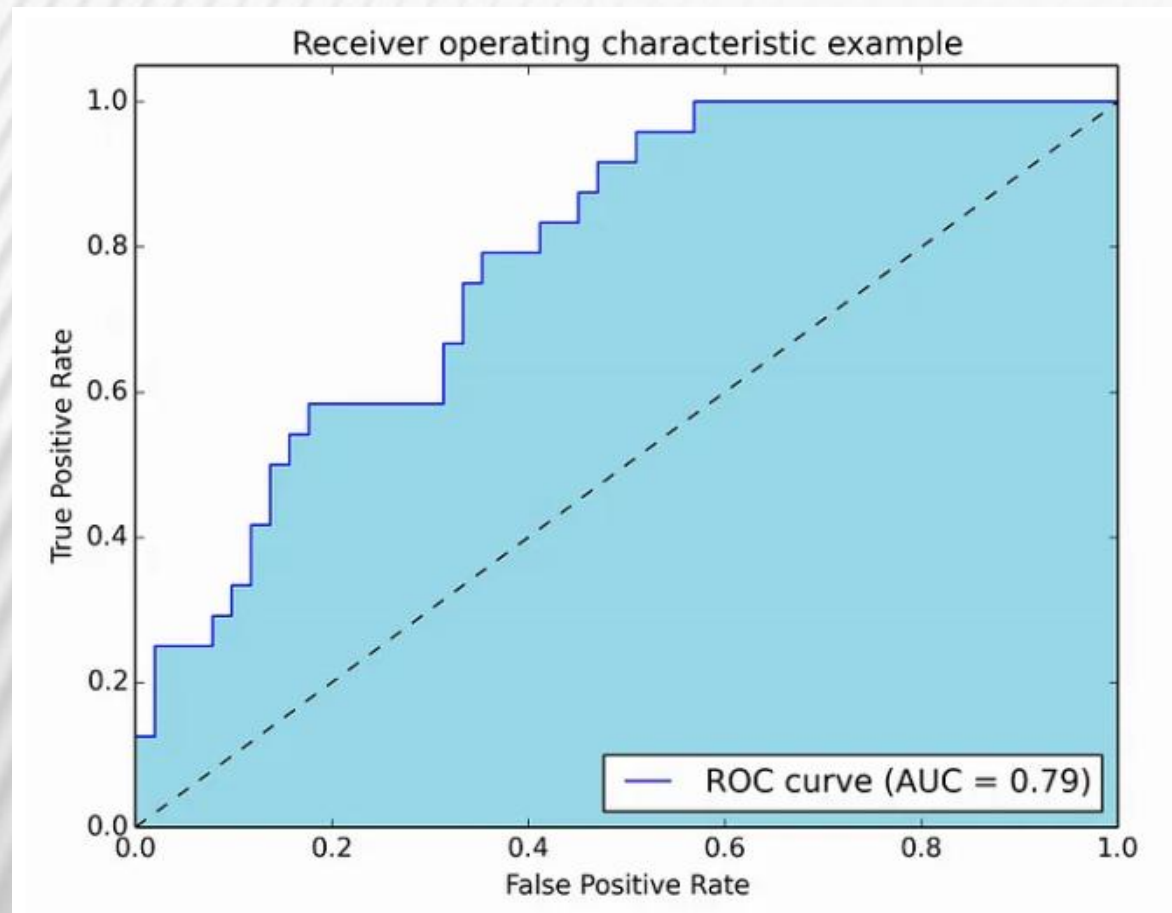
ROC

- **Taxa de Verdadeiro Positivo (TPR).**
- $TPR = VP / (VP + FN)$
- **Taxa de Falso Positivo (FPR).**
- $FPR = FP / (FP + VN)$

AUC (Área Sobre a Curva, Area Under Curve). Resume a ROC em um único valor de 0 – 1, agregando todos os limiares. Quanto maior, melhor.

CLASSIFICAÇÃO BINÁRIA

× 5. Curvas AUC e ROC



CLASSIFICAÇÃO BINÁRIA

× Exercício.



CLASSIFICAÇÃO MULTICLASSE.

✖ -- EM DESENVOLVIMENTO --