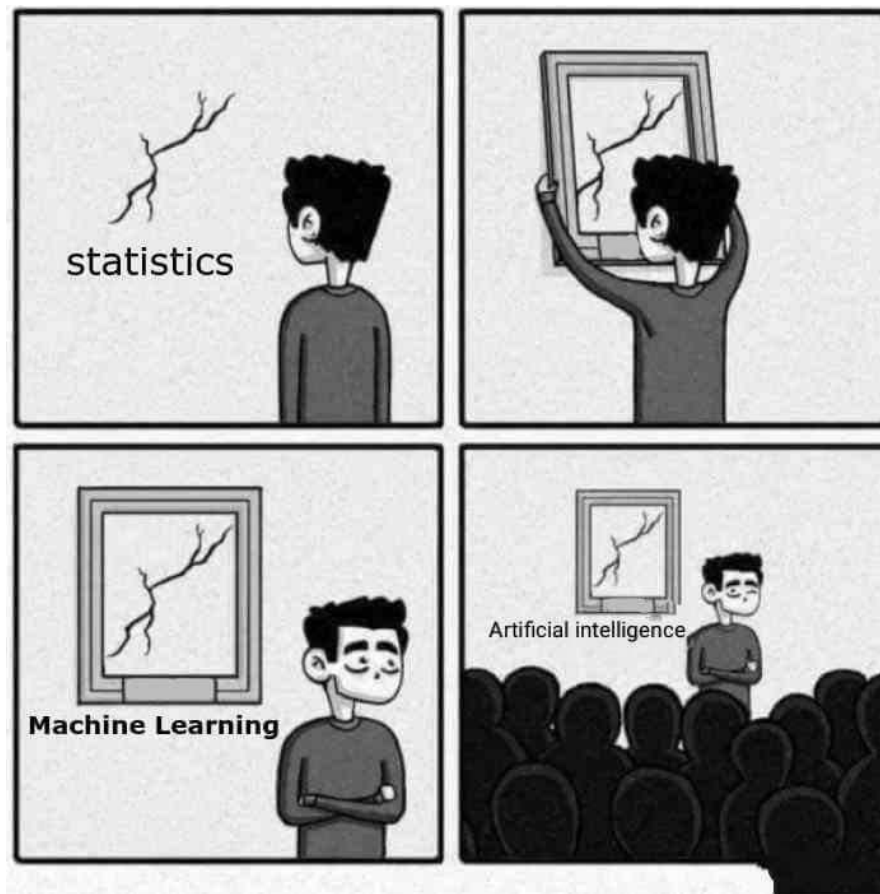


Inteligência Artificial e Machine Learning II

Aula 1

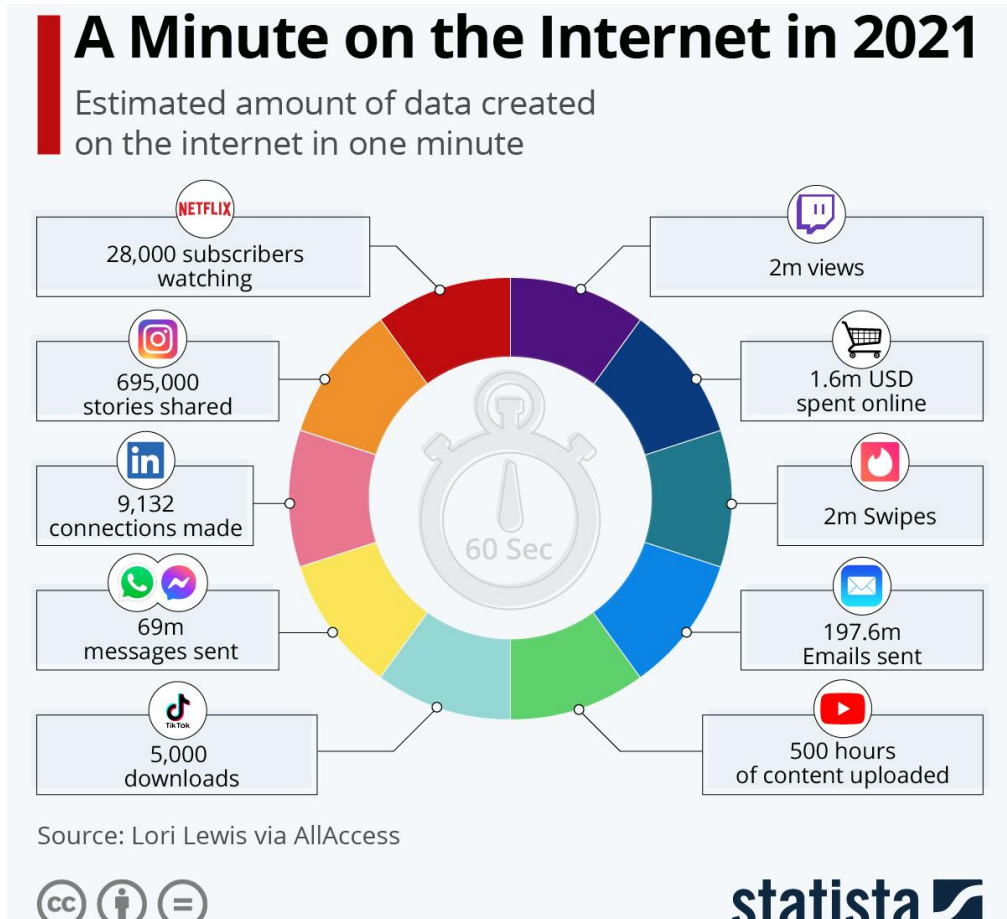
Professor: Alex Pereira

Desmistificando a Inteligência Artificial



Por que estamos aqui (neste curso) ?

Mais dados, mais processamento



ML como quebra de paradigma



Sem IA

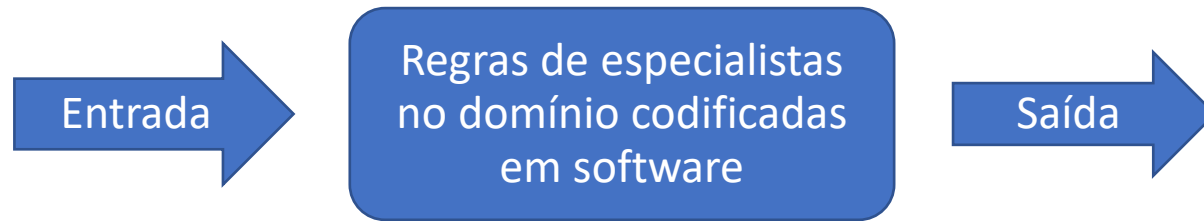


Com IA

Qual paradigma foi quebrado ?

- O da automação por meio de software

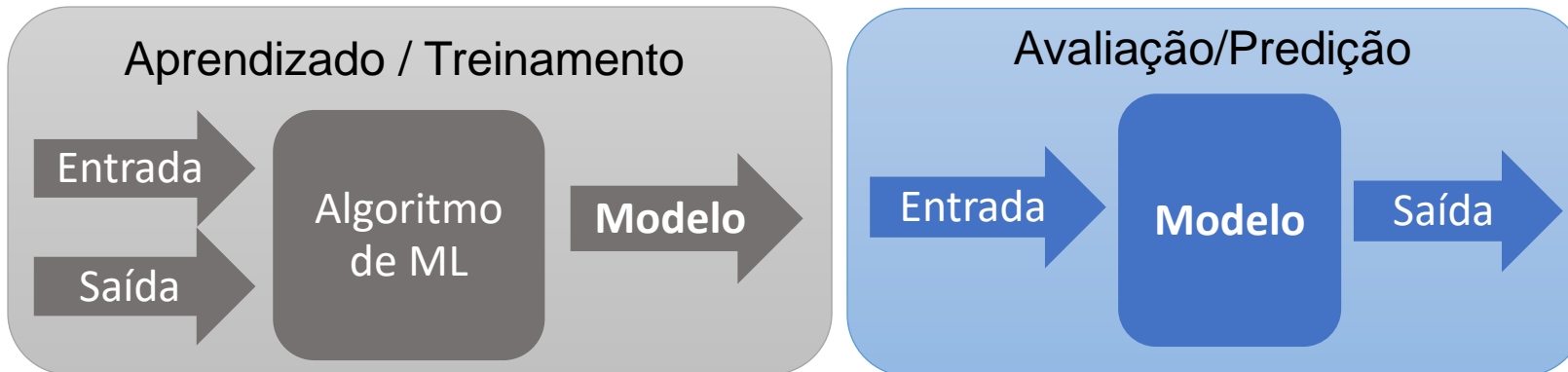
- Paradigma da codificação de regras



**Generalização precária;
Se baseia em:**

- Lógica
- Teoria estabelecida
- Bom senso

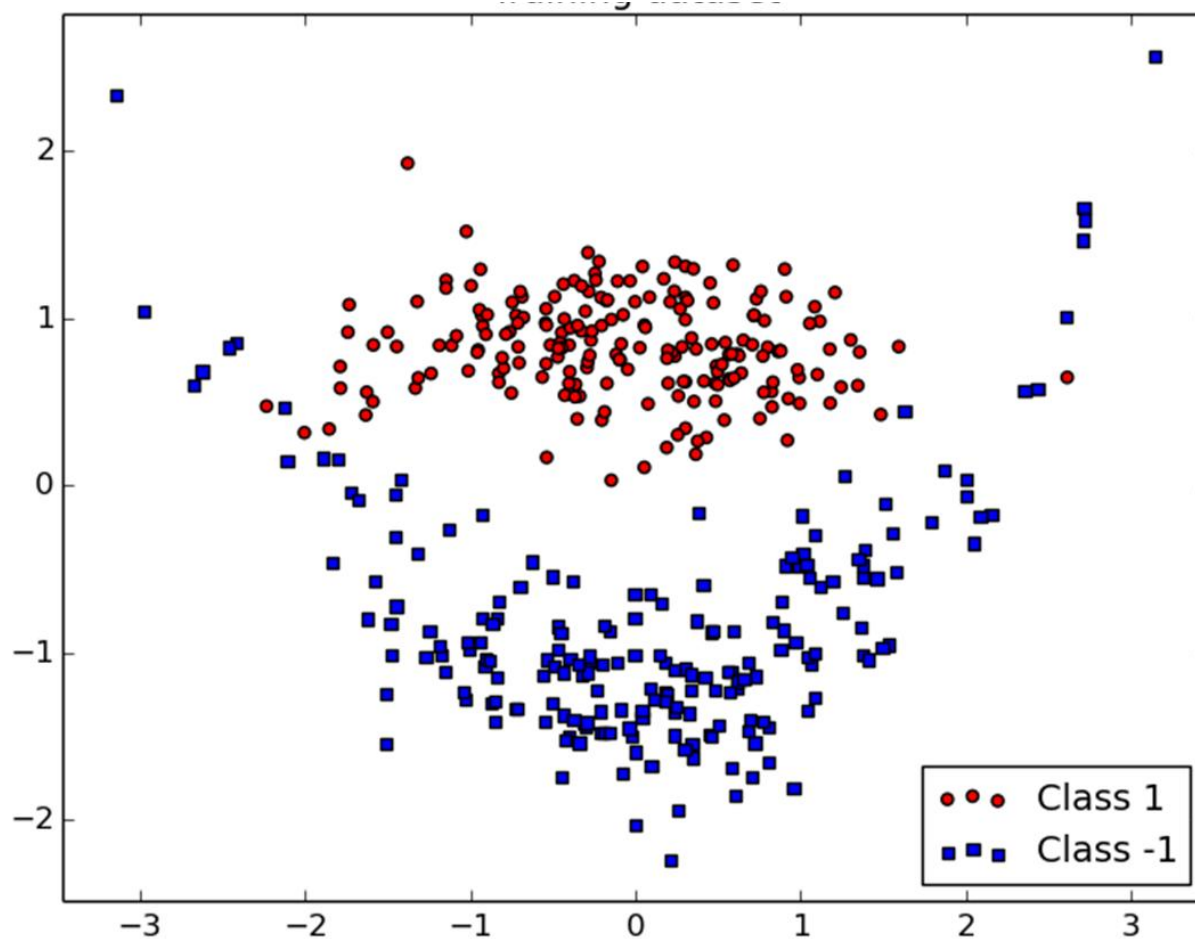
- Paradigma do aprendizado por exemplos



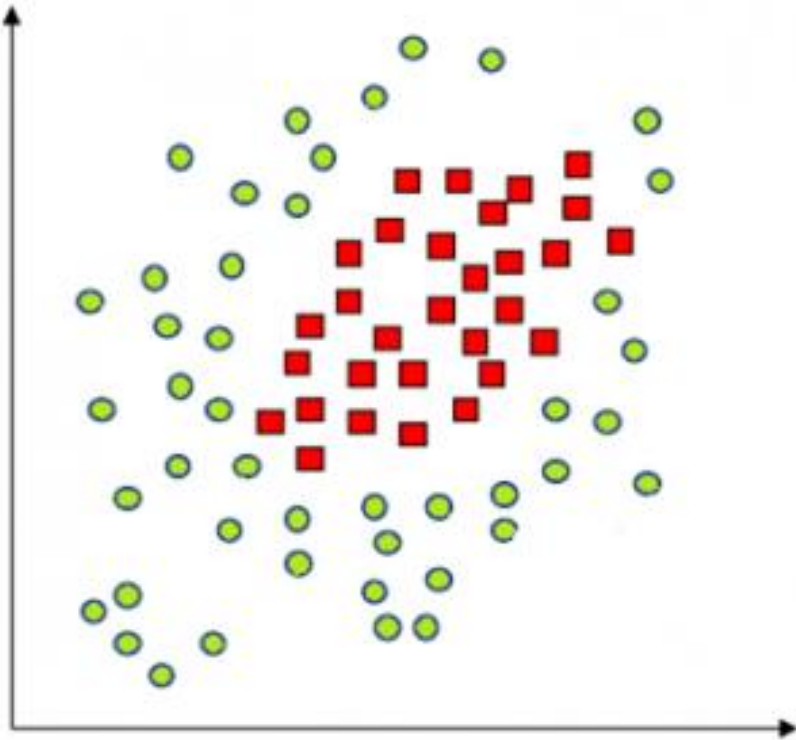
**Boa generalização;
Se baseia em:**

- Experimentação
- Estatística
- Dedução por exemplos

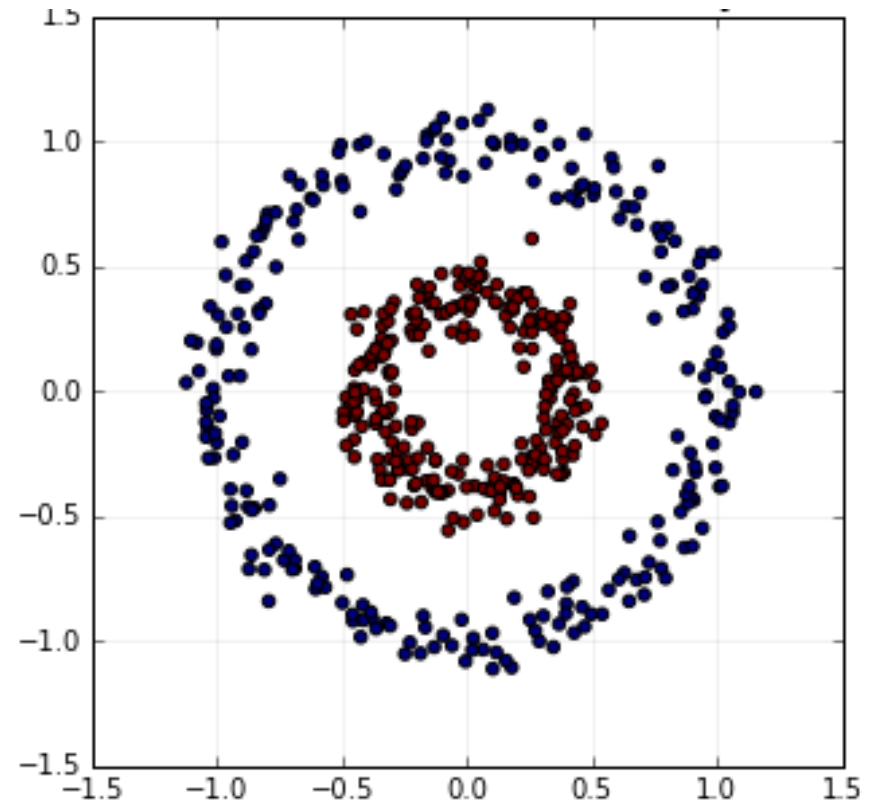
Trace uma reta que separe as classes (1)



Trace uma reta que separe as classes (2)



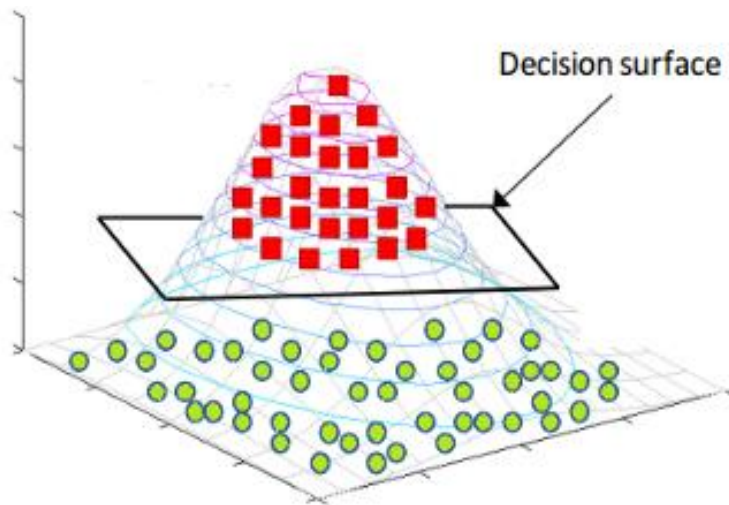
[Fonte da imagem](#)



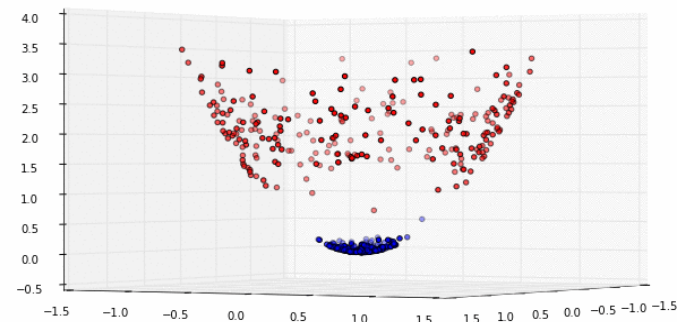
[Fonte da imagem](#)

Trace um plano que separe as classes (3)

- Adicionar uma dimensão
 - Pode facilitar o problema de classificação
 - ✓ Neste contexto, chamamos dimensão de **Feature (atributo/variável)**
 - Problemas mais complexos, exigem mais features.
 - Os símbolos ■ ● e ● ● representam as **classes (labels/rótulos)**



[Fonte da imagem](#)



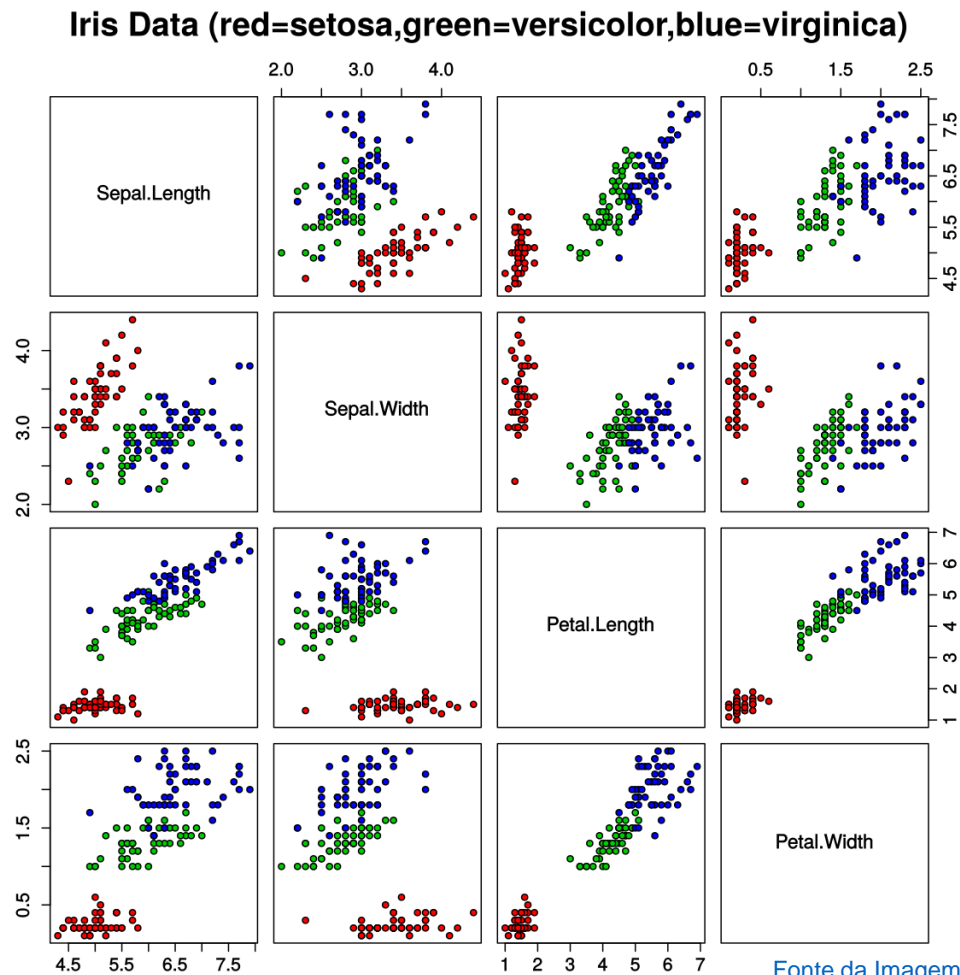
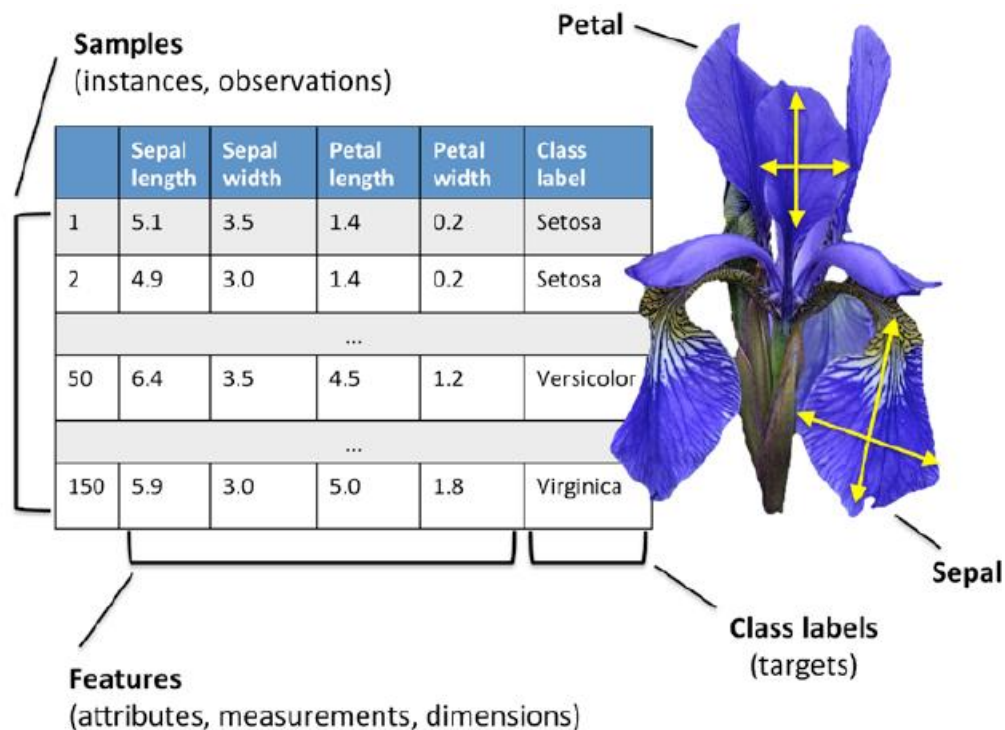
[Fonte da imagem](#)

Features (Variáveis/Atributos)

- São métricas (quantitativas)
- Usadas como entrada dos algoritmos
- Têm um potencial preditivo dos resultados/efeitos.
- As features devem ser independentes entre si
 - Ou seja, não use uma feature que seja uma combinação linear de outra(s)
 - ✓ Exemplo de combinação linear: $F_3 = 2 * F_2 + F_1$
 - Combinação linear não agrega novidade (informação nova)
 - Apenas deixa seu modelo mais complexo.

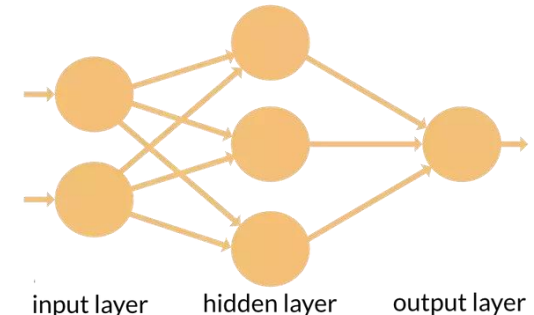
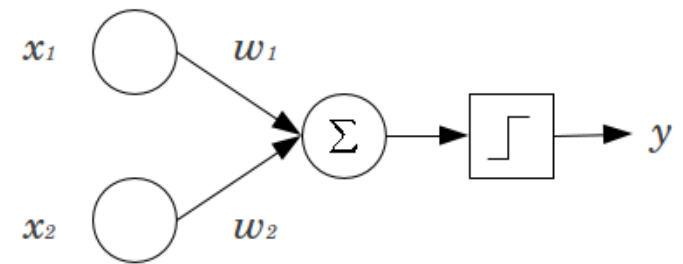
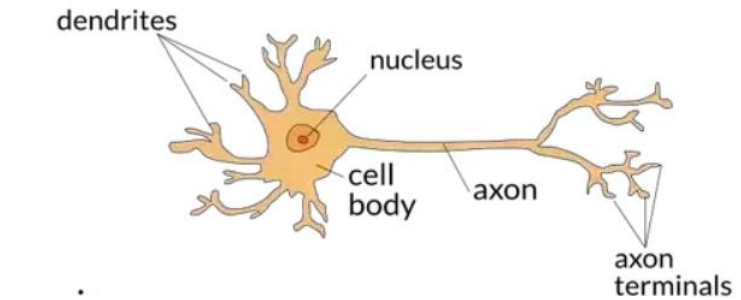
Representação Tabular de um dataset (features e labels)

- Dataset de plantas Iris



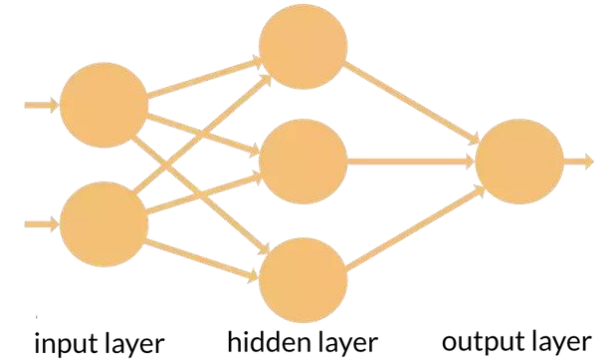
Redes Neurais Artificiais: Feed-forward Network

- Analogia com as Redes Neurais cerebrais
 - Dendritos recebem diversos sinais
 - O corpo celular processa esses sinais
 - Axônio envia um sinal
 - ✓ para outros neurônios
- Exemplo de uma rede neural artificial
 - Informação flui da esquerda p/ direita
 - ✓ Informação = Números Reais
 - RN não processam dados categóricos
- Entender primeiro a Predição
 - e depois o Treinamento



Redes Neurais Artificiais: Feed-forward Network

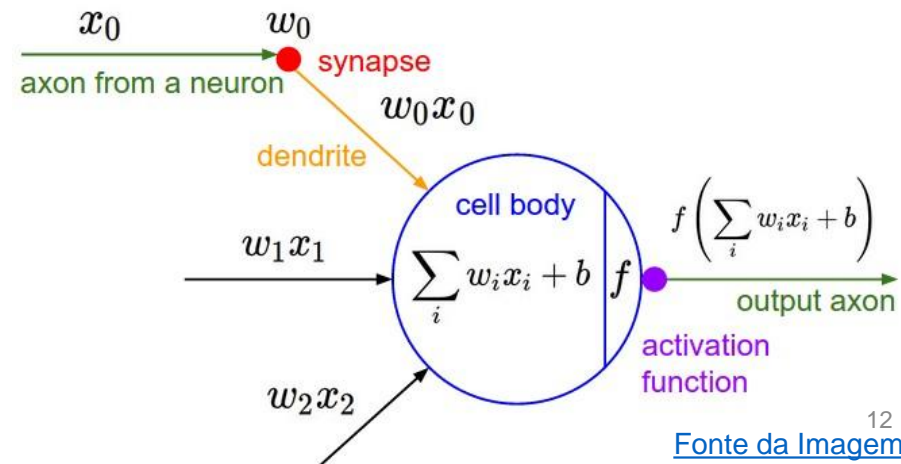
- Na Cada neurônio (unidade) está conectado
 - a todos os neurônios da camada seguinte,
 - ✓ e não se conecta a nenhum outro
 - Multi-Layer Perceptron (MLP) é um tipo de FFN
- A quantidade ótima de camadas e neurônios (nas hidden layers)
 - é determinada empiricamente (experimentando)
- A quantidade de neurônios da camada de input
 - Depende da quantidade de features/atributos disponíveis
- A quantidade de neurônios da camada de output
 - Depende do problema abordado (1 neurônio p/ classificação binária)
 - ✓ 1 neurônio p/ cada label, num classificador multi-classes



[Fonte da Imagem](#)

Feed-Forward: Entendendo a Predição

- Cada neurônio (exceto da camada de entrada) aplica uma soma ponderada mais um bias
 - $z = x_0 * w_0 + x_1 * w_1 + ... + x_n * w_n + b$
 - ✓ os valores dos sinais x são recebidos dos neurônios da camada anterior
 - a ponderação é o peso (w) de cada conexão
 - ✓ bias é um viés que impulsiona (reforça) uma tendência
 - E aplica uma função de ativação sobre o resultado da soma (z)
 - ✓ Esta é a função que faz a Rede
 - resolver problemas não lineares
- Os neurônios da camada de entrada
 - Simplesmente repassam
 - ✓ o valor das features/atributos

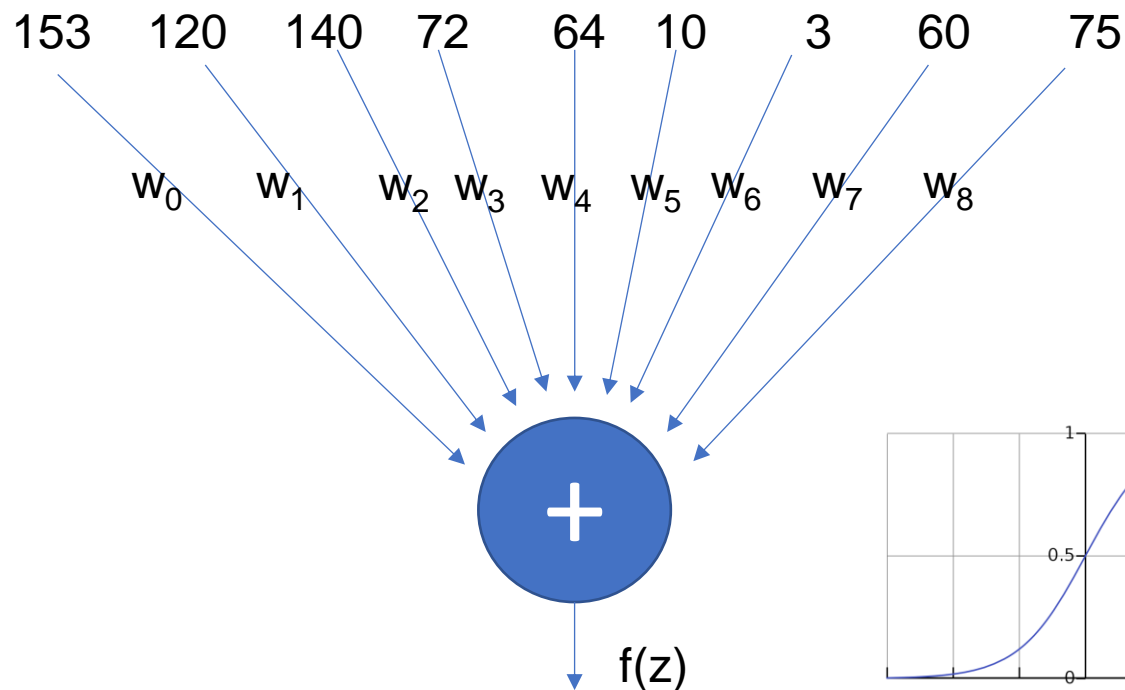


Uma rede com um neurônio

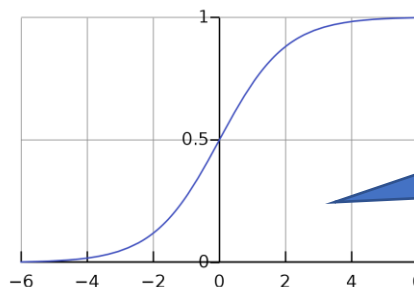
- Quais pesos w (bias $b=0$) fariam esta rede reconhecer que a 1ª linha
○ é mais clara do que a 3ª linha, ou seja, $f(z) > 0,5$?

✓ $z = 153 \cdot w_0 + 120 \cdot w_1 + \dots + 75 \cdot w_8 + b$

- Obs.: Tons de cinza variam de 0 (preto) a 255 (branco)



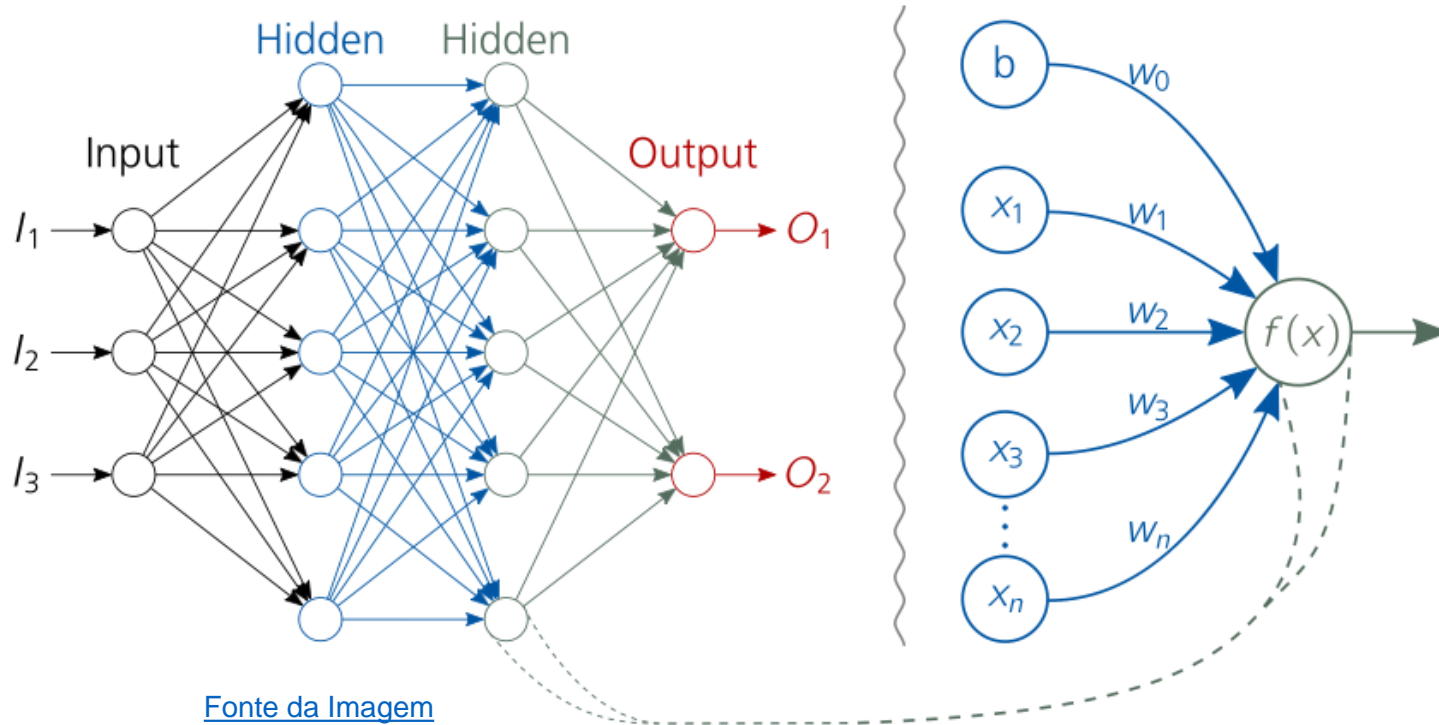
153	120	140
72	64	10
3	60	75



f é a Função sigmoid
(existem outras)

Uma Rede Neural completa

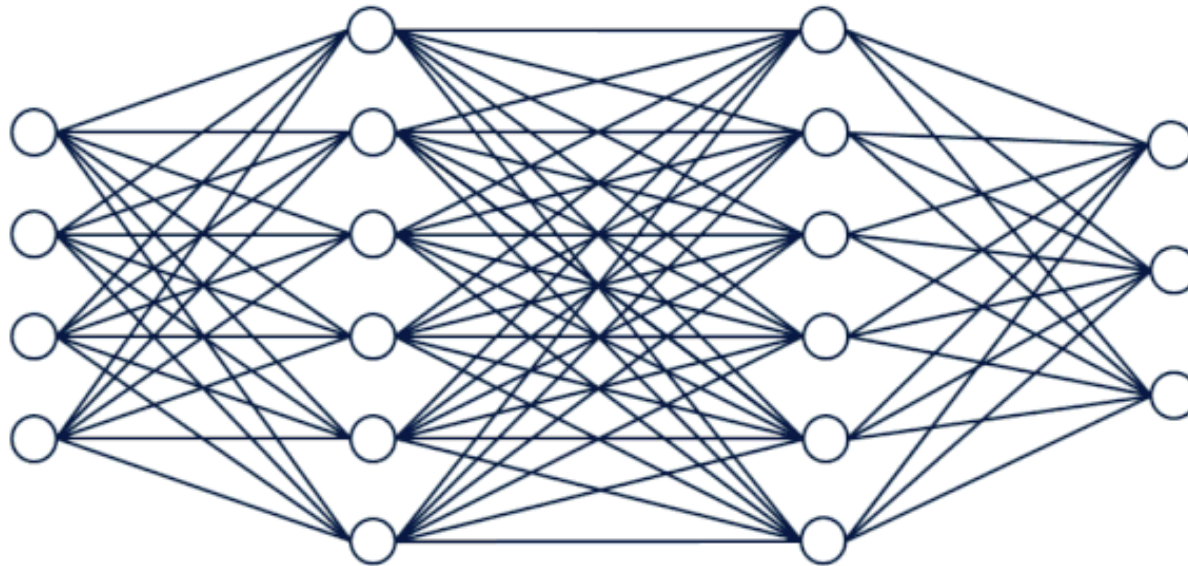
- Se $f_{O1} > f_{O2}$, então a classificação será o label de O1
 - Caso contrário, a classificação será o label de O2



Feed-Forward: Entendendo o Treinamento da Rede

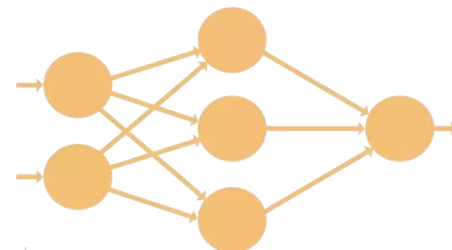
- Treinamento é a etapa na qual dizemos que a rede "aprende"
 - O valor agregado (aprendizado) pela computação realizada no treinamento
 - ✓ se materializa nos valores encontrados para os pesos (w) e para os bias (b)
 - que juntamente com a função de ativação, conseguem replicar de forma aproximada
 - uma função matemática que resolve o problema em questão
- Os valores dos pesos (w) e dos bias (b)
 - São iterativamente melhorados utilizando o algoritmo
 - ✓ Do Gradiente Descendente (ou outro semelhante)
 - O vetor gradiente (que aponta para a direção crescente da função)
 - é calculado por um algoritmo chamado Backpropagation

Exemplo de uma rede em funcionamento (1)



Analogia ao Treinamento de uma rede neural

- Encontrar os mínimos ajustes (pesos)
 - Da mesa de som (rede neural) que
 - ✓ geram o som (resultado) desejado



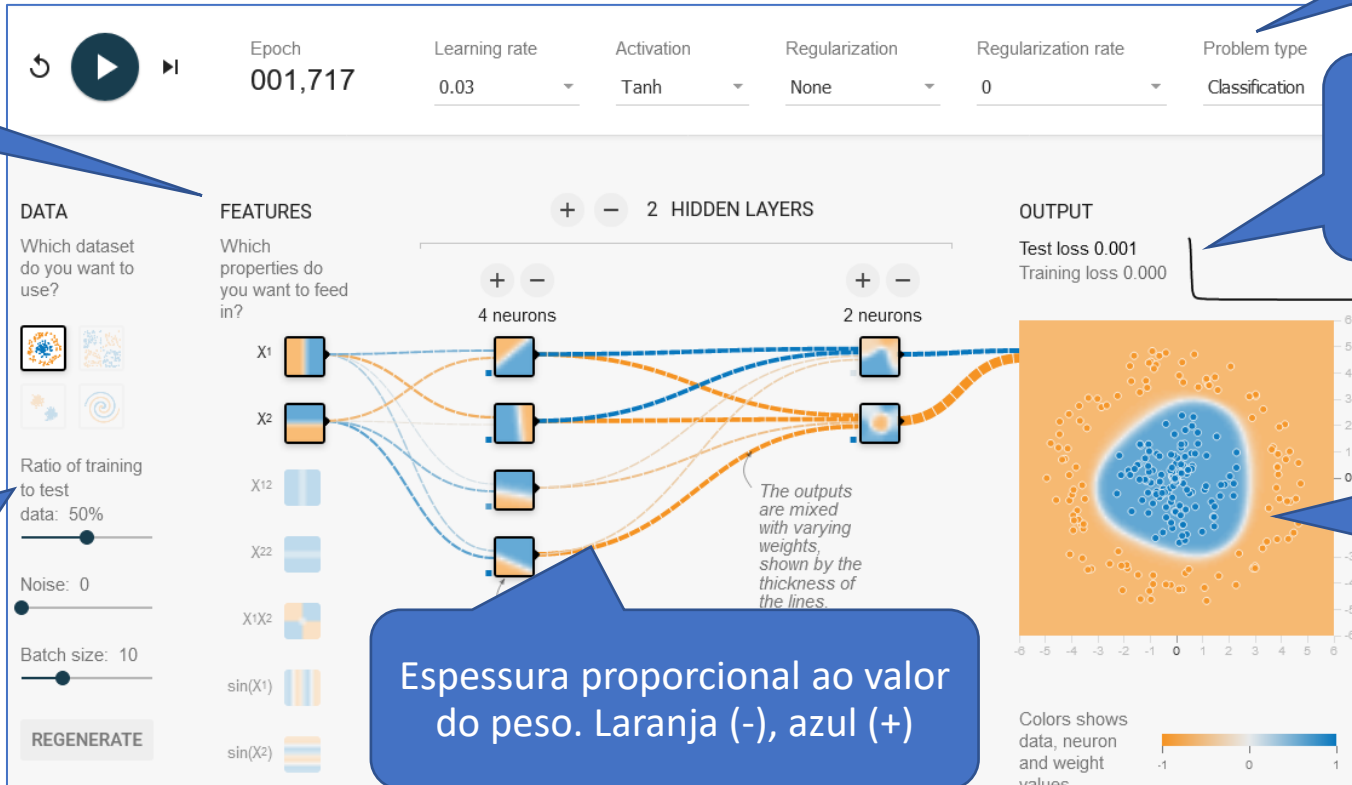
Playground para Testes de parâmetros de Redes Neurais

- <https://playground.tensorflow.org>
 - Mais vocabulário para entender os parâmetros

Features

Dataset

Proporção
dados
Treinamento /
Teste



Tipo de Problema

Erro de
treinamento e de
Generalização

Resultado da
Decision
Boundary

Espessura proporcional ao valor
do peso. Laranja (-), azul (+)

Procedimento simplificado de avaliação

- Particiona-se o Dataset em Treinamento e Teste
 - Por exemplo: 80%/20% ou 70%/30%
 - ✓ Não há valor ideal. Depende do problema. Empírico.
 - Também conhecido como Método Holdout
- Garante-se que o modelo aprende e **generaliza para dados não vistos**
 - Avaliar o seu modelo sobre o dado de treinamento não é uma boa prática
 - ✓ Equivalente a avaliar o aprendizado escolar com uma prova igual a lista de exercícios



Preparação dos dados para construção de Modelos de ML

Garbage-in, garbage-out

A qualidade da saída é determinada pela qualidade da entrada

Operações básicas

- Remover registros duplicados
- Tratar valores N/A (missing values)
 - Remover linhas (observações)
 - Remover colunas (atributos/features)
 - Inutar dados gerados/criados
 - ✓ Média, mediana, moda, vizinhos mais próximos (KNN)

Insert missing records Replace with 0 Replace with last known value Replace with mean Interpolate based on splines Remover linhas/colunas

	DATE	air_mv	air_mv_zero	air_mv_previous	air_mv_mean	air_expand
1	JAN49	112	112	112	112	112
2	FEB49	118	118	118	118	118
3	MAR49	132	132	132	132	132
4	APR49	129	129	129	129	129
5	MAY49		0	129	284.54385965	128.29783049
6	JUN49	135	135	135	135	135
7	JUL49		0	135	284.54385965	144.73734152
8	AUG49	148	148	148	148	148

Fonte da imagem

Feature Scaling (Reescalonamento de Atributos)

- Motivação: melhorar o desempenho do modelos
 - Alguns modelos não funcionam bem sem reescalar (Ex.: cálculo de distâncias)
- O que significa reescalar os atributos?
 - Colocar todos os atributos numa mesma escala (Ex.: no intervalo 0 a 1)

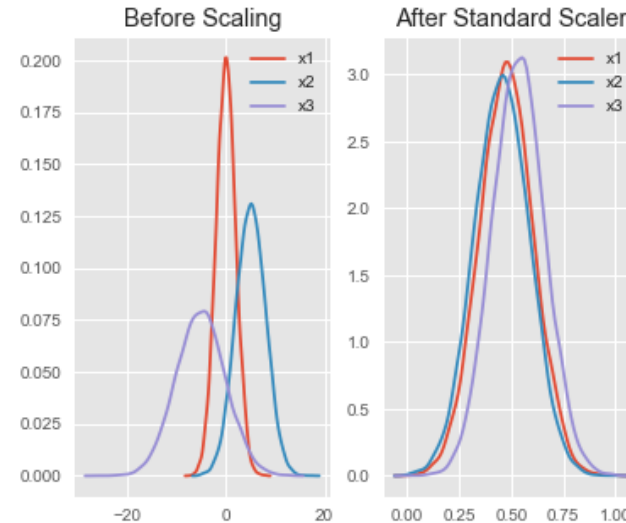
✓ Exemplos

Padronização
Standard

$$z = \frac{x - \mu}{s}$$

Normalização

$$z = \frac{x - x_{min}}{x_{max} - x_{min}}$$

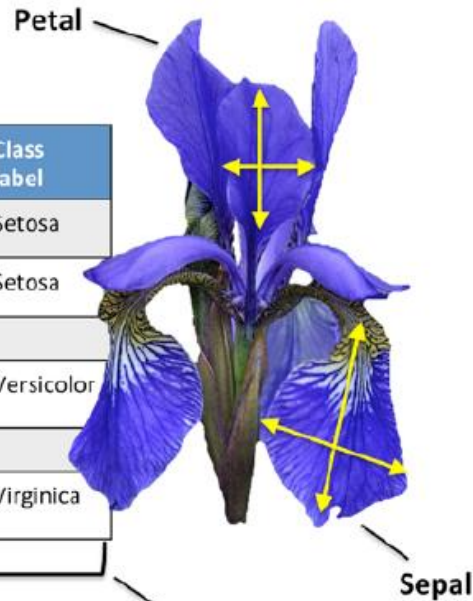


Experimento com o dataset Iris

Samples
(instances, observations)

	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

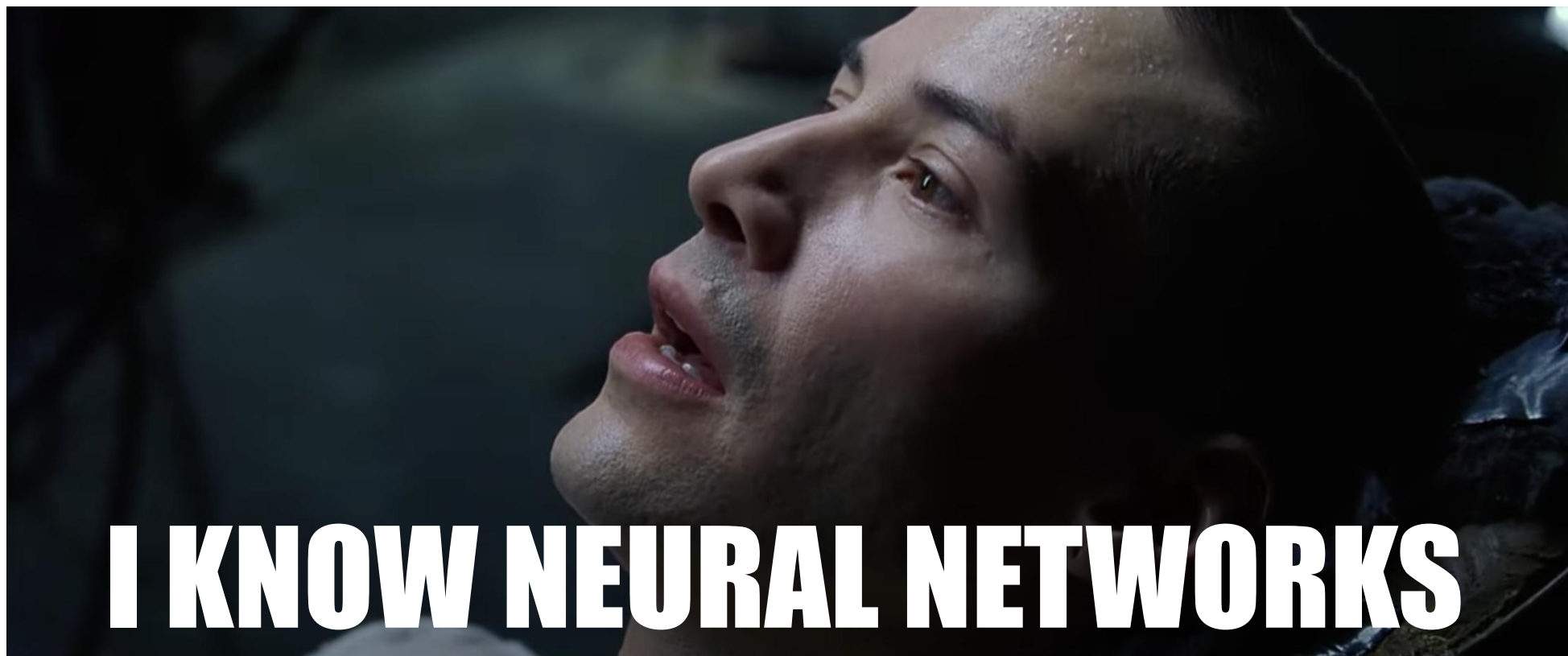
Features
(attributes, measurements, dimensions)



Class labels
(targets)

Três tipos de flores
Virginiza, Setosa e versicolor

Notebook



Preparação dos dados para construção de Modelos de ML

Garbage-in, garbage-out

A qualidade da saída é determinada pela qualidade da entrada

Codificação de Atributos (Feature Encoding)

- Sinônimo: One-hot encoding
- Transformar atributos categóricos em numéricos

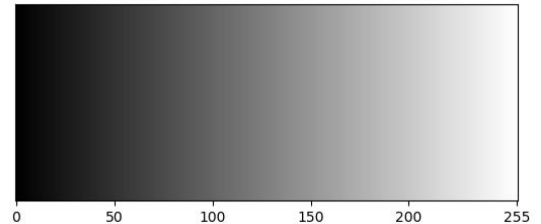
cpf	nome	salario	departamento
98654345821	Antônio Freitas	15000	Vendas
78621232185	Alessandra Rezende	10000	Produção
89632189251	Amanda Silva	8000	RH
21326478932	Anderson Pereira	12000	Produção



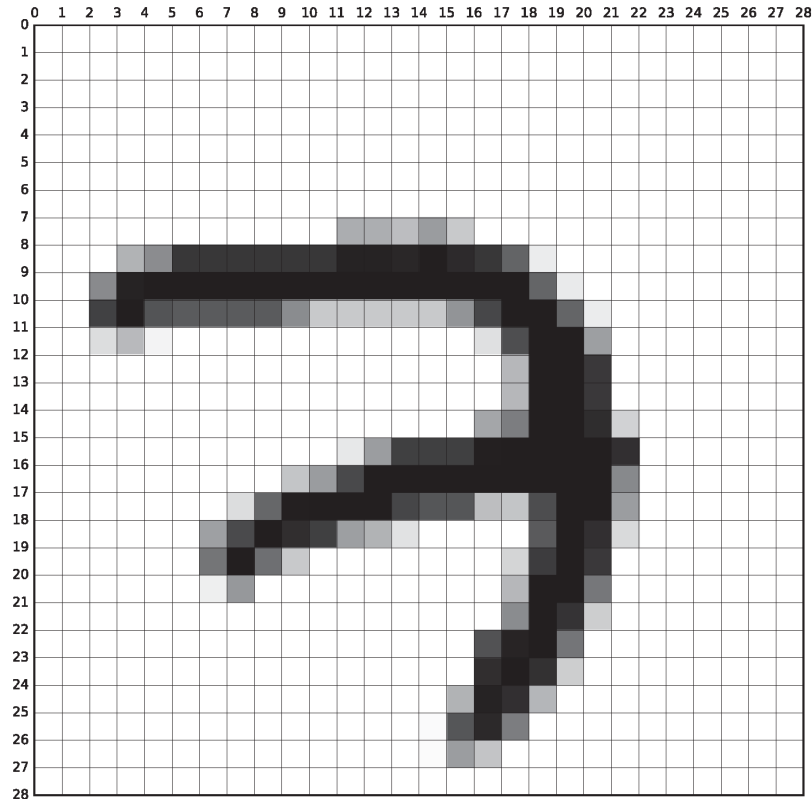
cpf	nome	salario	Vendas	Produção	RH
98654345821	Antônio Freitas	15000	1	0	0
78621232185	Alessandra Rezende	10000	0	1	0
89632189251	Amanda Silva	8000	0	0	1
21326478932	Anderson Pereira	12000	0	1	0

Dataset MNIST

- Base de dados de dígitos manuscritos
 - A MNIST contém 60.000 imagens de treino e 10.000 imagens de teste de dígitos manuscritos.
 - O conjunto de dados inclui imagens em escala de cinzentos de 28x28 pixéis.
 - As imagens são normalizadas para caberem numa caixa delimitadora de 28x28 pixels
 - ✓ níveis de escala de cinza (0 a 255)
- Em 1989, Yann LeCun (Meta)
 - Aplicou redes neurais e backpropagation neste problema
 - ✓ Que mais tarde deu origem ao MNIST.



Dataset MNIST (Modified National Institute of Standards and Technology)

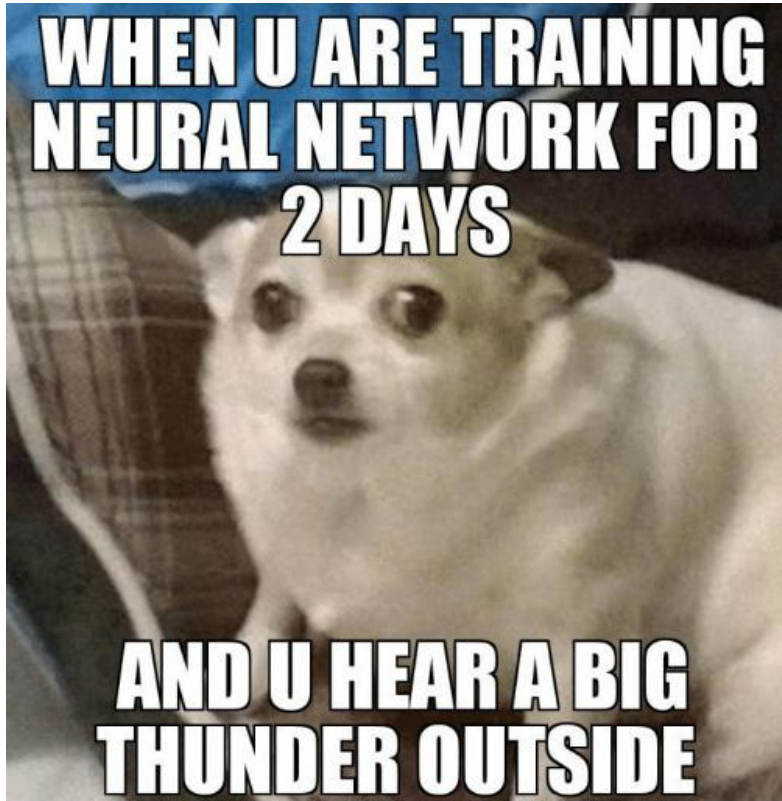


(a) MNIST sample belonging to the digit '7'.



(b) 100 samples from the MNIST training set.

Se entender a piada é porque já está falando a língua dos nerds



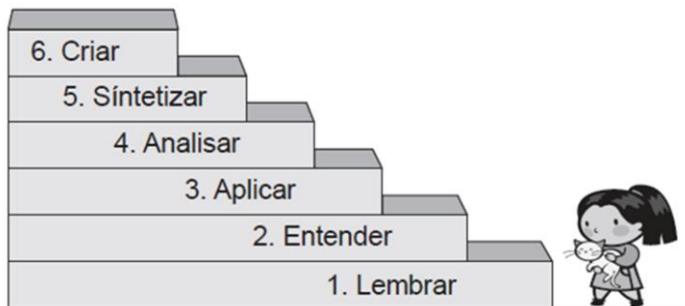
[Fonte da Imagem](#)



[Fonte da Imagem](#)

Se entender a piada é porque já está falando a língua dos nerds

Taxonomia de Bloom



Novo teste para substituir o Teste de Turing



[Fonte da Imagem](#)