



PPGI PROGRAMA
DE PÓS-GRADUAÇÃO
EM INFORMÁTICA
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

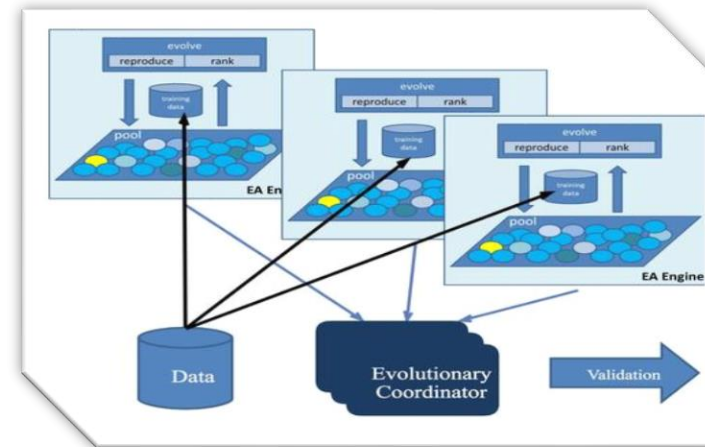
Machine Learning



Prof. Jorge Zavaleta
Prof. Sergio Serra

RJ, Março de 2021

Machine Learning (ML)



Machine Learning

“Aprendizado de máquina é o campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados”
– Arthur Samuels, 1959.



Machine Learning

O que resolve ML?



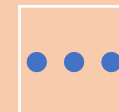
Problemas com grande número de regras ou processos manuais repetitivos.



Problemas complexos sem boas soluções.

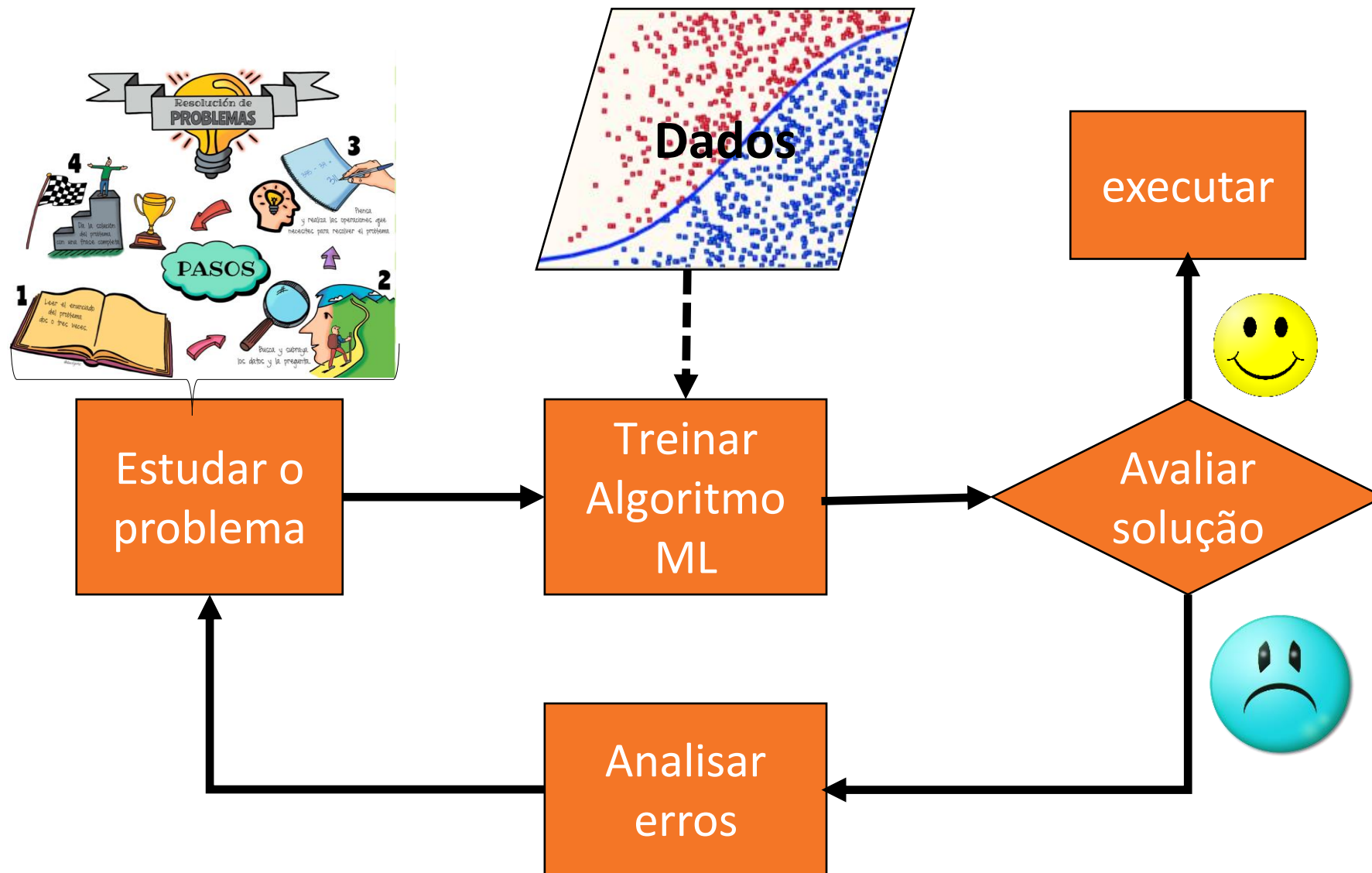


Obter soluções de grandes volumes de dados .



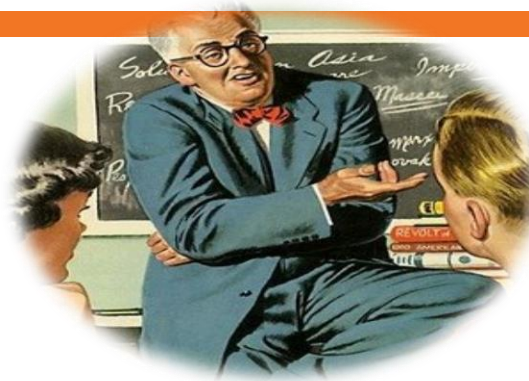
...

FUNDAMENTOS DE CIÊNCIA DE DADOS



ML – Tipos de Aprendizado

Aprendizagem



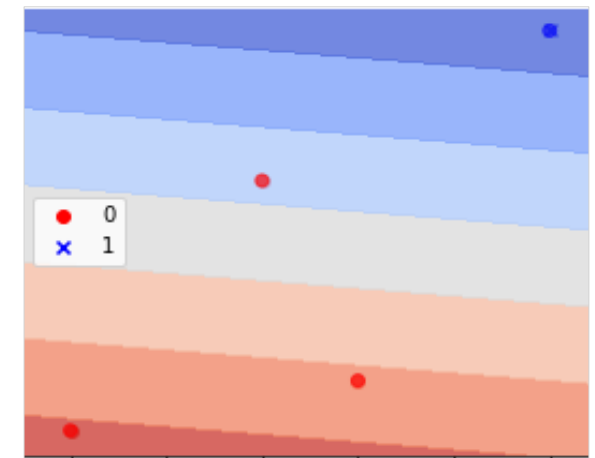
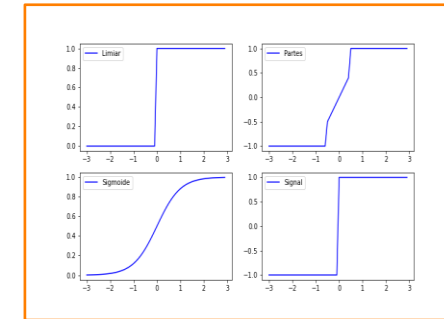
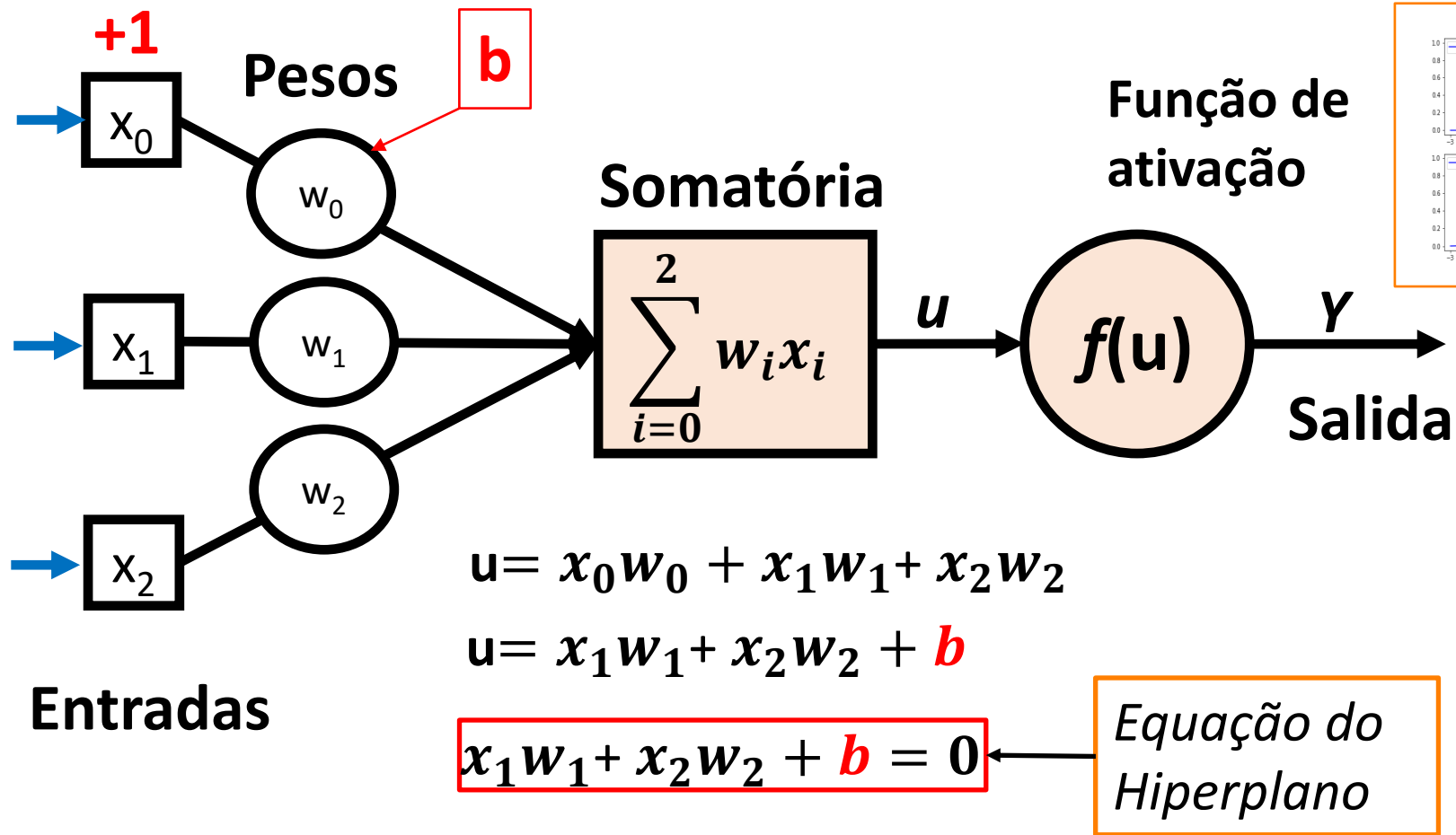
Supervisionado



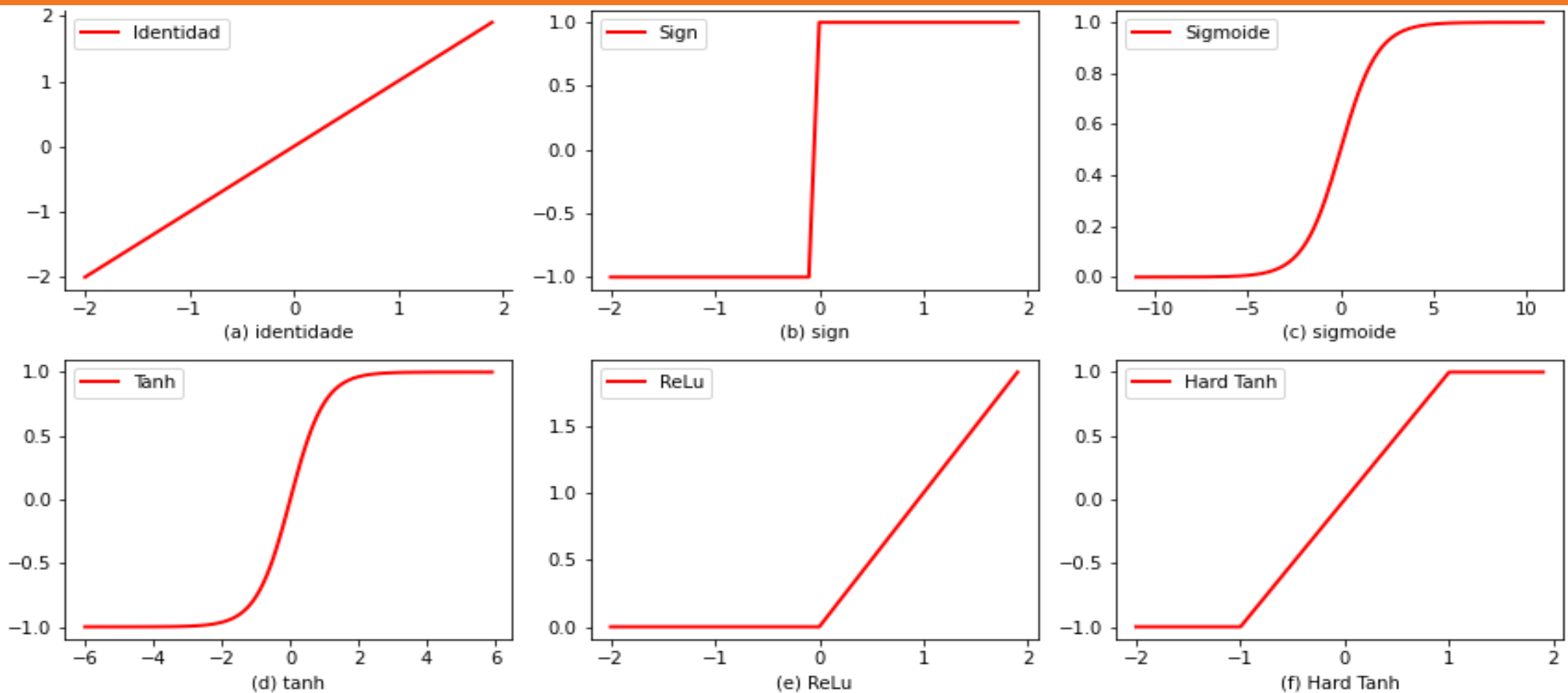
Não Supervisionado

Por Reforço

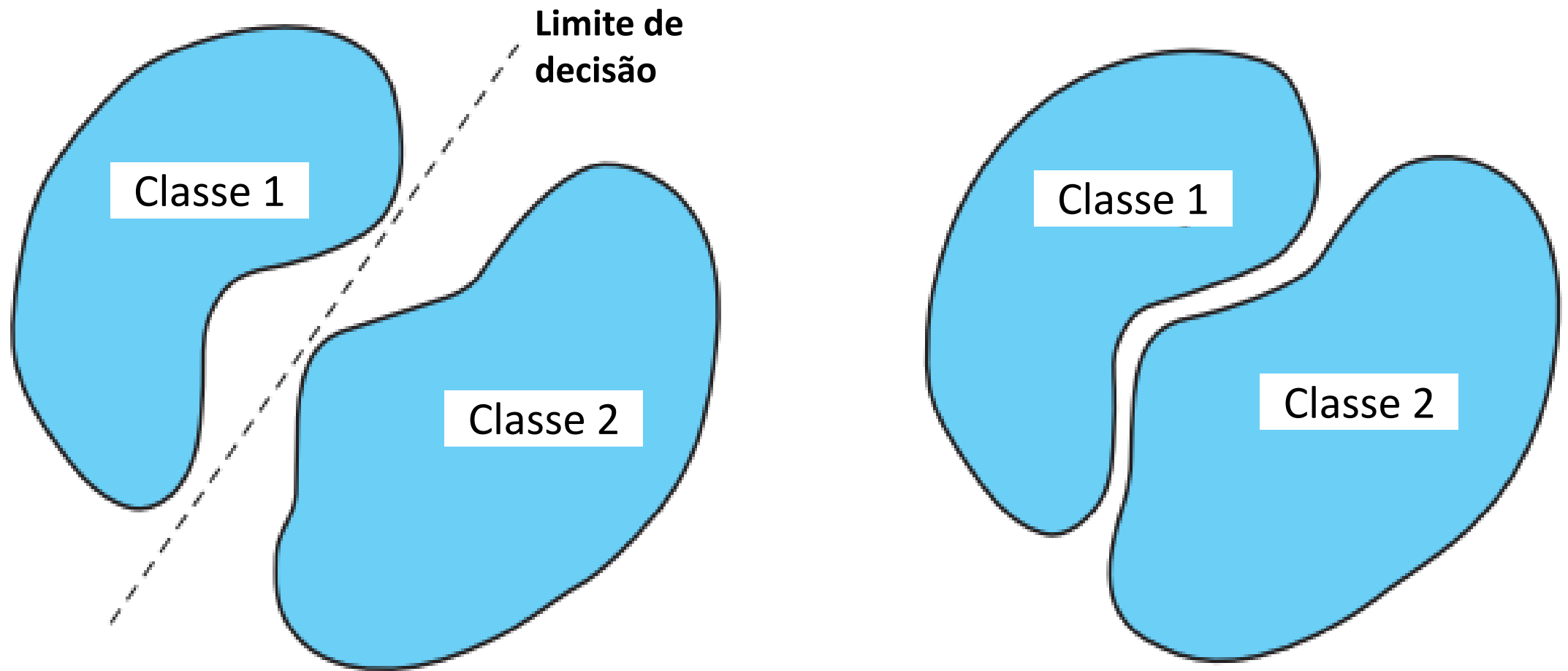
Aprendizagem – Modelo matemático



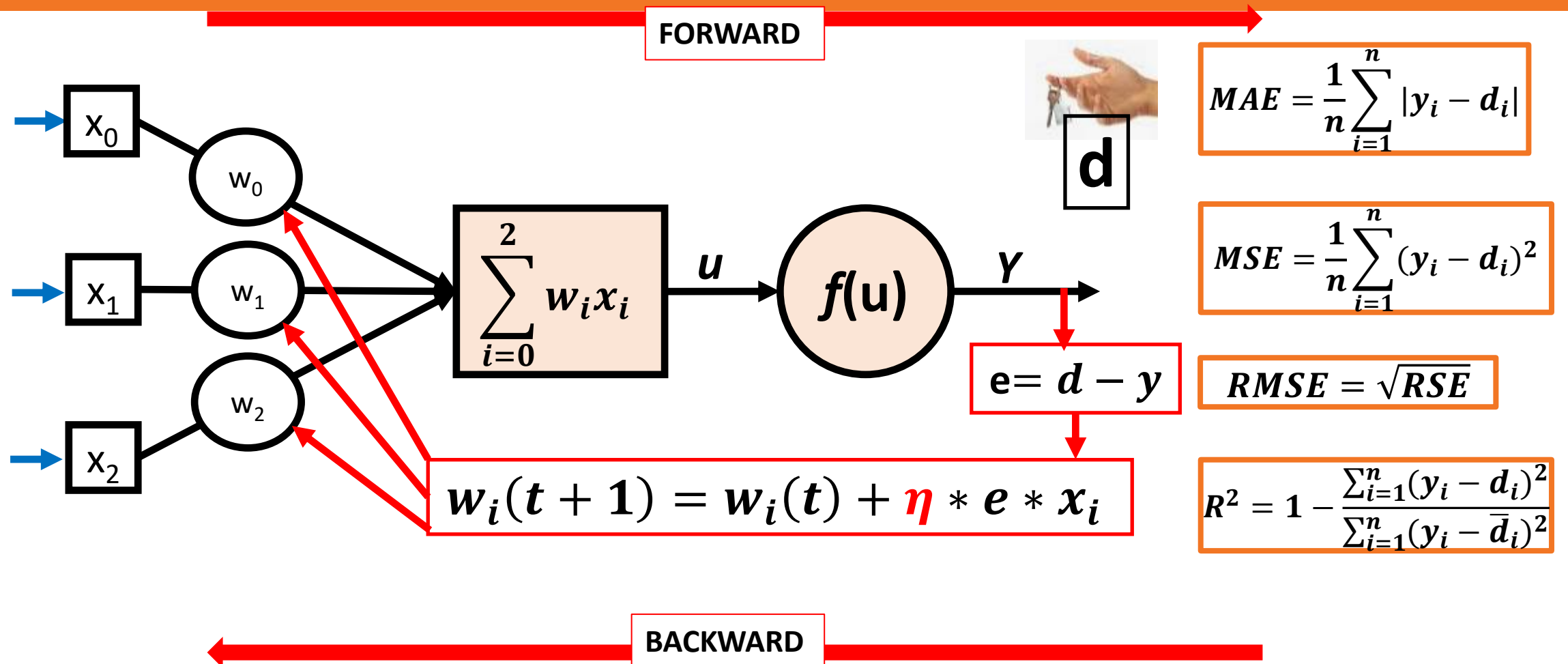
Funções de Ativação



Hiperplano



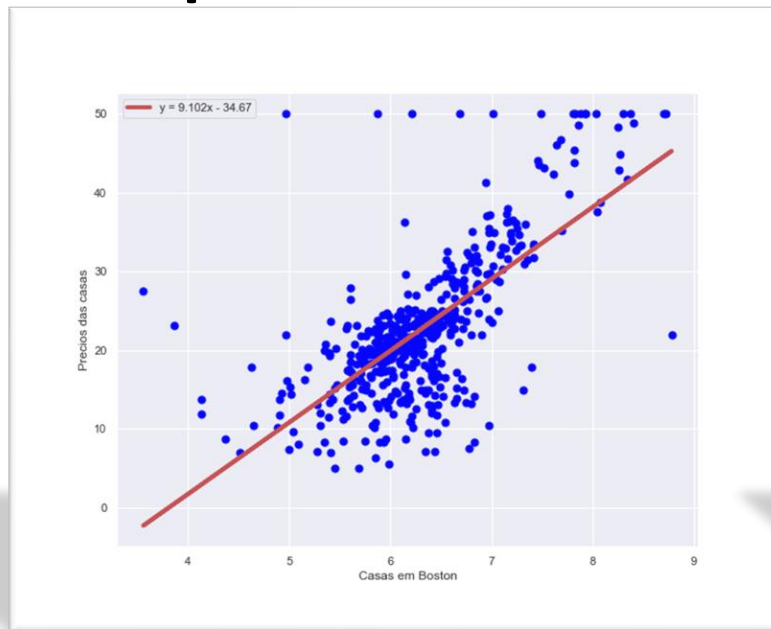
Aprendizagem (perceptron)



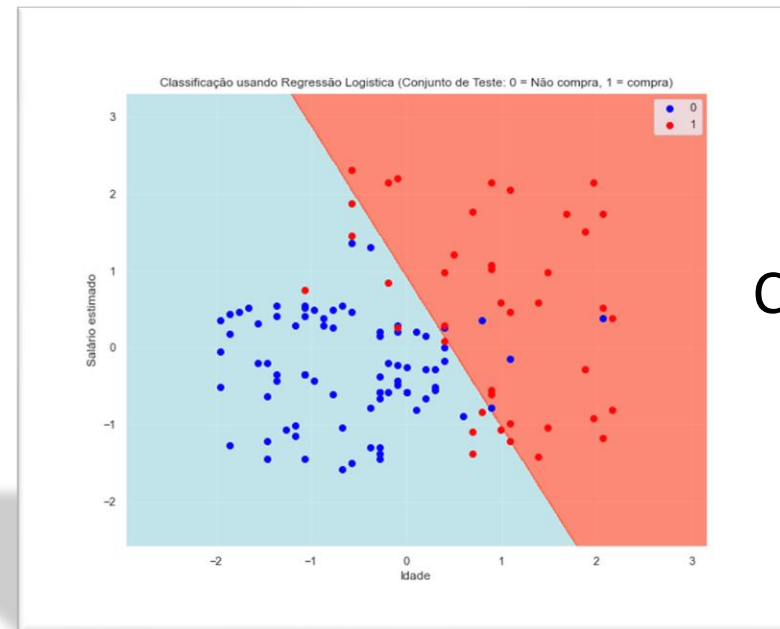
Aprendizado Supervisionado

- Os algoritmos de aprendizado supervisionados fazem previsões com base em um conjunto de exemplos de entrada. Neste tipo de aprendizagem existe um “professor/tutor” que avalia a resposta da rede ao padrão atual de entradas.

Regressão

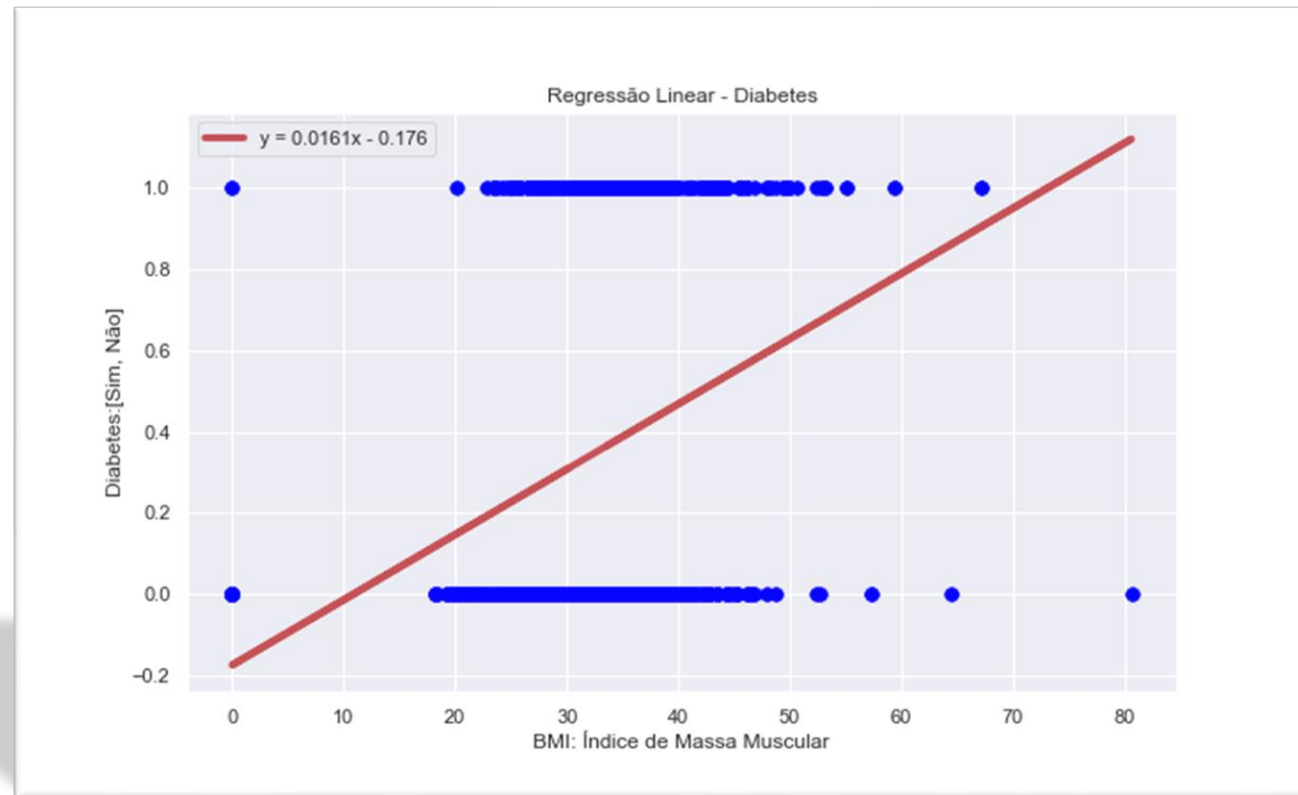


Classificação



Aprendizado Supervisionado

- **Regressão:** métodos de regressão buscam encontrar como uma variável evolui em relação a outras.



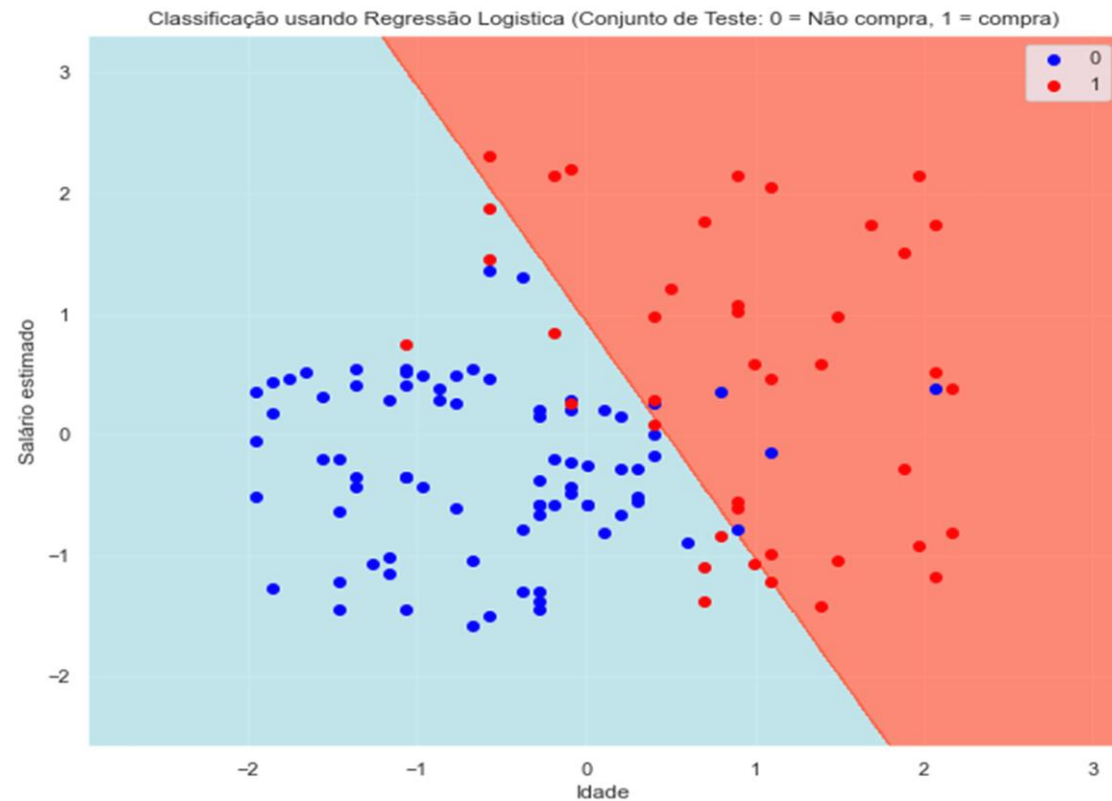
Diabetes

Regressão

- Previsão do mercado de ações,
- Previsão de demanda,
- Estimativa de preços,
- Otimização de lances de anúncios,
- Gerenciamento de riscos,
- Gerenciamento de ativos,
- Previsão do tempo (clima),
- Previsão de esportes

Aprendizado Supervisionado

- **Classificação:** são métodos que buscam explicar uma variável categórica, com duas categorias (variável binária) ou mais.

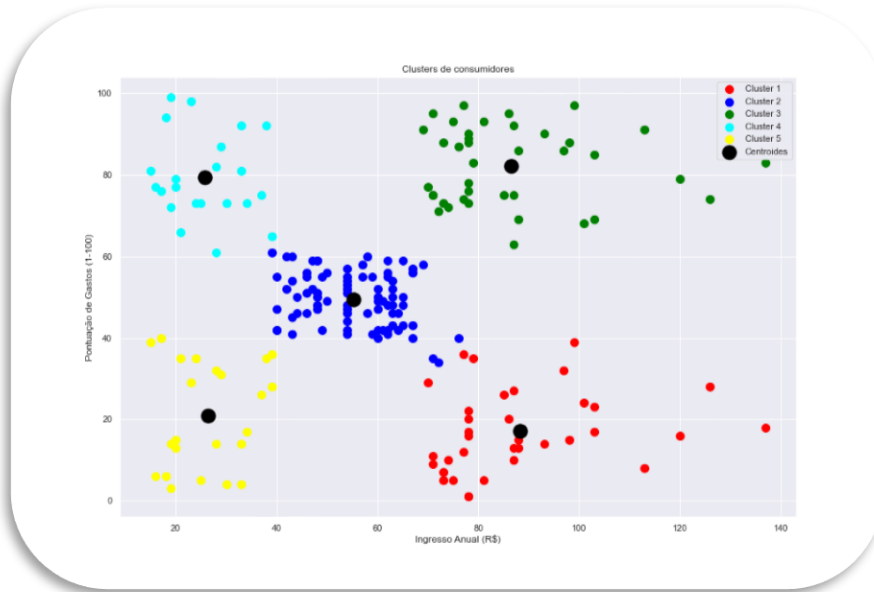


Classificação - Aplicações

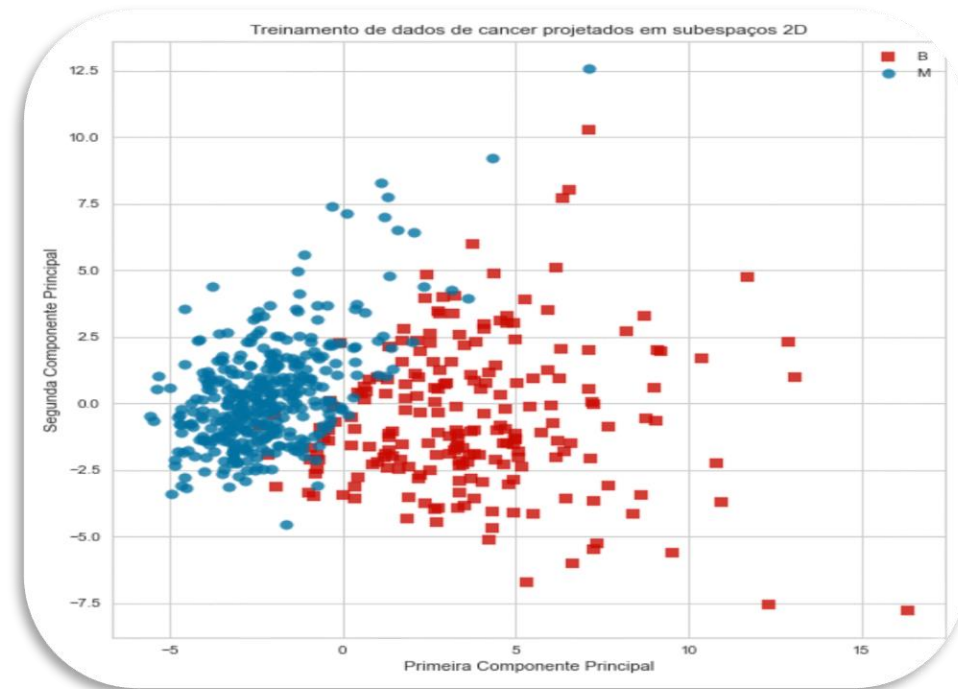
- Filtros de spam, análise de sentimentos, detecção de fraude, segmentação de anúncios de clientes, previsões de rotatividade, sinalização de casos de suporte, personalização de conteúdo, segmentação de clientes, descoberta de eventos, genômica, eficácia de medicamentos, detecção de defeitos de fábrica etc.

Aprendizado Não Supervisionado

- Não existe “**professor/tutor**”. A rede tem de descobrir sem ajuda as relações, padrões, regularidades ou categorias nos dados que lhe vão sendo apresentados e codificá-las nas saídas.



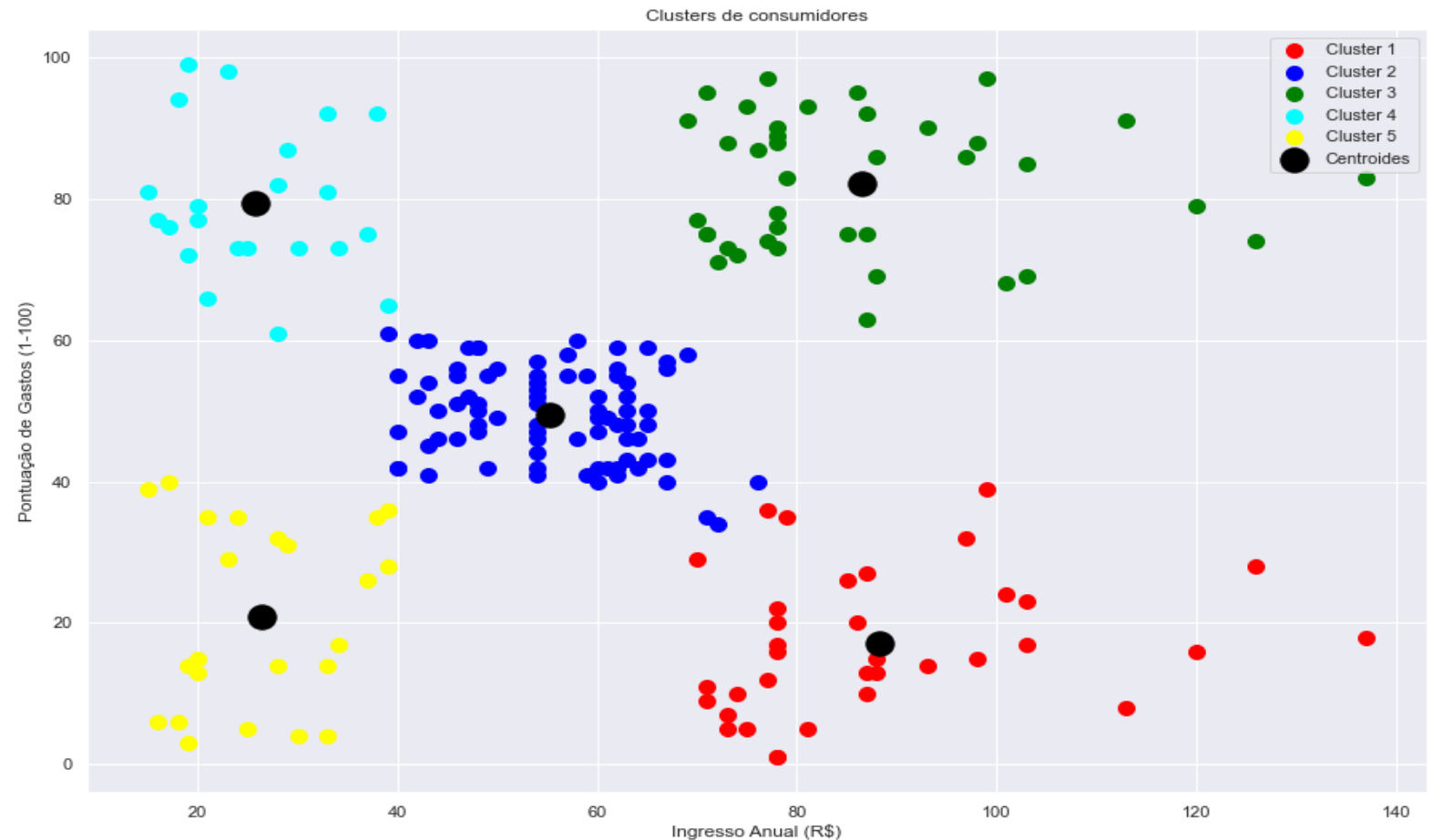
Clusterização



PCA

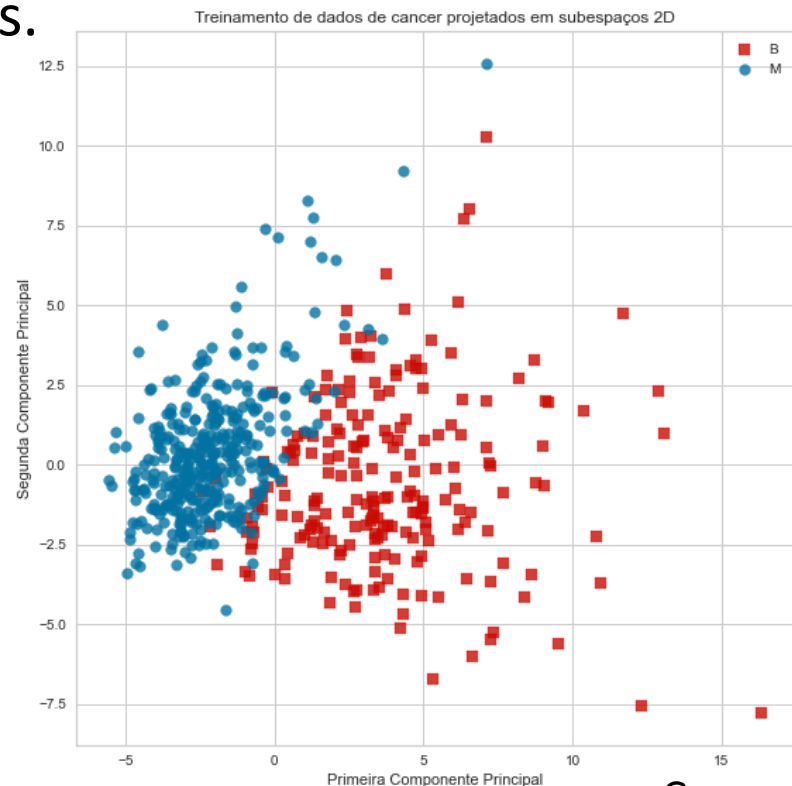
Aprendizado Não Supervisionado

- **Clusterização** é a tarefa de dividir os pontos de dados em vários grupos com características semelhantes.
- Cada grupo possui um ponto central, denominado de **centroide**.

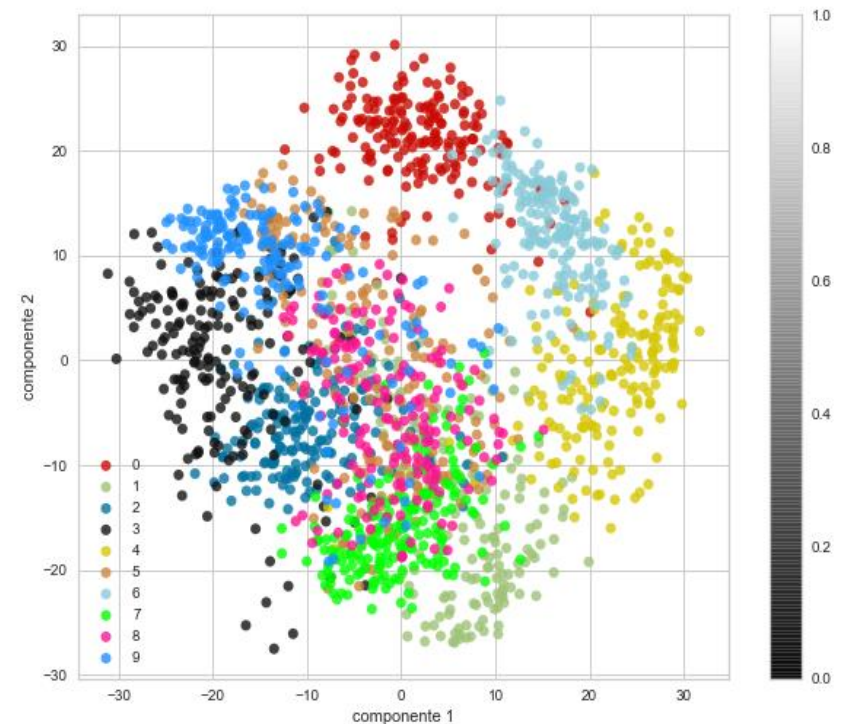


Aprendizado Não Supervisionado

- **Principal Component Analysis (PCA -Análise de Componentes Principais)** é um procedimento matemático que permite reduzir variáveis n-dimensionais em variáveis lineares.



Cancer



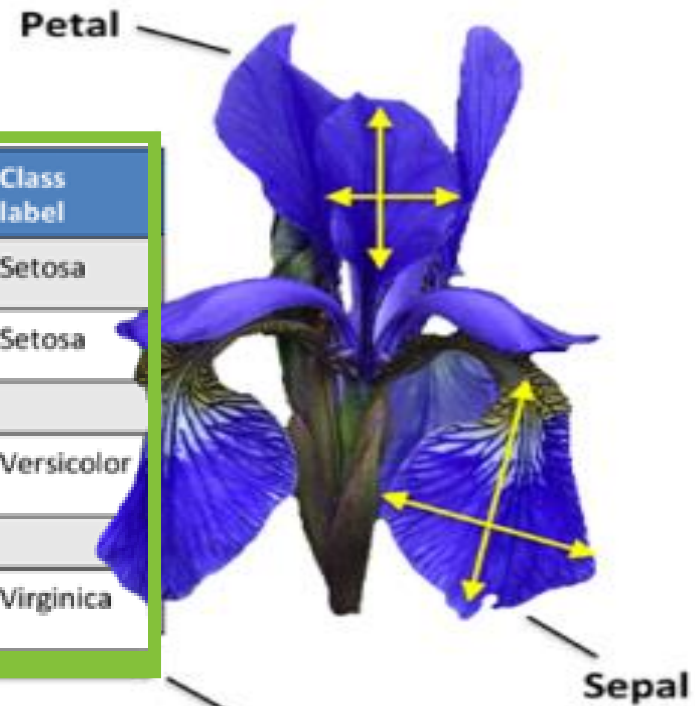
Dígitos

Matriz de Exemplos-Características

**Exemplos
(instâncias, observações)**

	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
...
50	6.4	3.5	4.5	1.2	Versicolor
...
150	5.9	3.0	5.0	1.8	Virginica

**Características
(atributos, medidas, dimensões)**



**Rótulo de classe
(objetivo)**

Prática Usando o Scikit-learn

