Modelos Preditivos Conexionistas

Vitor Casadei

Especialização em Dados - 2022.02

Outline

Definição de Machine Learning

Arthur Samuel(1959)

Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.

Definição de Machine Learning

Tom Mitchell(1998)

Well posed Learning Problem: A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

Definição de Machine Learning

Tom Mitchell(1998)

Well posed Learning Problem: A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

Spam Classification:

- E: Watching you label emails as spam or not spam
- T: Classifying emails as spam or not spam
- P: The number (or fraction) of emails correctly classified as spam/not spam

Aprendizado Supervisionado

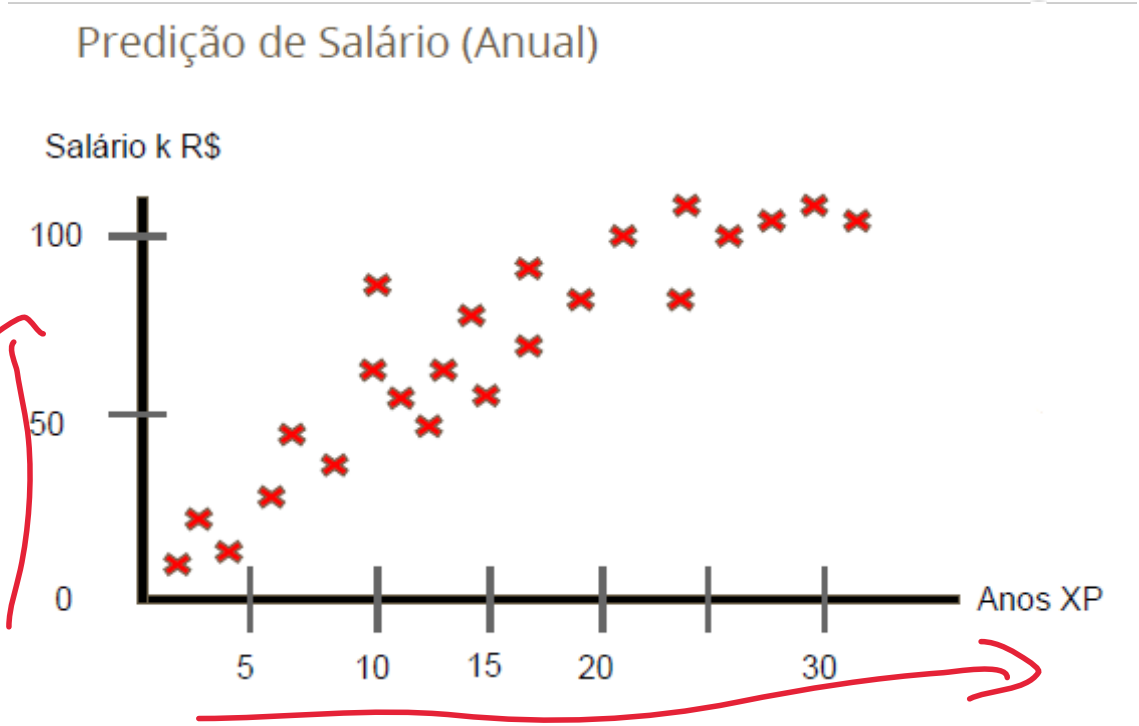


Figure: Exemplo de Aprendizado Supervisionado

Aprendizado com base nas "respostas certas"

Regressão: Predição de valores contínuos

Aprendizado Supervisionado

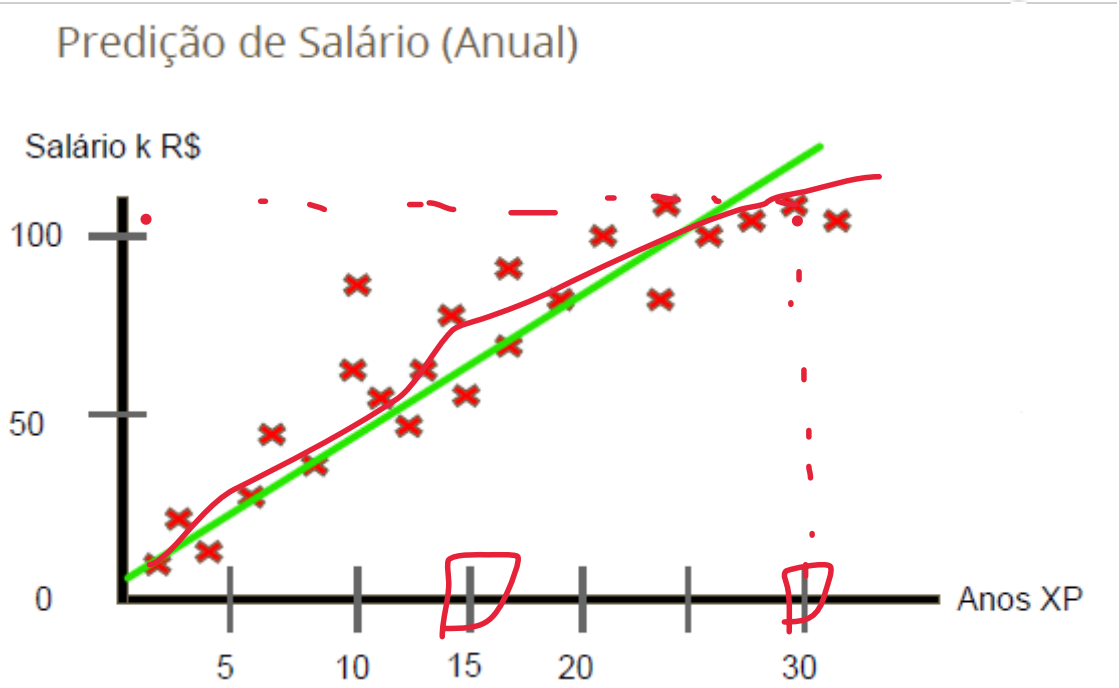
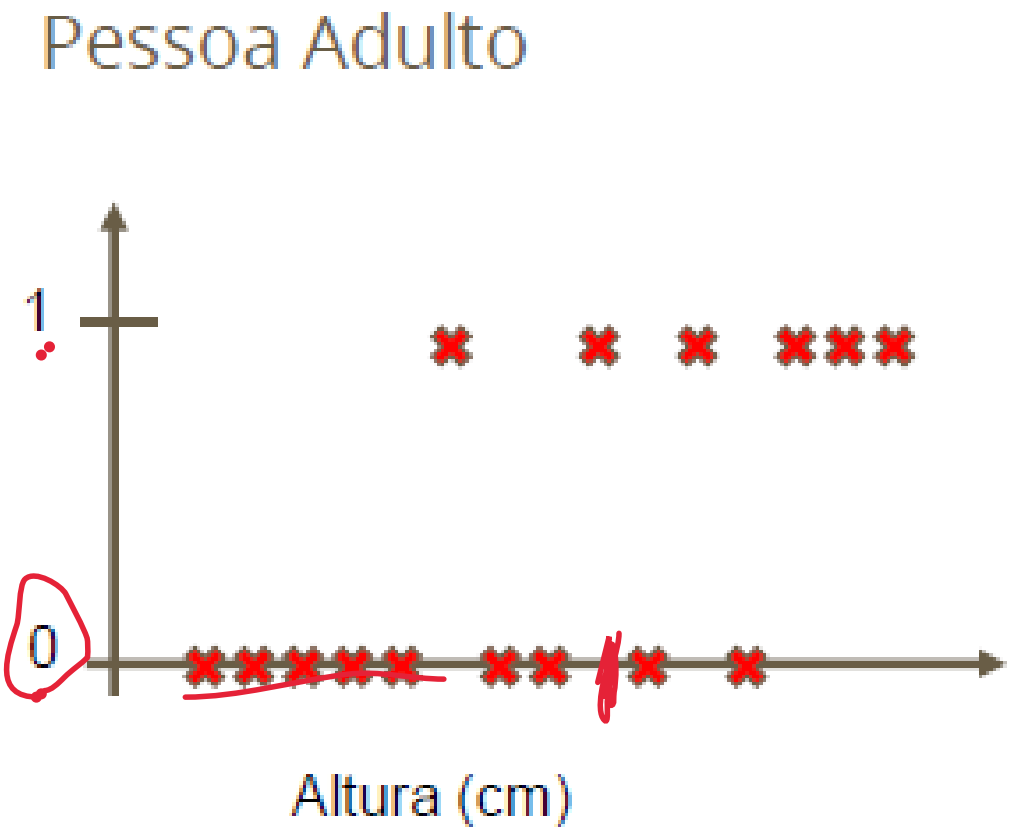


Figure: Exemplo de Aprendizado Supervisionado

Aprendizado com base nas "respostas certas"

Regressão: Predição de valores contínuos

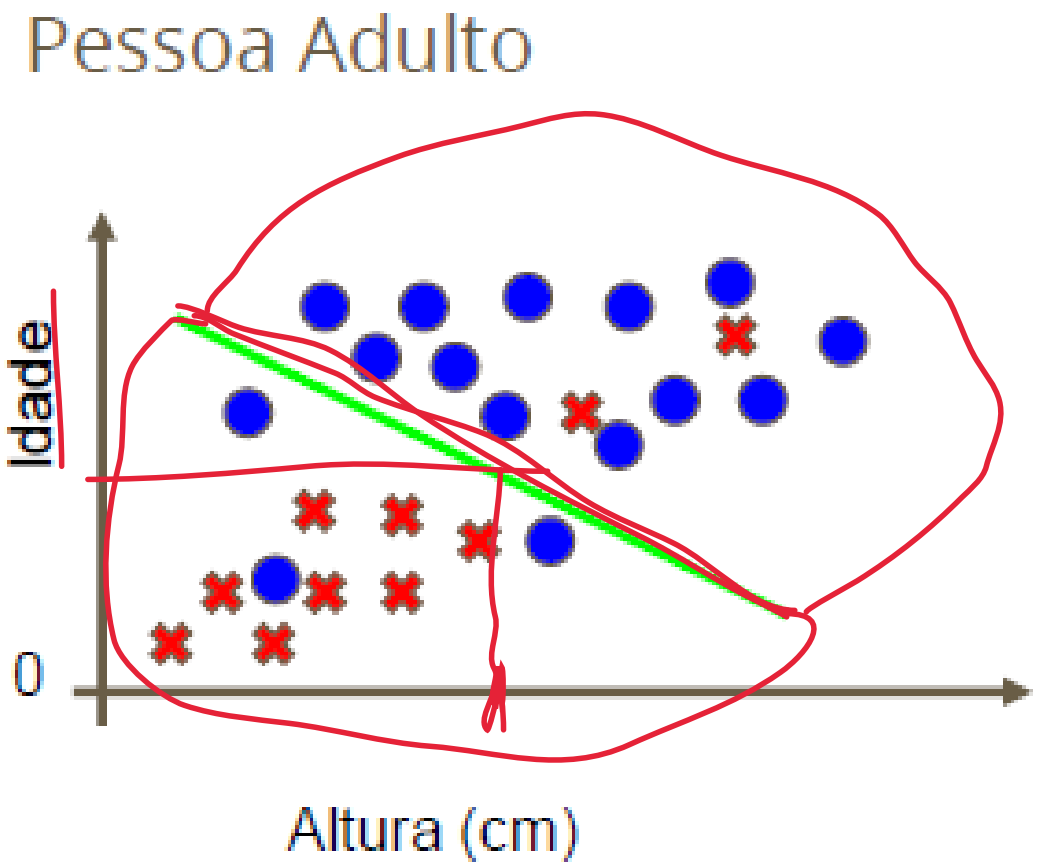
Aprendizado Supervisionado



Classificação: **Predição** de valores discretos

Figure: Exemplo de Aprendizado Supervisionado

Aprendizado Supervisionado



Classificação: **Predição** de valores discretos

Figure: Exemplo de Aprendizado Supervisionado

Aprendizado Não Supervisionado

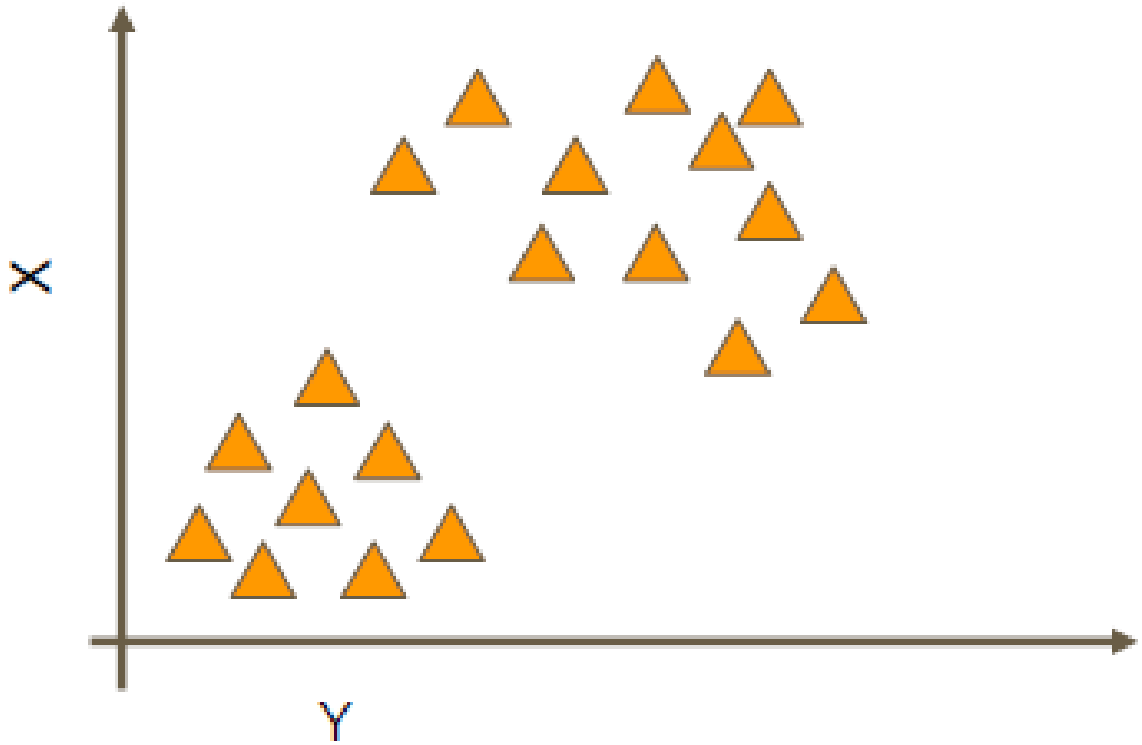
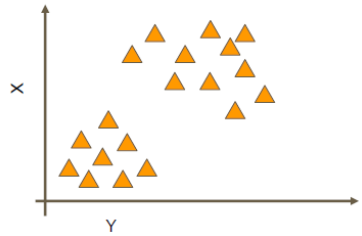


Figure: Exemplo de Aprendizado Não Supervisionado

Aprendizado Não Supervisionado - Exemplos

- An advertising platform segments the U.S. population into smaller groups with similar demographics and purchasing habits so that advertisers can reach their target market with relevant ads.
- Airbnb groups its housing listings into neighborhoods so that users can navigate listings more easily.



Aprendizado Não Supervisionado - Exemplos

Research

Interaction With Mobile Devices by Elderly People: The Brazilian Scenario
Ricardo Leme, Luciana Zaina, and Vitor Casadei ACHI 2014 Barcelona

Marina

Retired woman, 80 years.

"I never had a cell phone!"

Although they enjoy good health, due to the limitations imposed by age, Marina depends on the family to perform basic tasks such as call to the grandchildren or watch a DVD. Regrets are just not having completed primary education.
Carefree, proud to say I never laid a hand on a computer.
Despite the insistence of the children, is reluctant to have a mobile device by consider it impossible to use.

Personal goals:
Enjoy life with their grandchildren. If you need the technology they can help me.

Practical goals:
The devices are too complicated to operate. I'm afraid to use them.

Manoel

Retired man, 70 years.

"Technology is not just for the younger"

Has a college degree, is an active member of Third Age group and well respected by people.
When friends of Mr. Manoel has doubts about new technologies, he readily help because they know he can not live without your smartphone, with which ships with the latest news.

Personal goals:
Do not waste time with technology and enjoy it the best way possible.

Practical goals:
The text of the programs on the mobile device could be a little bigger.

Lucia

A worker retired, 60 years.

"I'm retired but I can't stay without working"

Lucia is a person who although already retired, as the value of benefit is not enough, continues to work.
A few months ago bought a mobile device and is asking all friends how to use it.
Regularly attend the computer classes and is finalizing the school.
Is becoming a fan of games on mobile.

Personal goals:
Learn more. Technology should support us constantly.

Practical goals:
At times I have trouble clicking on buttons, especially when the images are very small.

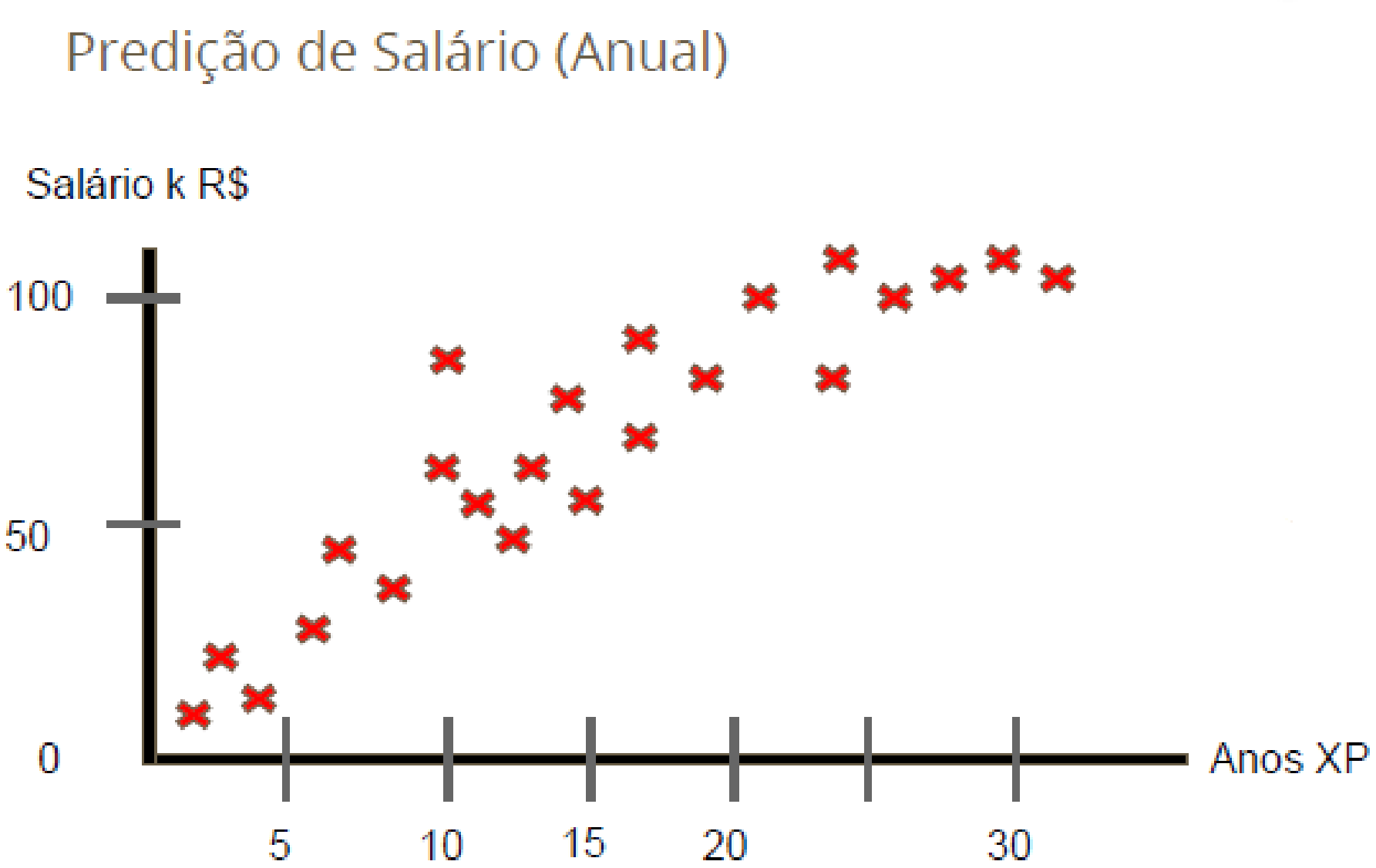


Figure: Exemplo de Aprendizado Não Supervisionado

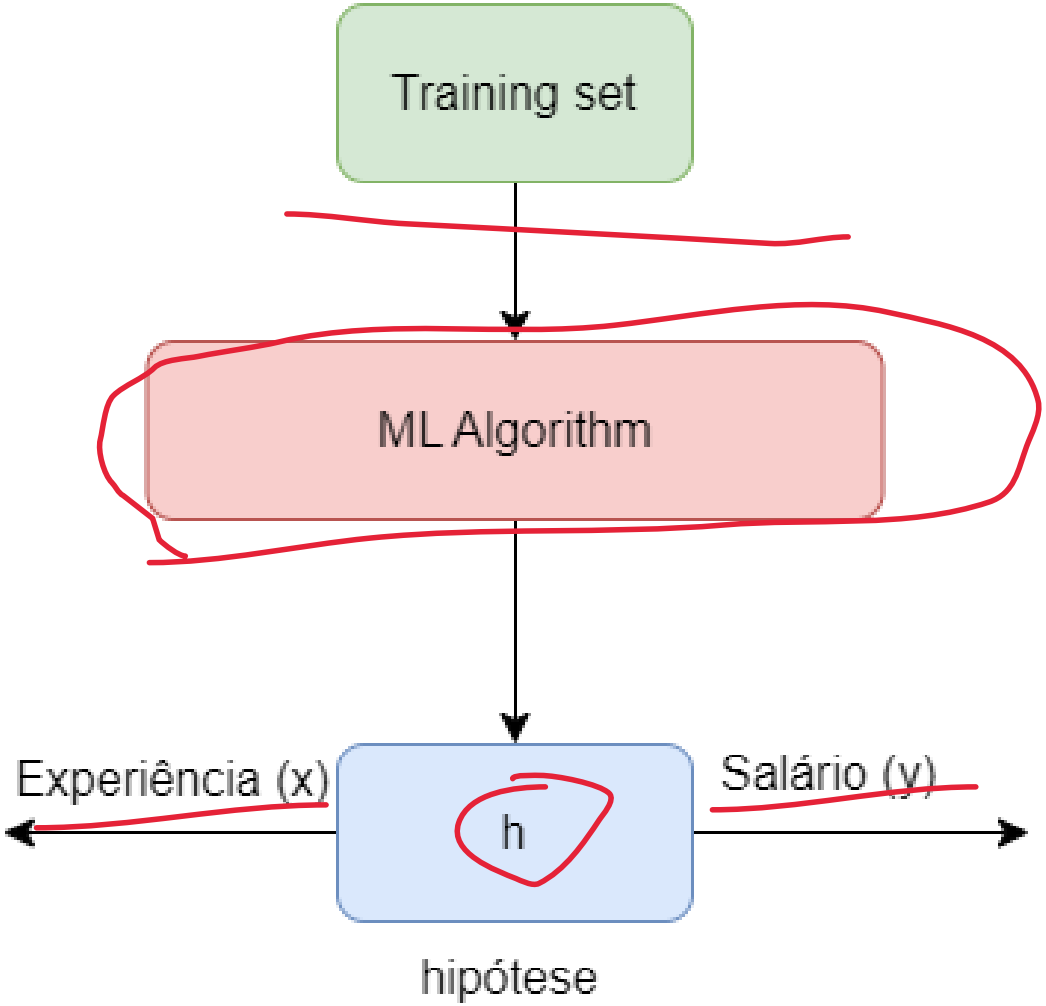
Regressão Linear - Uma variável

Notação:

- m = num. de exemplos de treinamento
- x = features
- y = target

Experiência (anos)	Salário Atual (K R$)
2	36
8	50
13	72
20	98
...	...

Regressão Linear - Treinamento



Regressão Linear - Função de Custo

Notação:

- m = num. de exemplos de treinamento
- x = features
- y = target

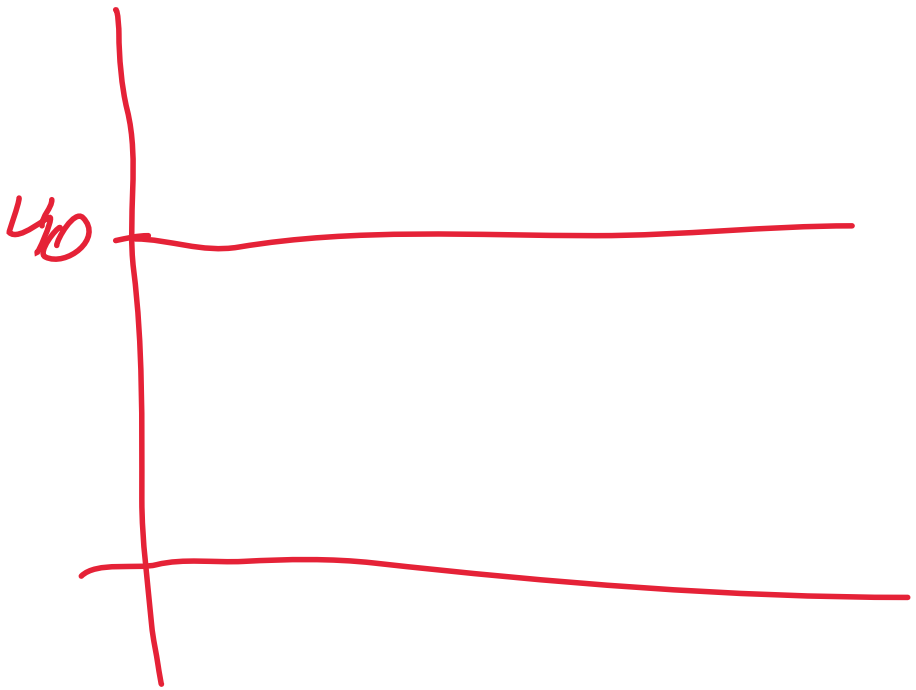
Experiência (anos)	Salário Atual (K R$)
2	36
8	50
13	72
20	98
...	...

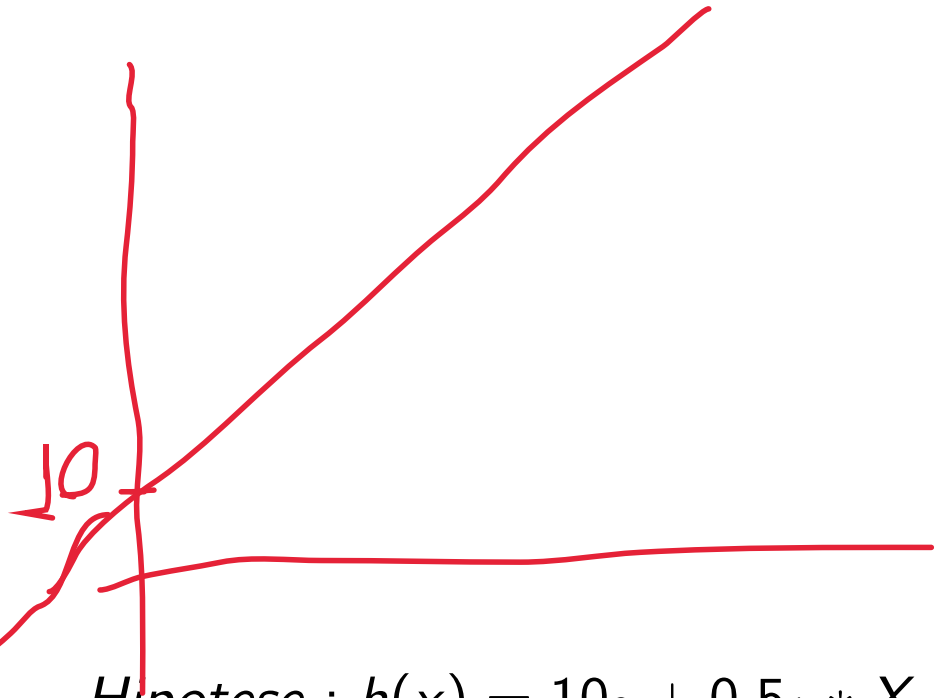
$$Hipotese : h(x) = \theta_0 + \theta_1 * X$$

$$\hat{Y}$$

Regressão Linear - Função de Custo

$$Hipotese : h(x) = \theta_0 + \theta_1 * X$$





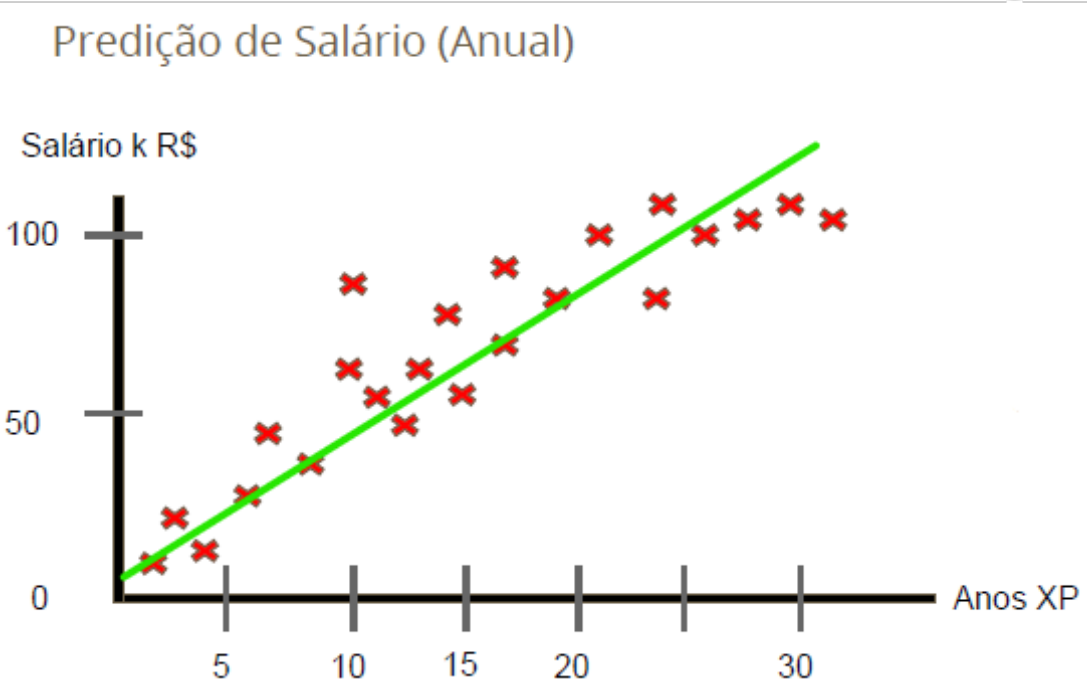
$$Hipotese : h(x) = 40_0 + 0_1 * X$$

$$Hipotese : h(x) = 10_0 + 0.5_1 * X$$

$$\theta_0 = 40 | \theta_1 = 0$$

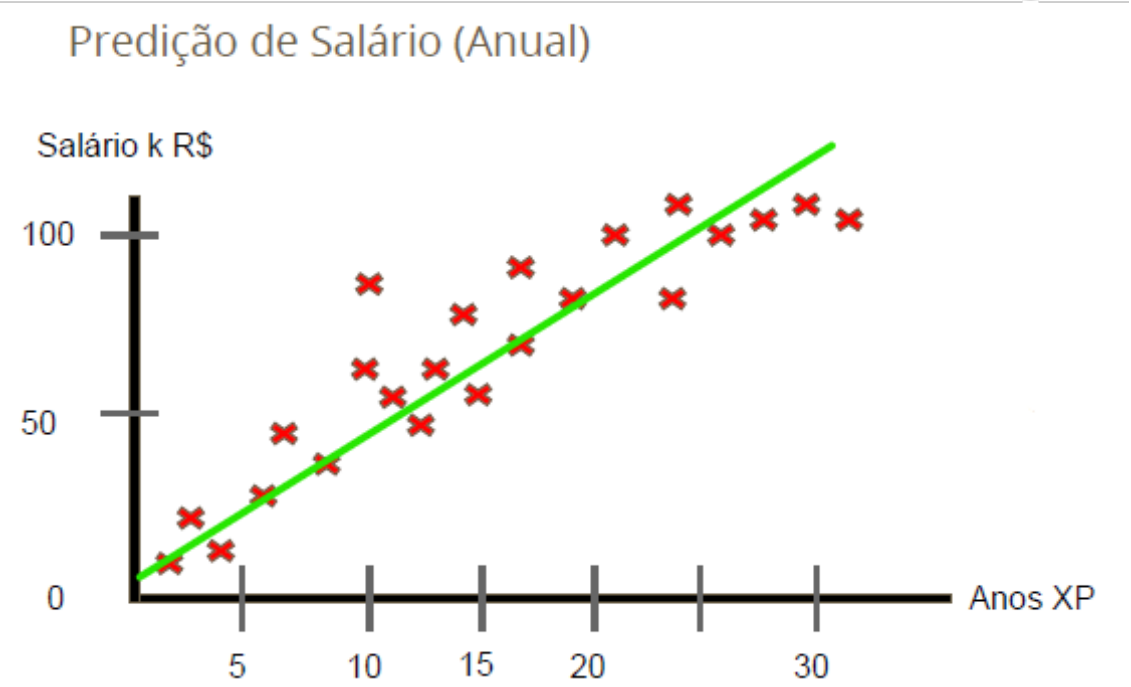
$$\theta_0 = 10 | \theta_1 = 0.5$$

Regressão Linear - Função de Custo



Escolher θ_0, θ_1 para que $h(x)$ seja perto de y com base no set de treinamento

Regressão Linear - Função de Custo



Escolher θ_0, θ_1 para que $h(x)$ seja perto de y com base no set de treinamento

$$Hipotese : h(x) = \theta_0 + \theta_1 * X$$

$$Minimize : \theta_0\theta_1$$

Mean Square Error Function

Regressão Linear - Gradiente Descendente

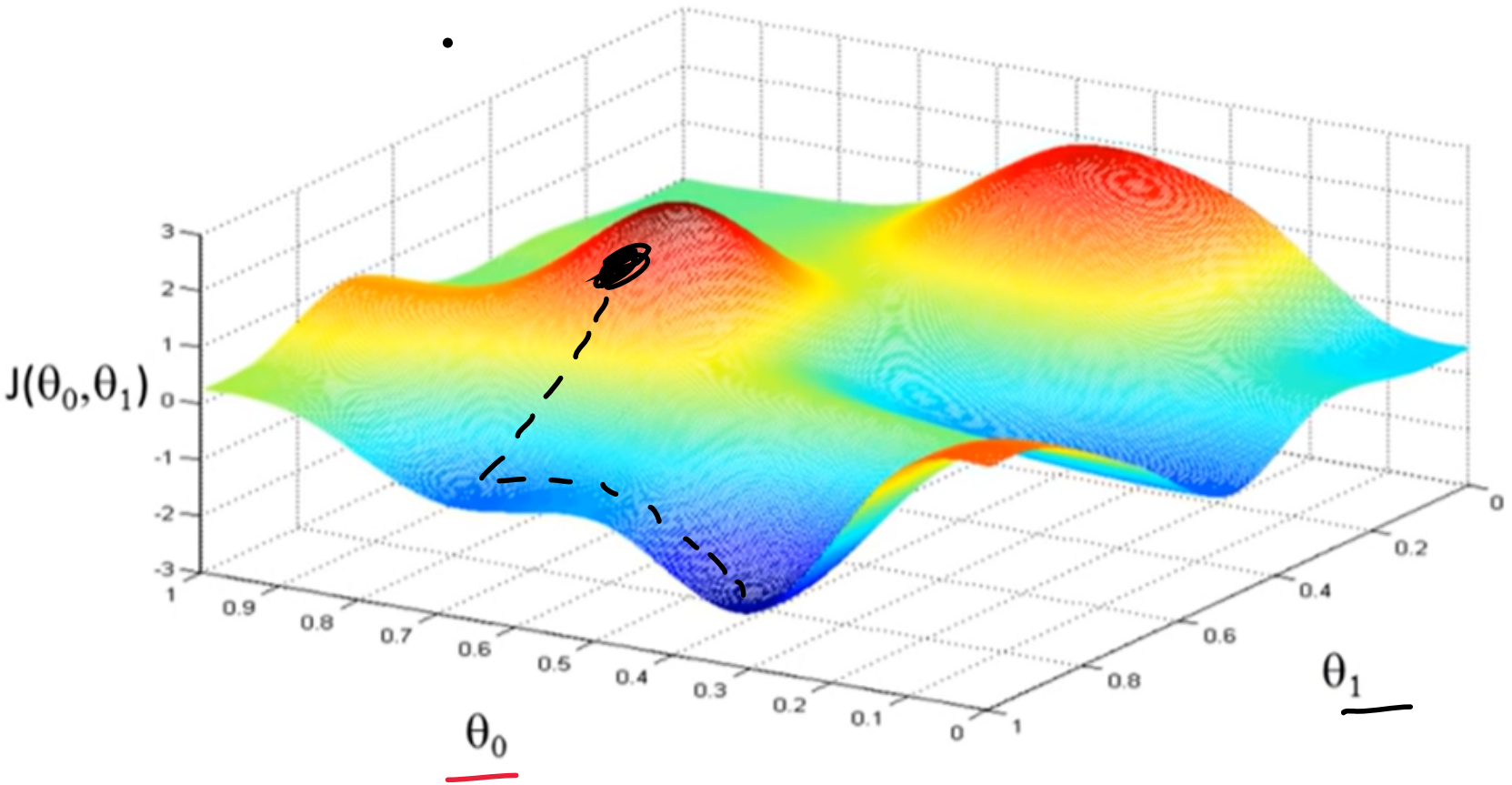
Objetivo

- Minimizar θ_0 e θ_1

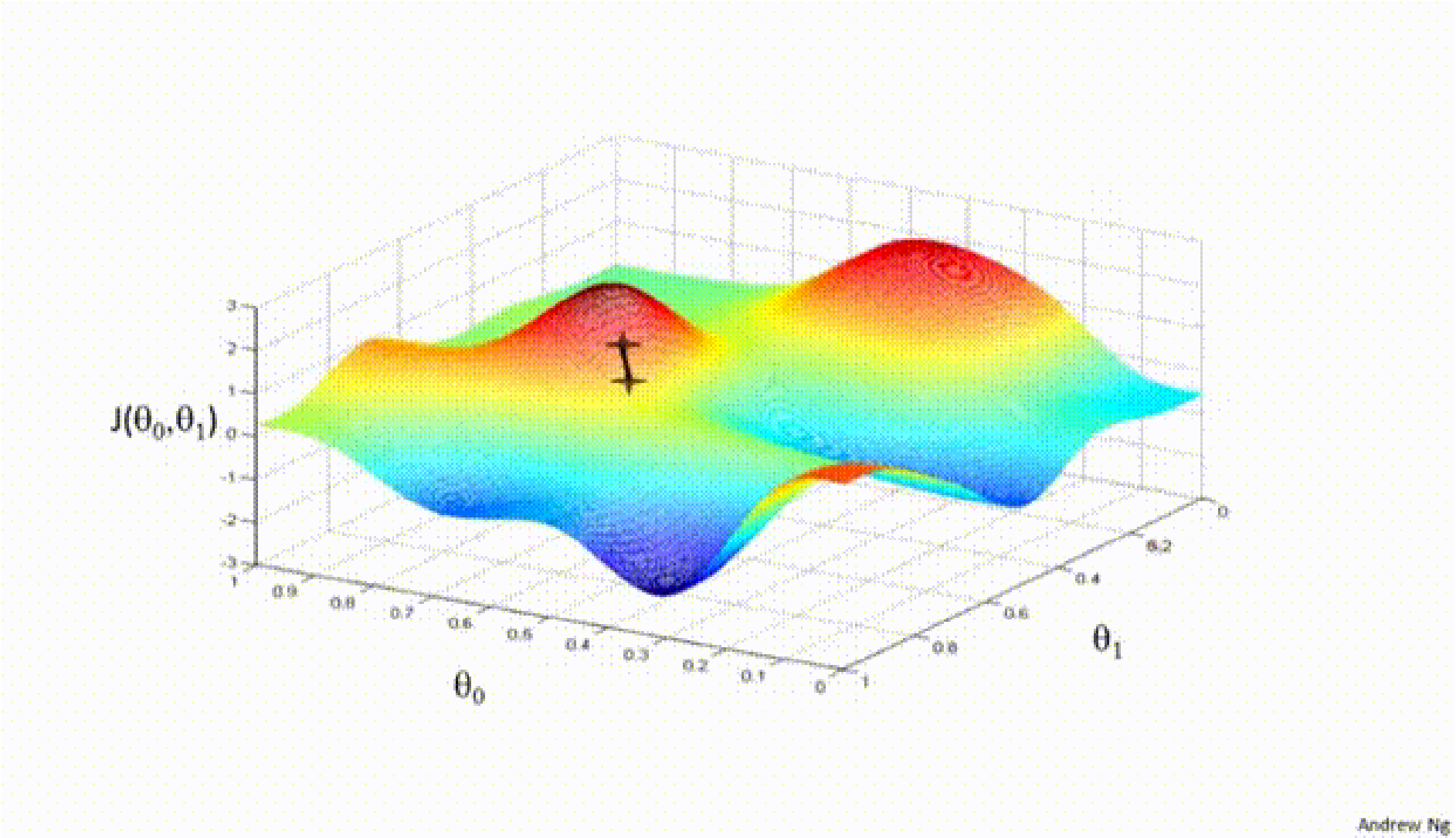
Como Fazer:

- Começar com um θ_0 e θ_1 qualquer
- Mudar os valores de θ_0 e θ_1 para reduzir o erro, até que se chegue a um mínimo

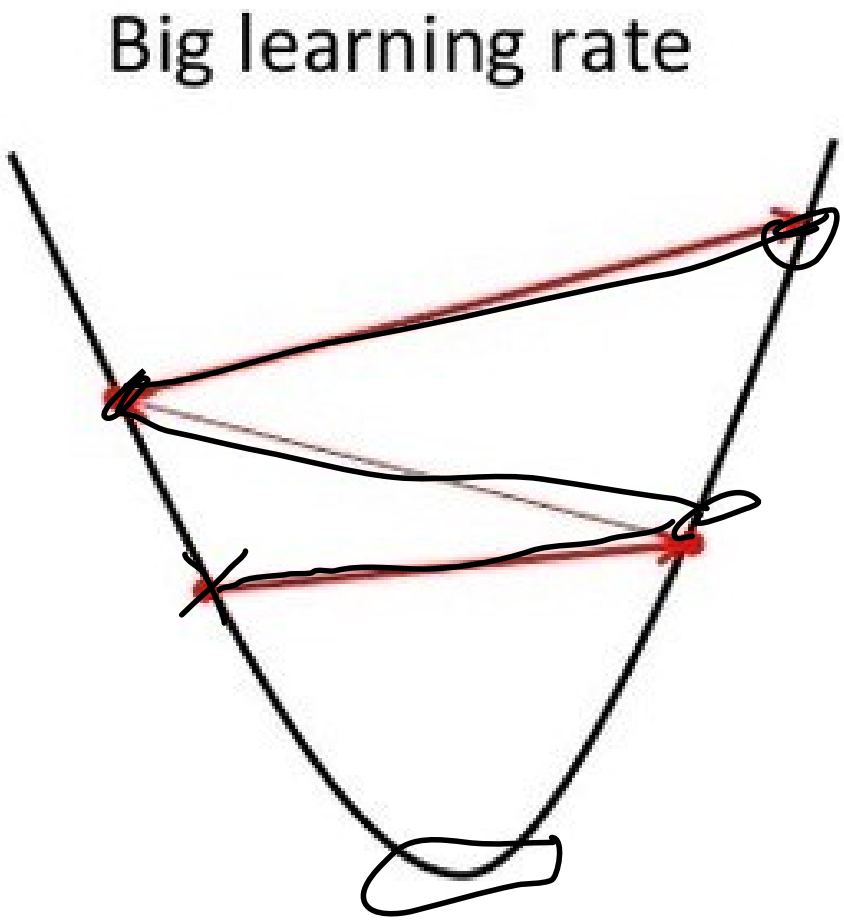
Regressão Linear - Gradiente Descendente

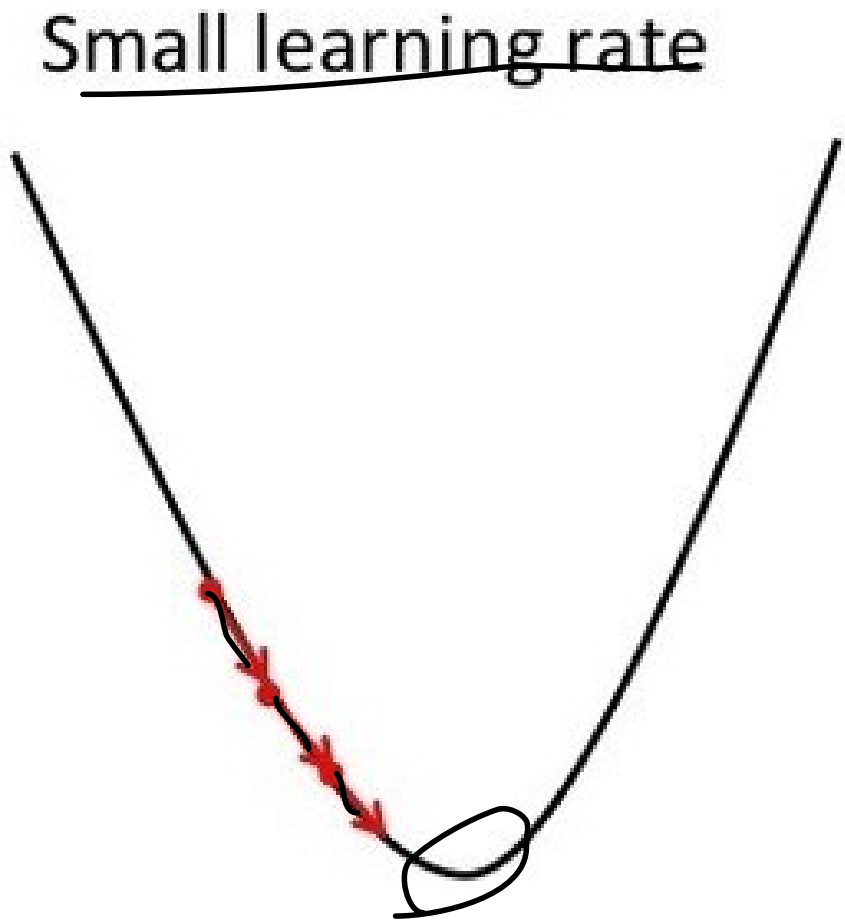


Regressão Linear - Gradiente Descendente



Regressão Linear - Gradiente Descendente





Regressão Logística

Classificação

- y é um valor discreto

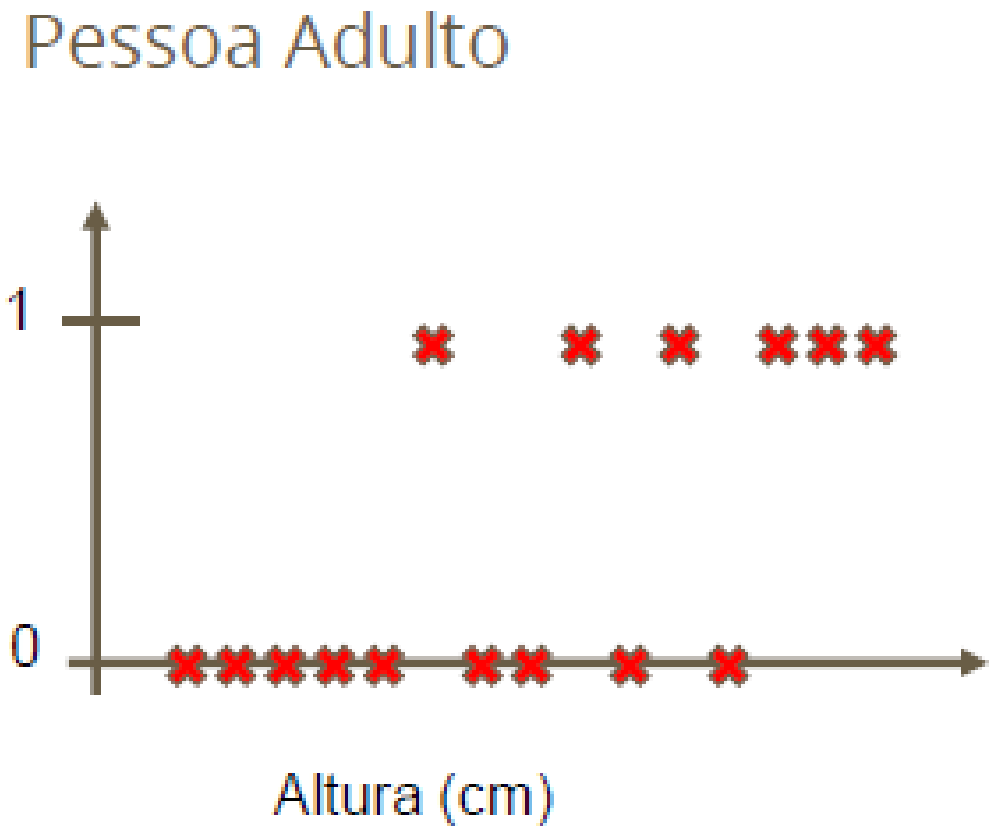
Exemplos:

- Email: Spam/Not Spam
- Online Transactions
- Tumor

Y

- is 0
- is 1

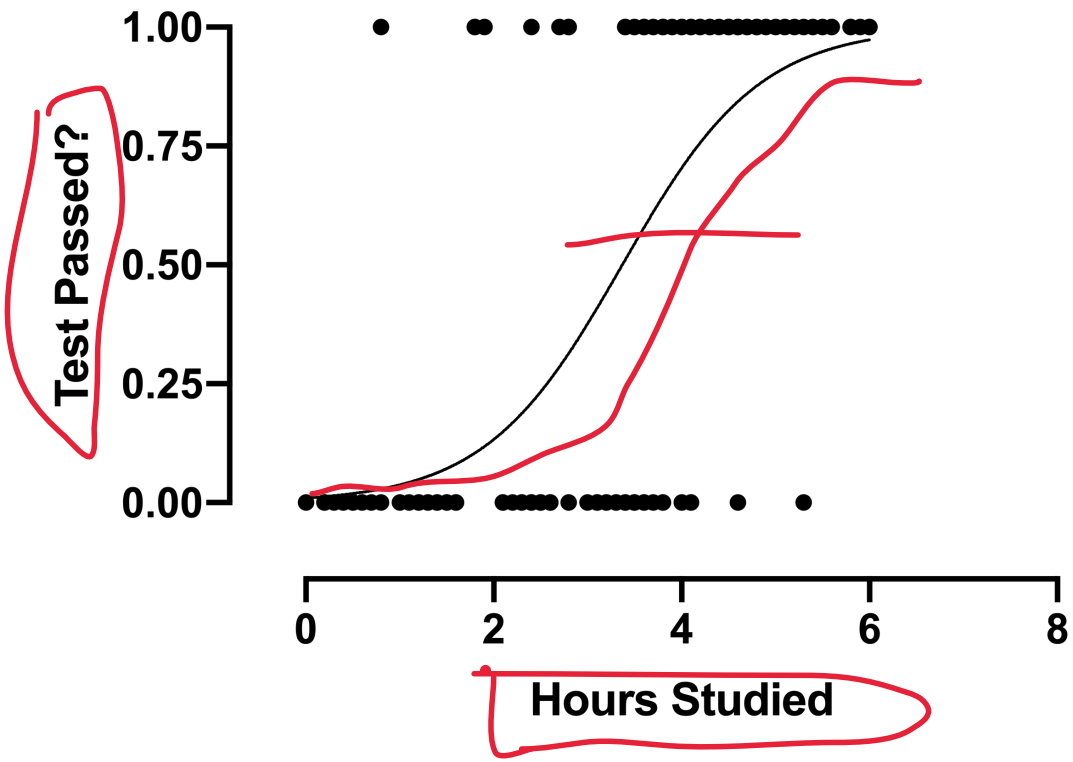
Regressão Logística



Classificação: **Predição** de valores discretos

Não é uma boa ideia usar Regressão Linear para esse tipo de problema

Regressão Logística



Objetivo

- $0 <= h(x) <= 1$
- $h(x) = g(\theta^T X)$
- Sigmoid Funcion

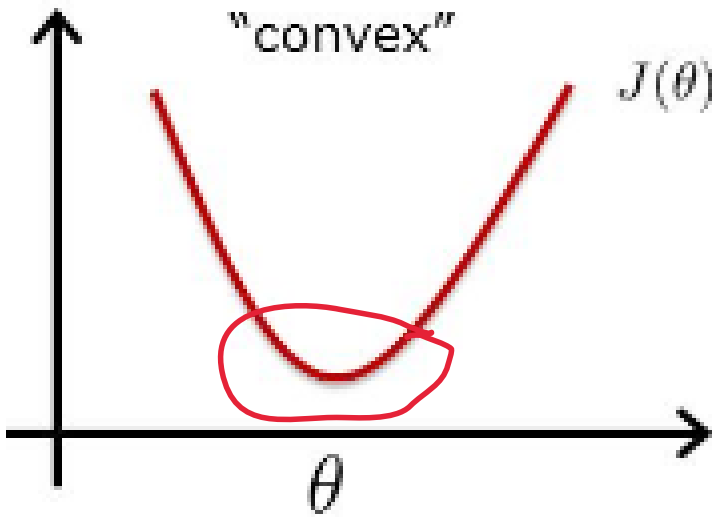
$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

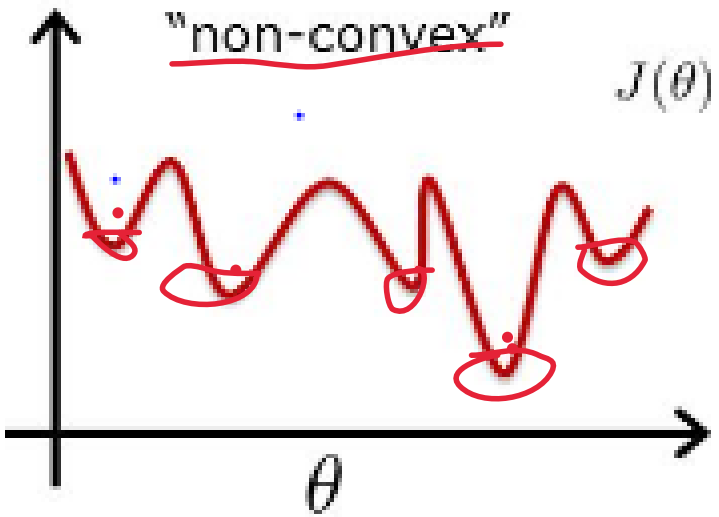
Regressão Logística - Função de Custo

Não podemos usar MSE

"because our prediction function is non linear (due to sigmoid transform). Squaring this prediction as we do in MSE results in a non convex function with many local minimums. If our cost function has many local minimums, gradient descent may not find the optimal global minimum."

Convex Vs Non-Convex





Regressão Logística - Função de Custo

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \text{Cost}(h_\theta(x^{(i)}), y^{(i)})$$

$$\text{Cost}(h_\theta(x), y) = -\log(h_\theta(x)) \qquad \text{if } y = 1$$

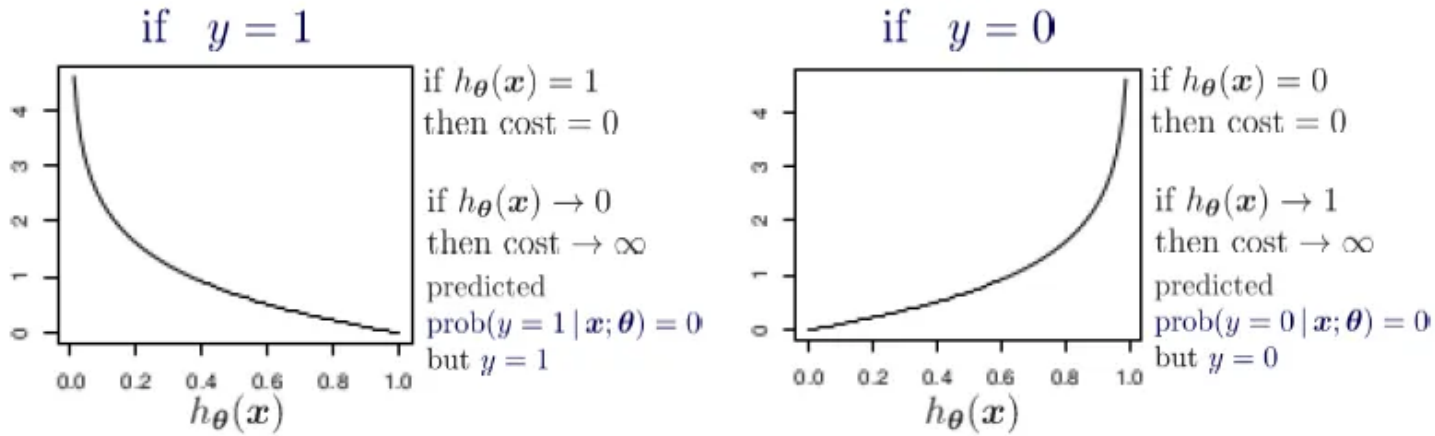
$$\text{Cost}(h_\theta(x), y) = -\log(1 - h_\theta(x)) \qquad \text{if } y = 0$$

Regressão Logística - Função de Custo

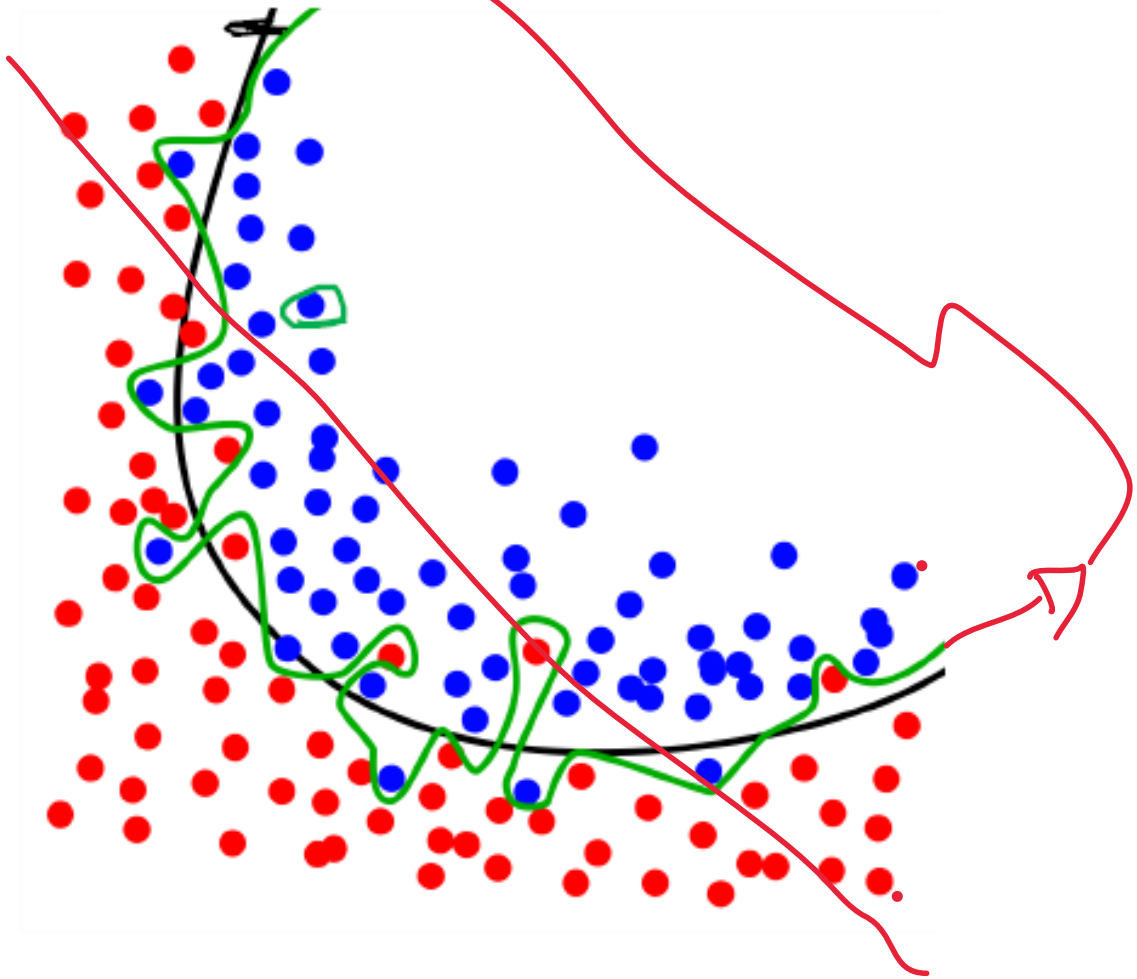
Cost Function for Logistic Regression

"The key thing to note is the cost function penalizes confident and wrong predictions more than it rewards confident and right predictions! The corollary is increasing prediction accuracy (closer to 0 or 1) has diminishing returns on reducing cost due to the logistic nature of our cost function."

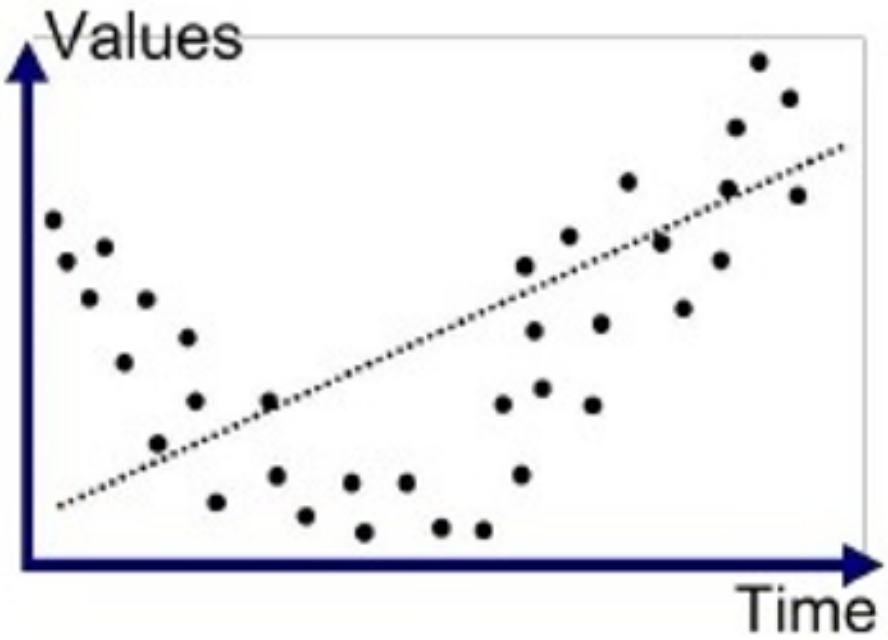
$$\text{cost}\left(h_{\boldsymbol{\theta}}(\boldsymbol{x}),\ y\right) = \begin{cases} -\log(h_{\boldsymbol{\theta}}(\boldsymbol{x})) & \text{if } y = 1 \\ -\log(1 - h_{\boldsymbol{\theta}}(\boldsymbol{x})) & \text{if } y = 0 \end{cases}$$



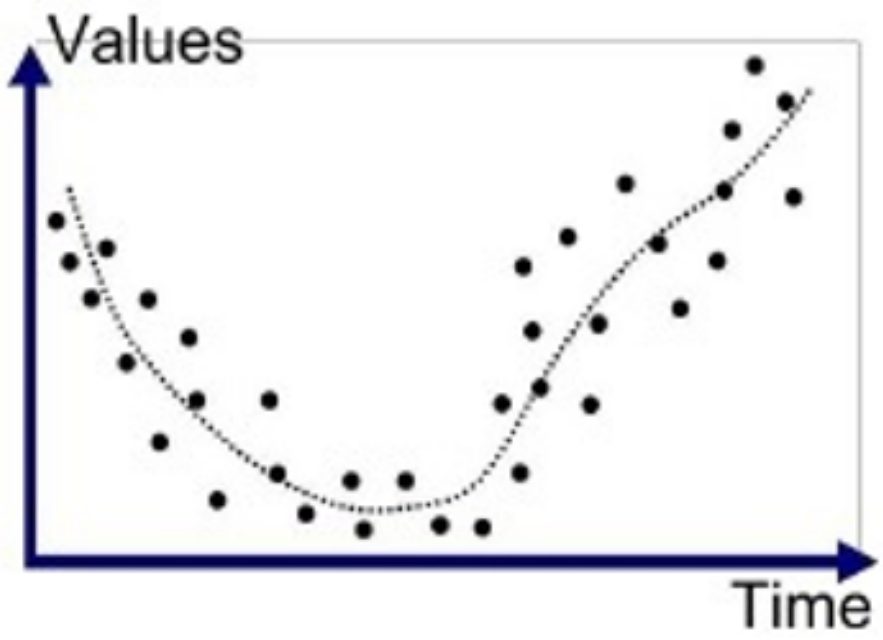
Overfitting e Underfitting



Overfitting e Underfitting

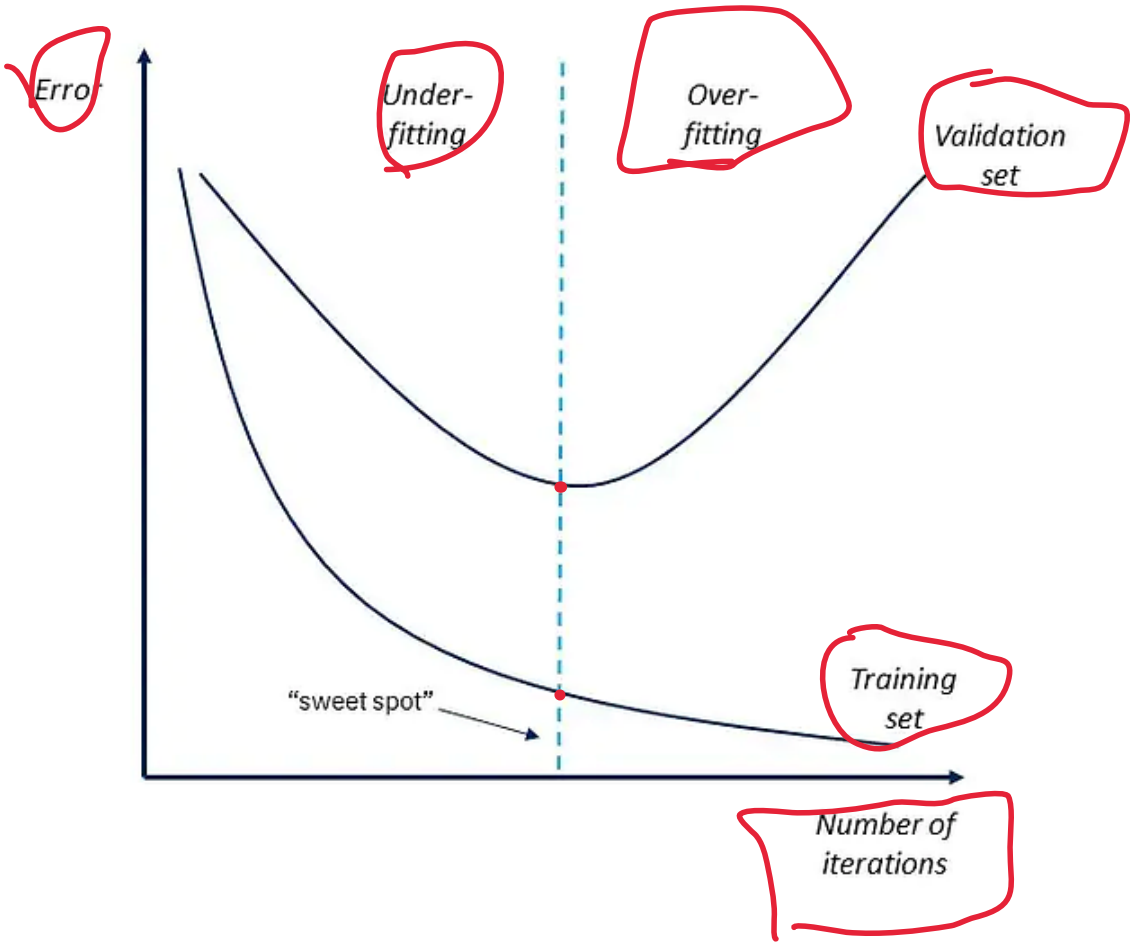


Underfitted

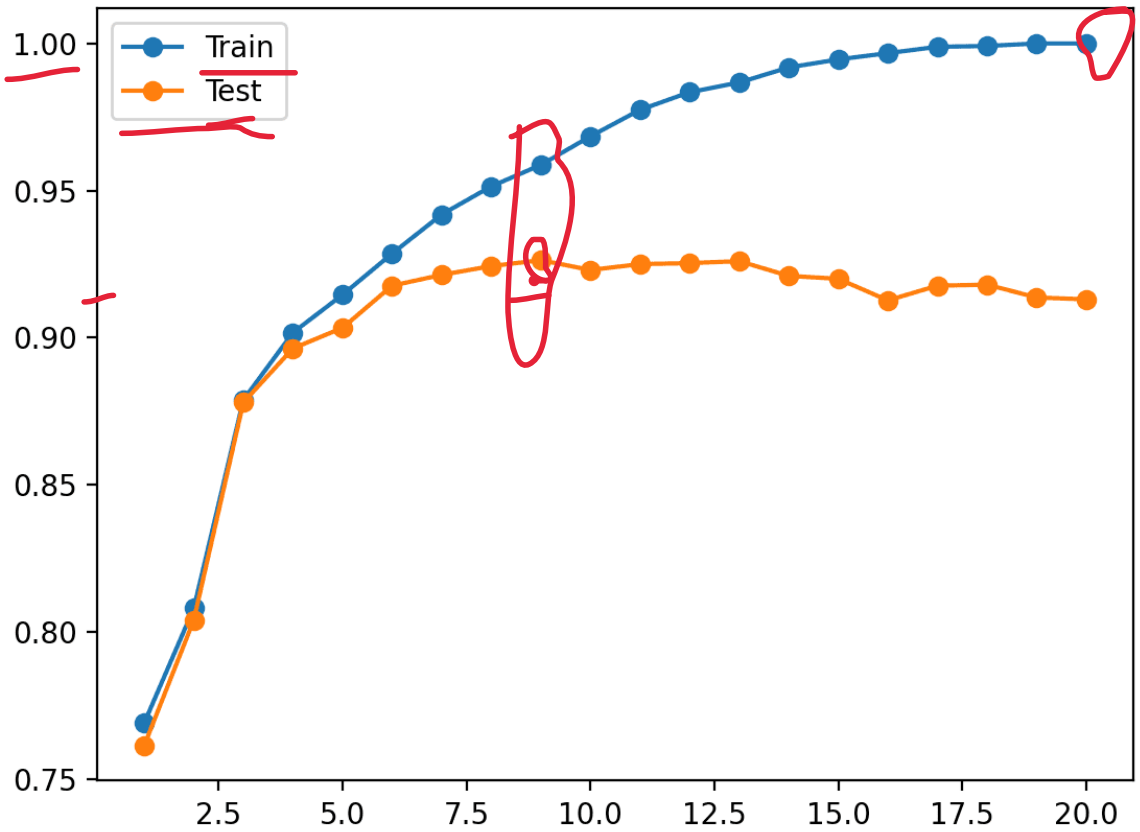


Good Fit/Robust

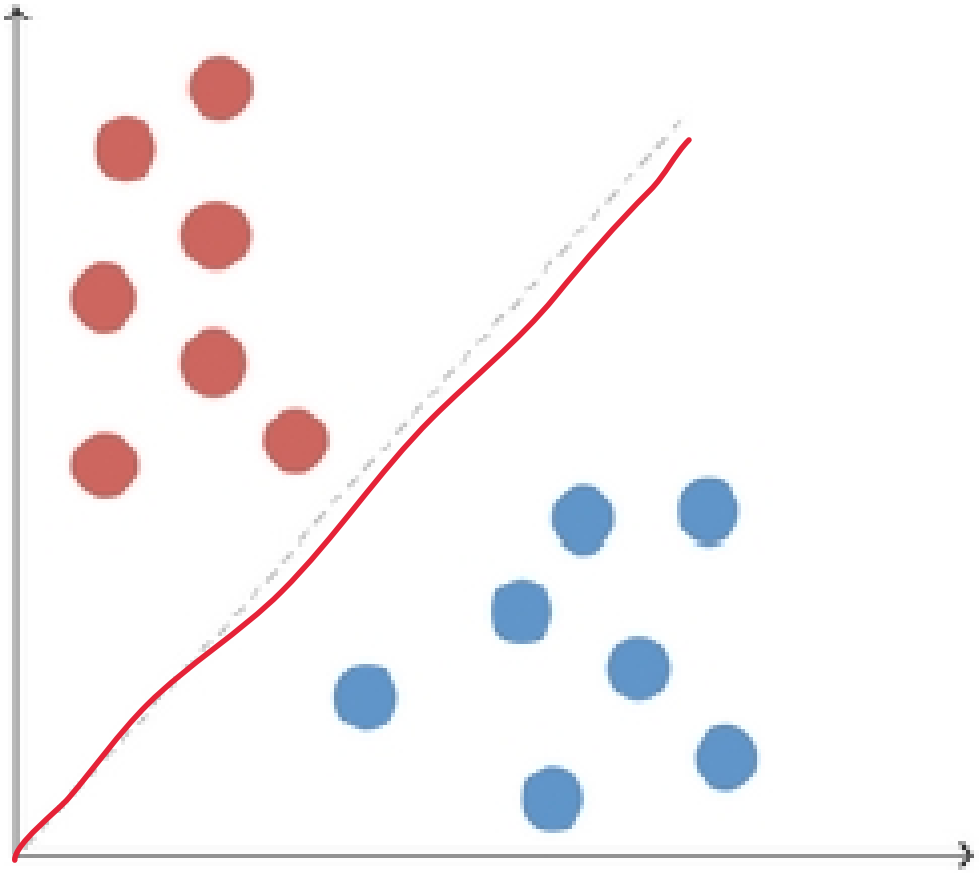
Overfitting e Underfitting

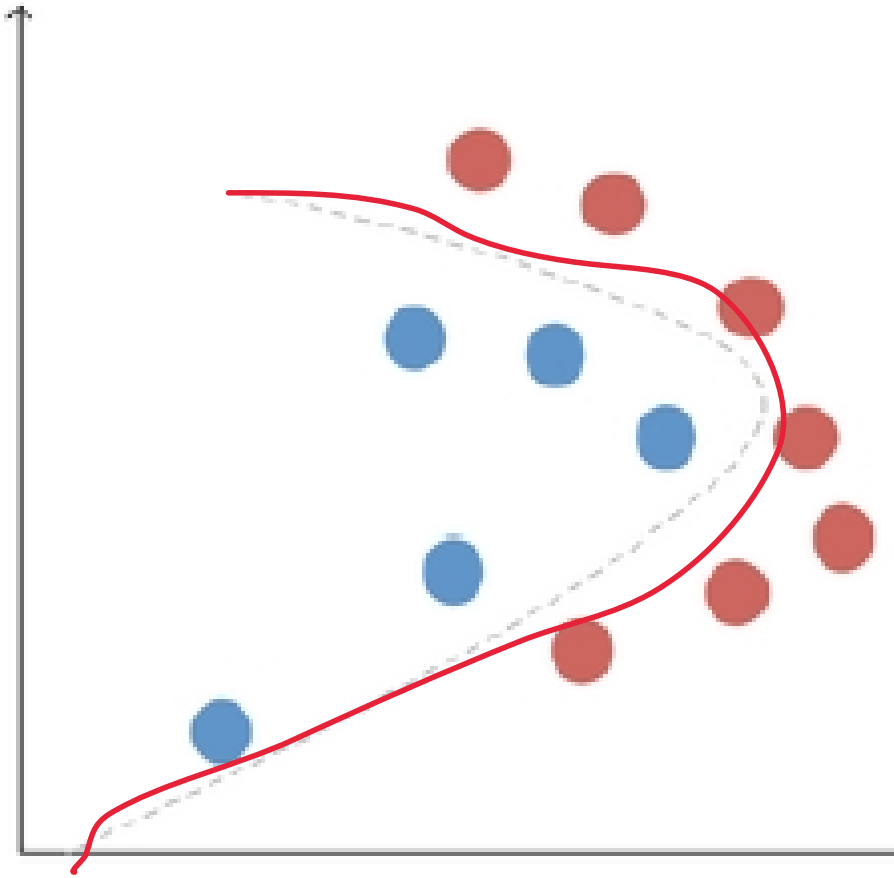


Overfitting

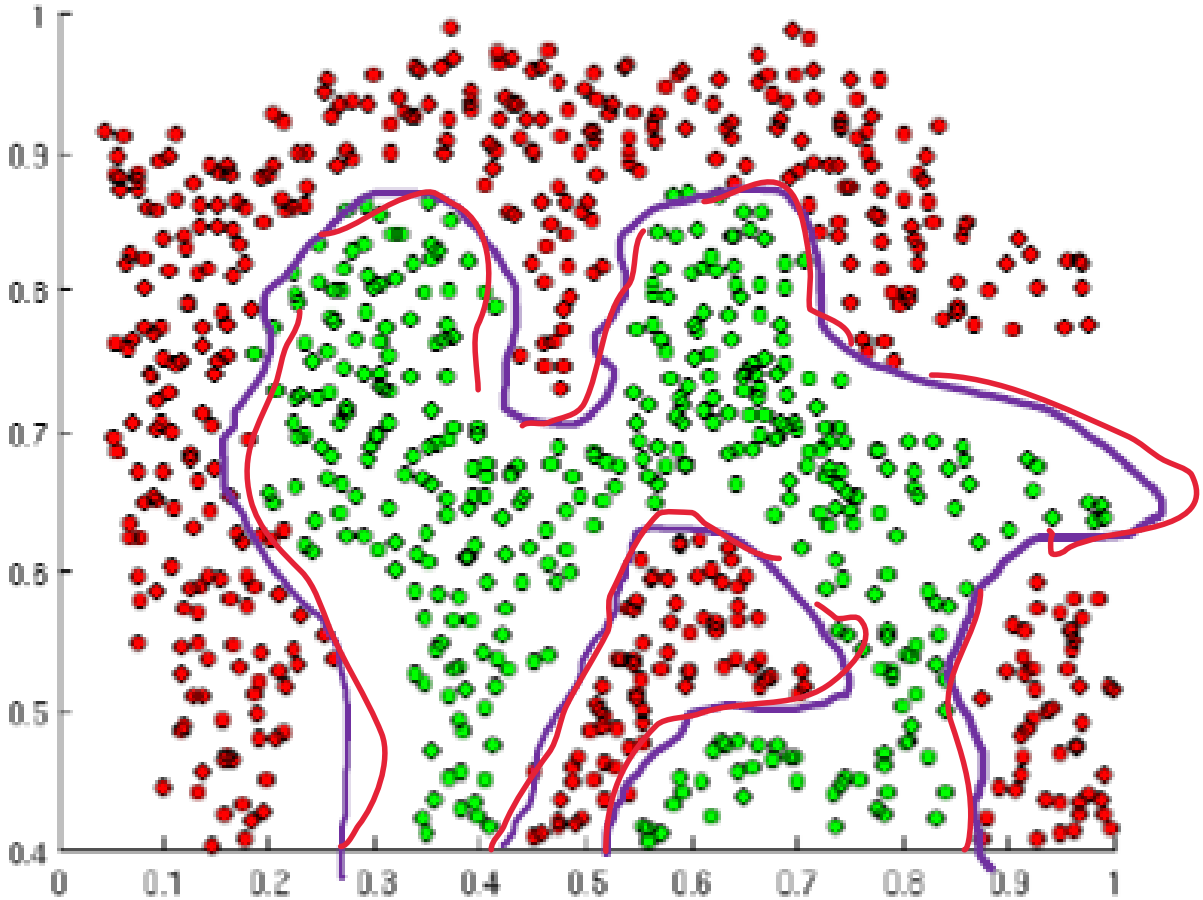


Redes Neurais - Hipótese não-linear

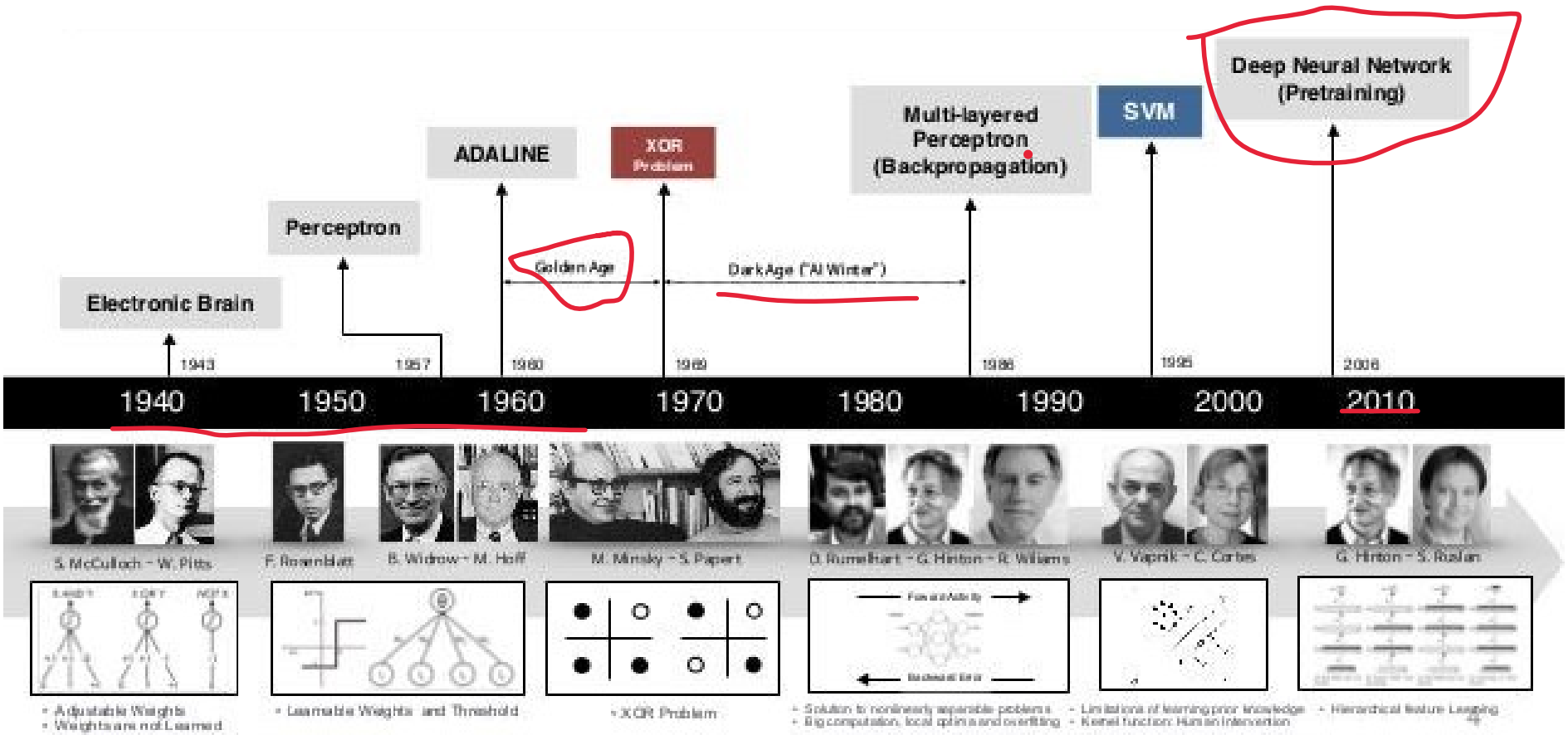




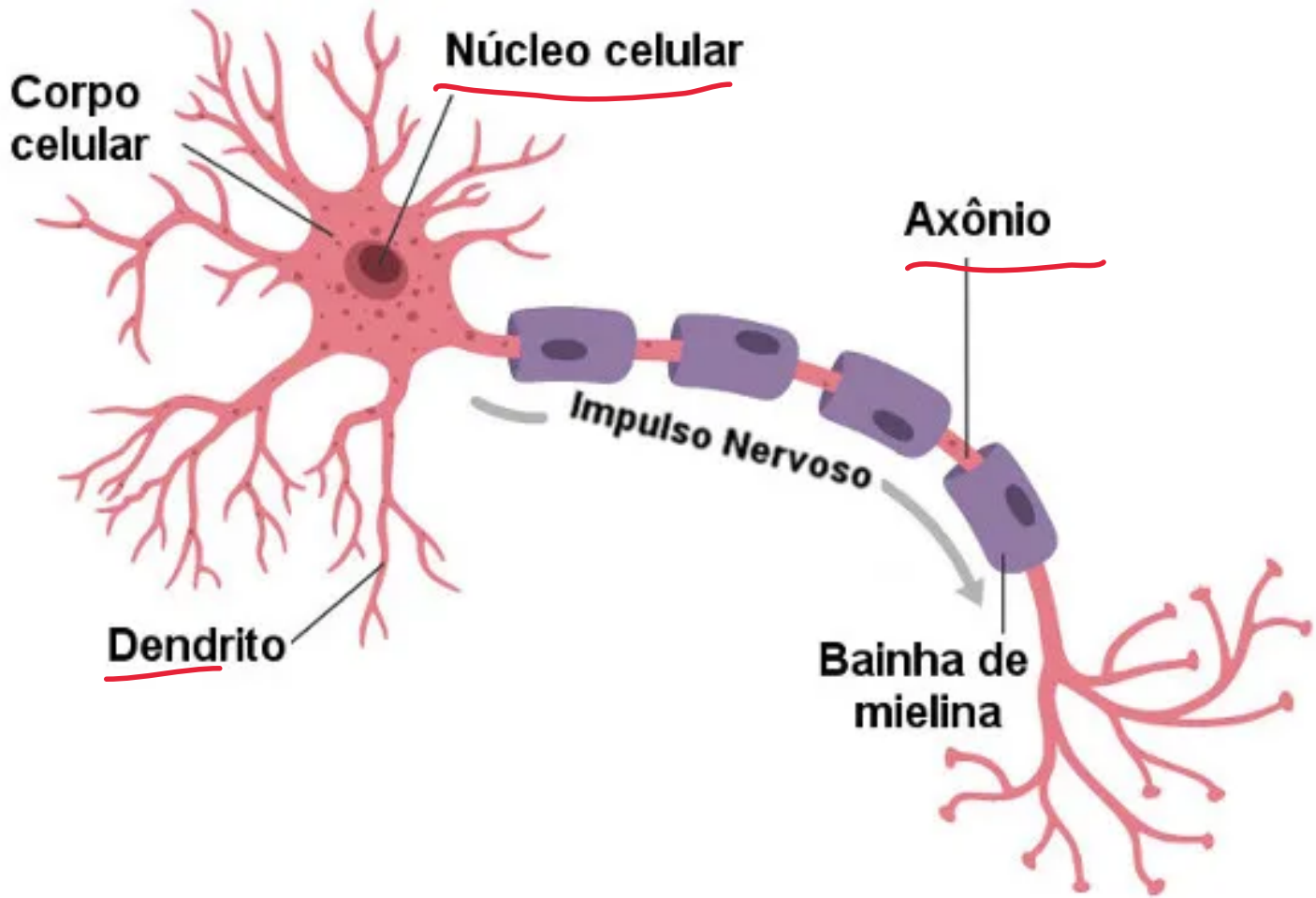
Redes Neurais - Hipótese não-linear



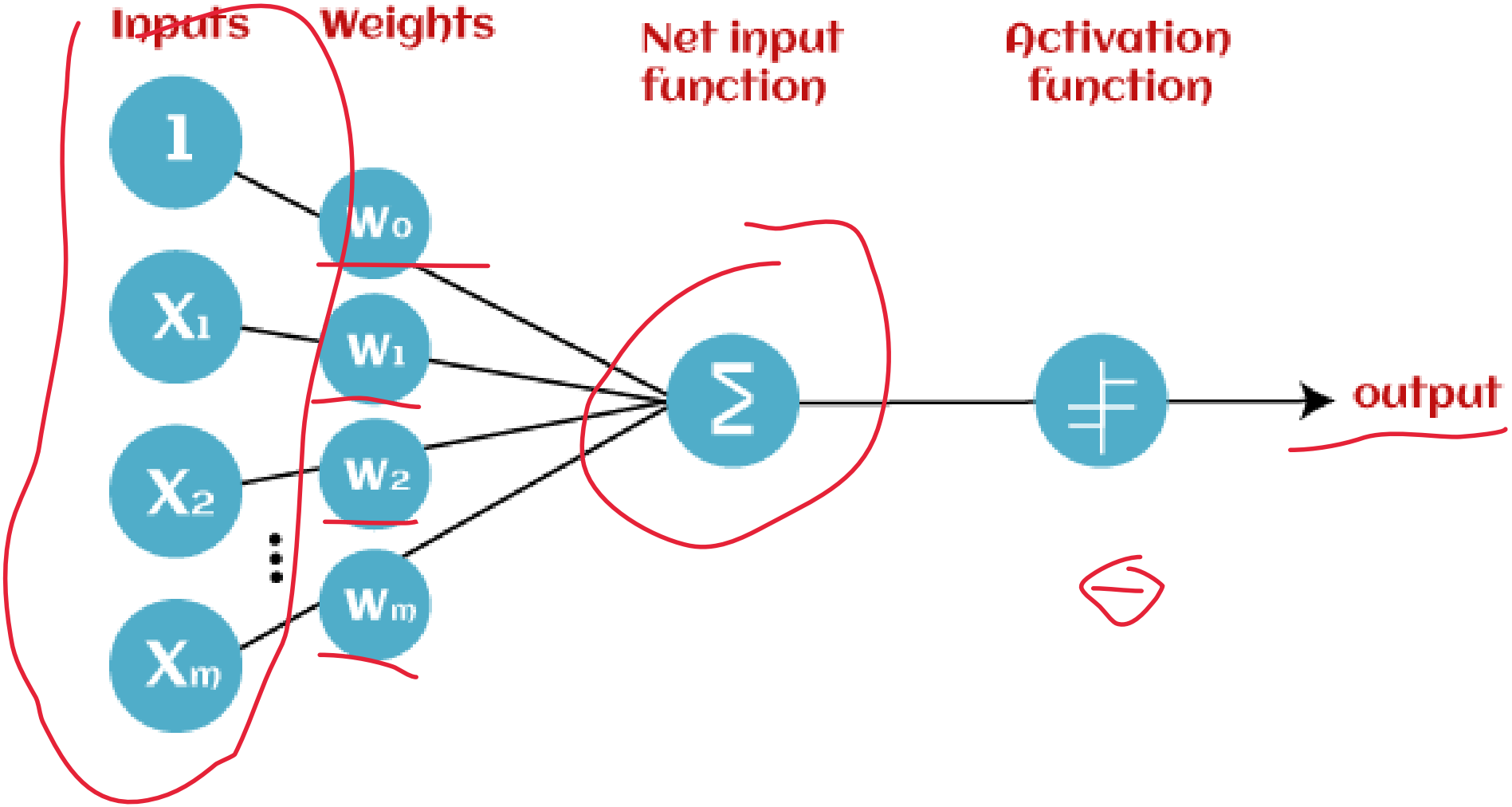
Redes Neurais - História



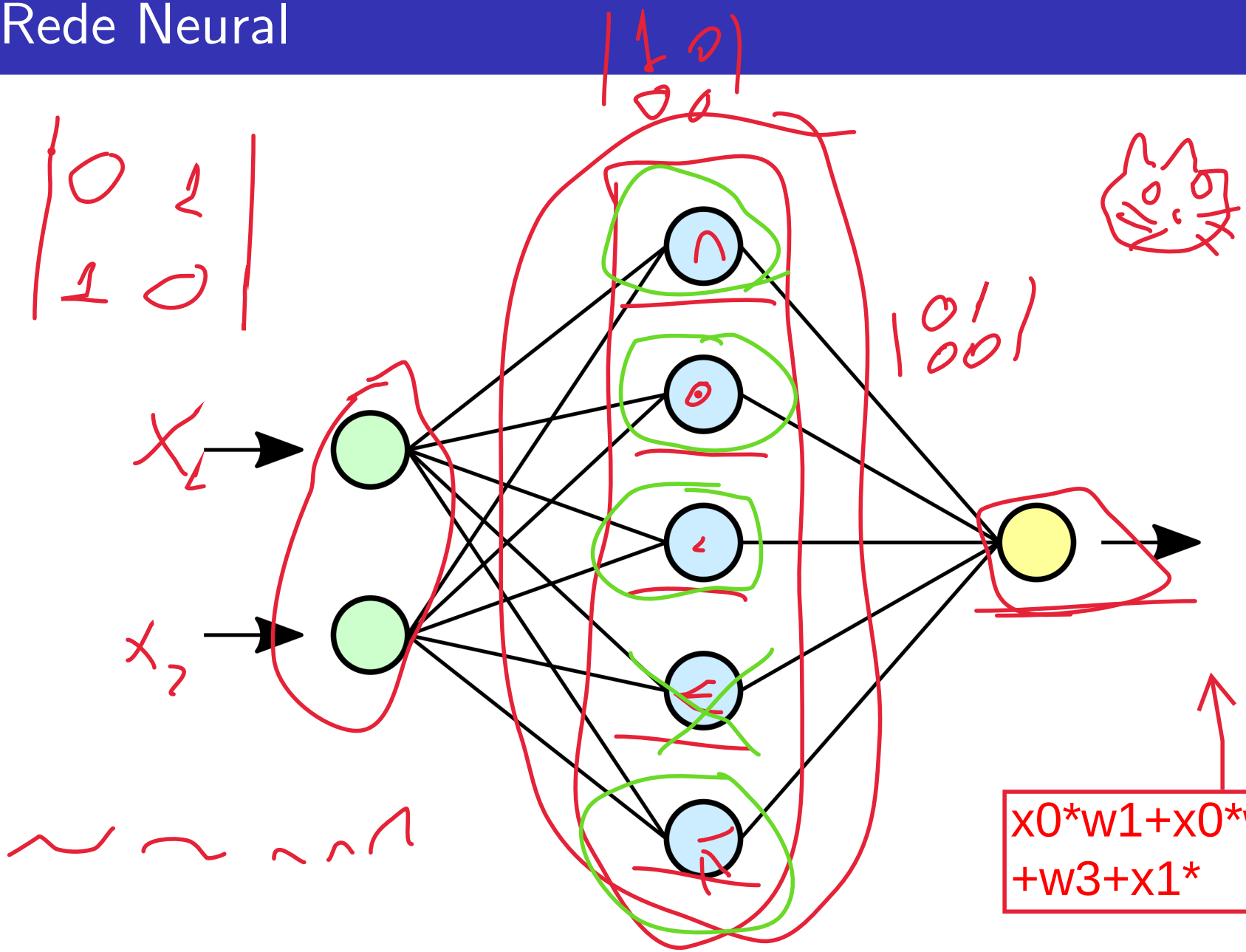
Redes Neurais - Neurônio



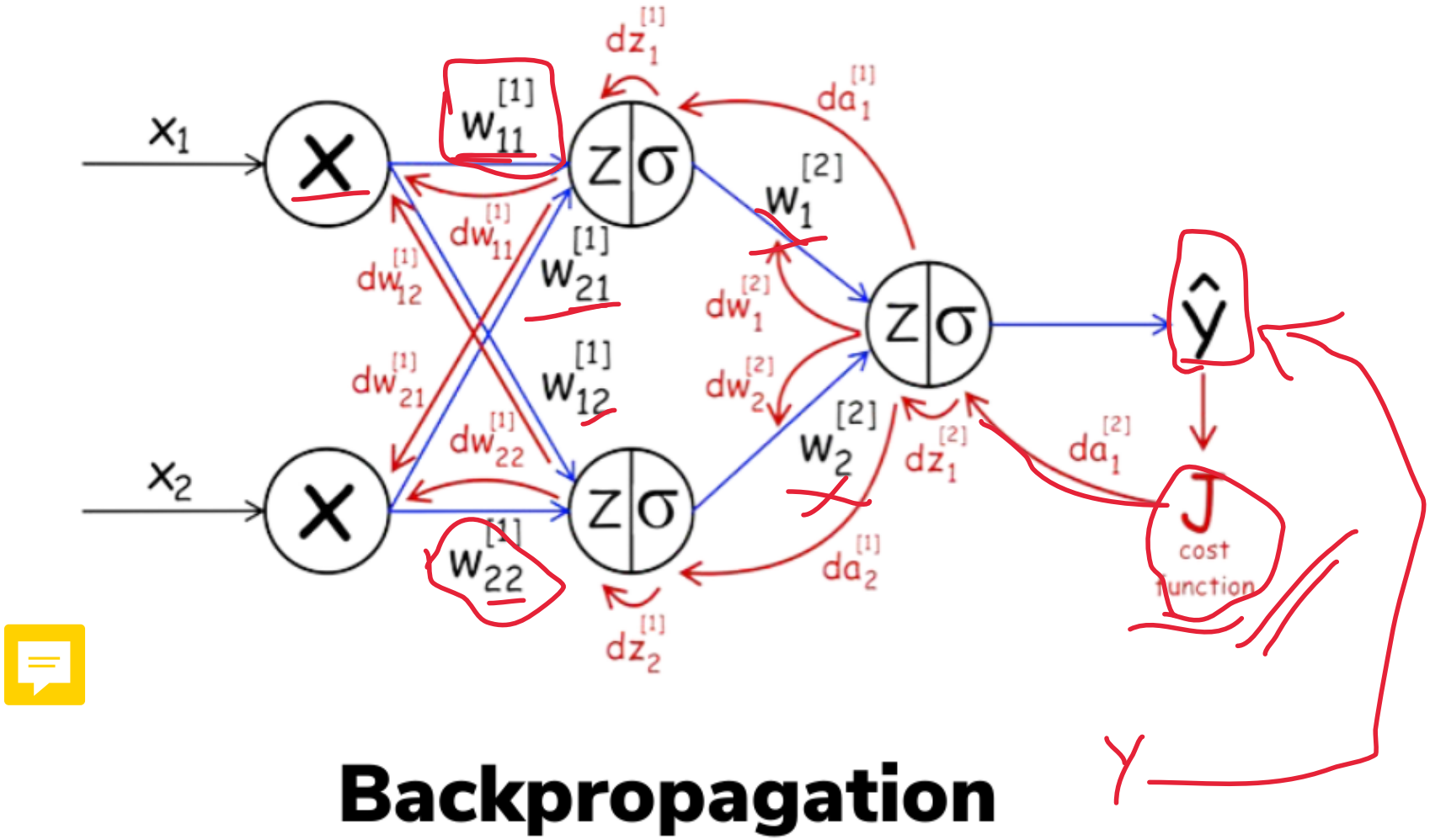
Redes Neurais - Perceptron



Rede Neural

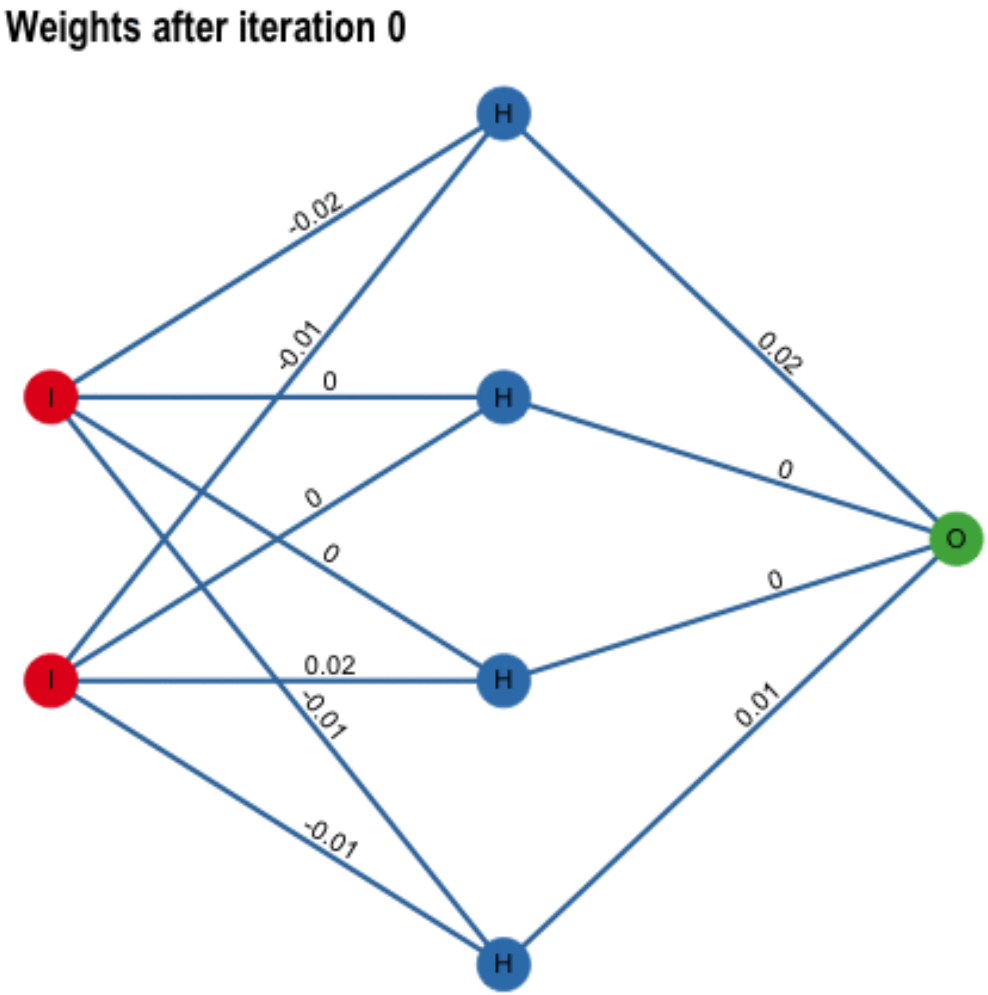


Redes Neurais - Backpropagation



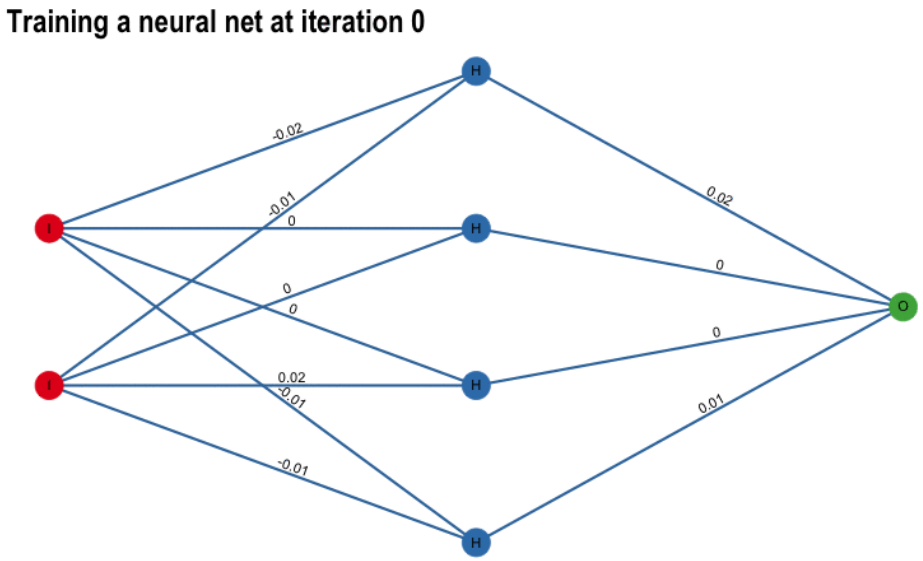
Backpropagation

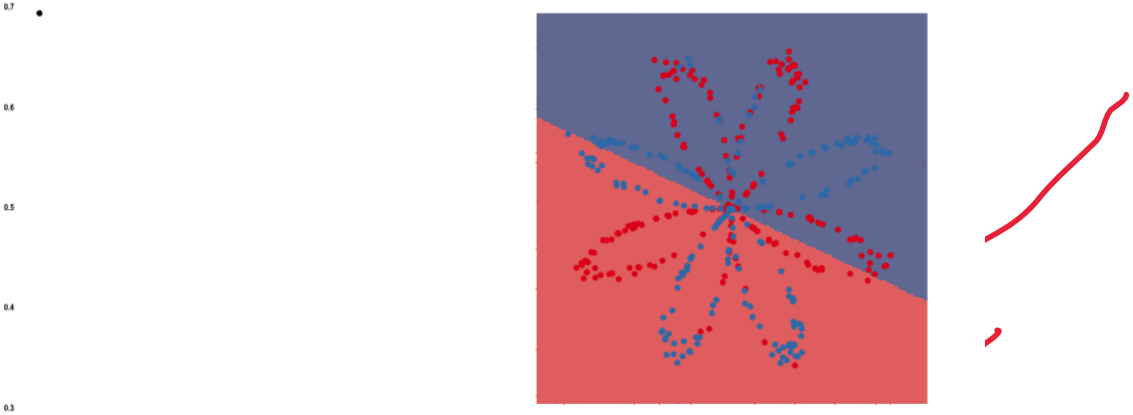
Redes Neurais - Training



Redes Neurais - Training

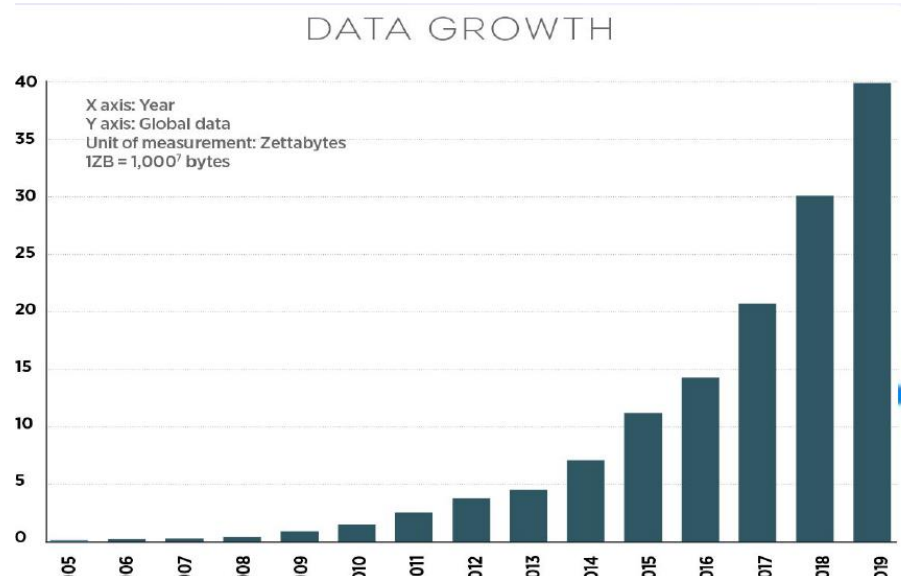
Redes Neurais - Training





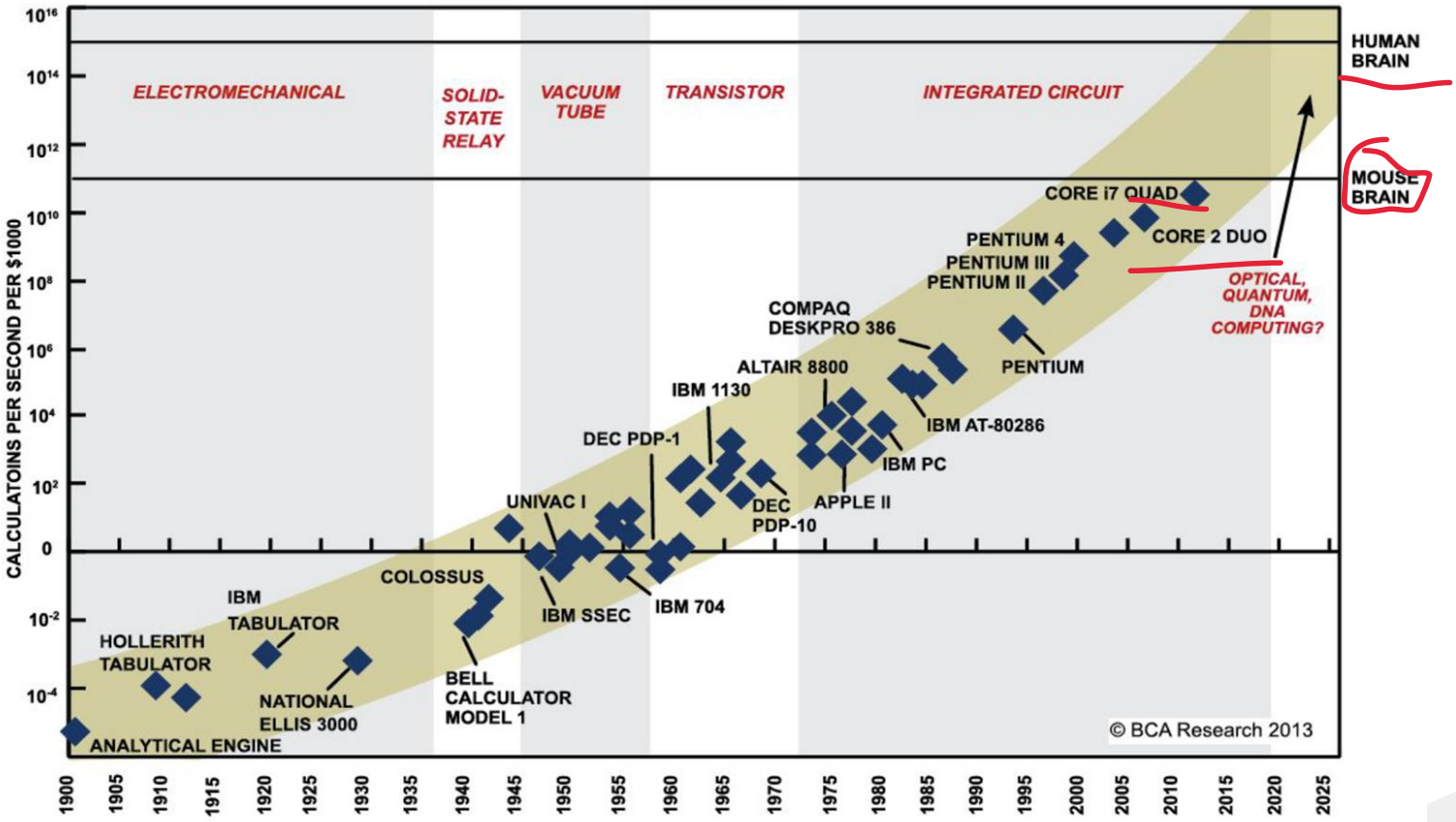
Por que Machine Learning está cada vez mais importante?

- Disponibilidade de muitos dados (Internet)
- Aumento do poder computacional (novas GPUs)
- Avanços nas técnicas e algoritmos



Facebook:
350 M imagens/dia
Youtube:
100 horas vídeo/minuto

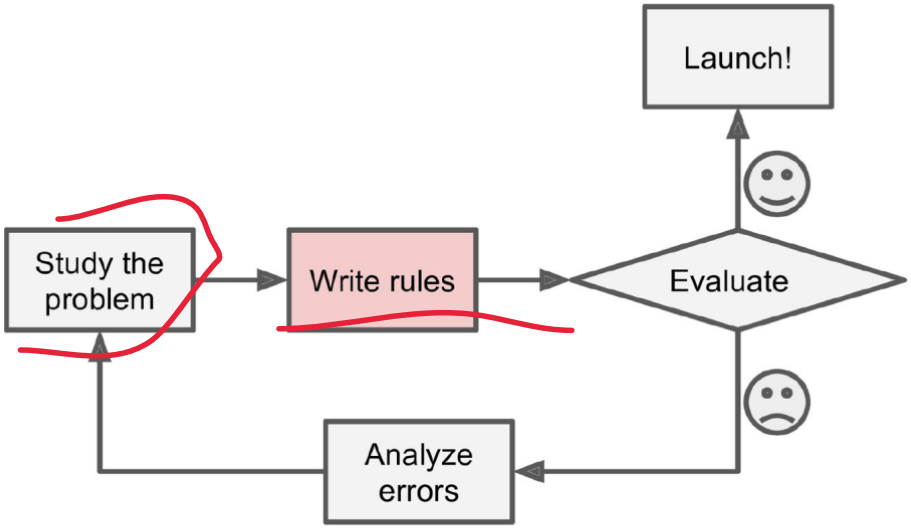
Por que Machine Learning está cada vez mais importante?



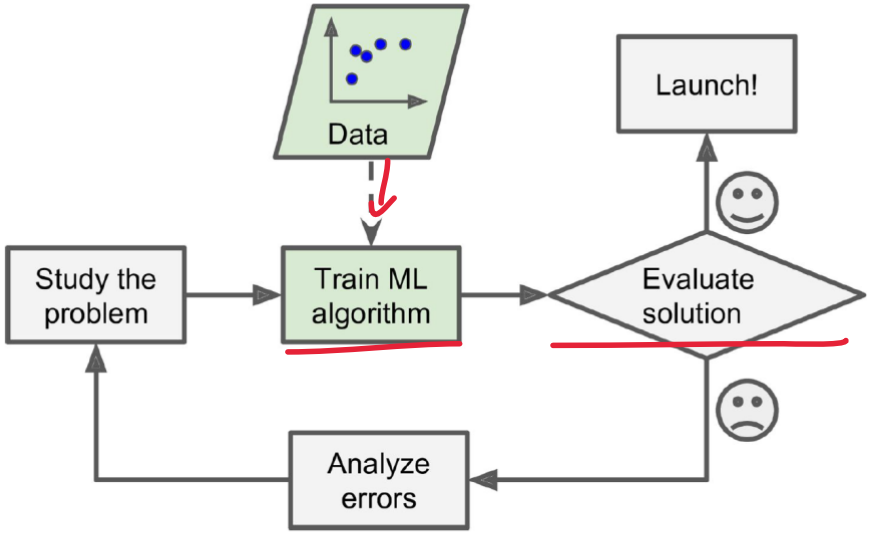
SOURCE: RAY KURZWEIL, "THE SINGULARITY IS NEAR: WHEN HUMANS TRANSCEND BIOLOGY", P.67, *THE VIKING PRESS*, 2006. DATAPOINTS BETWEEN 2000 AND 2012 REPRESENT BCA ESTIMATES.

O que é Machine Learning?

É o campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados. (Arthur Samuel, 1959)

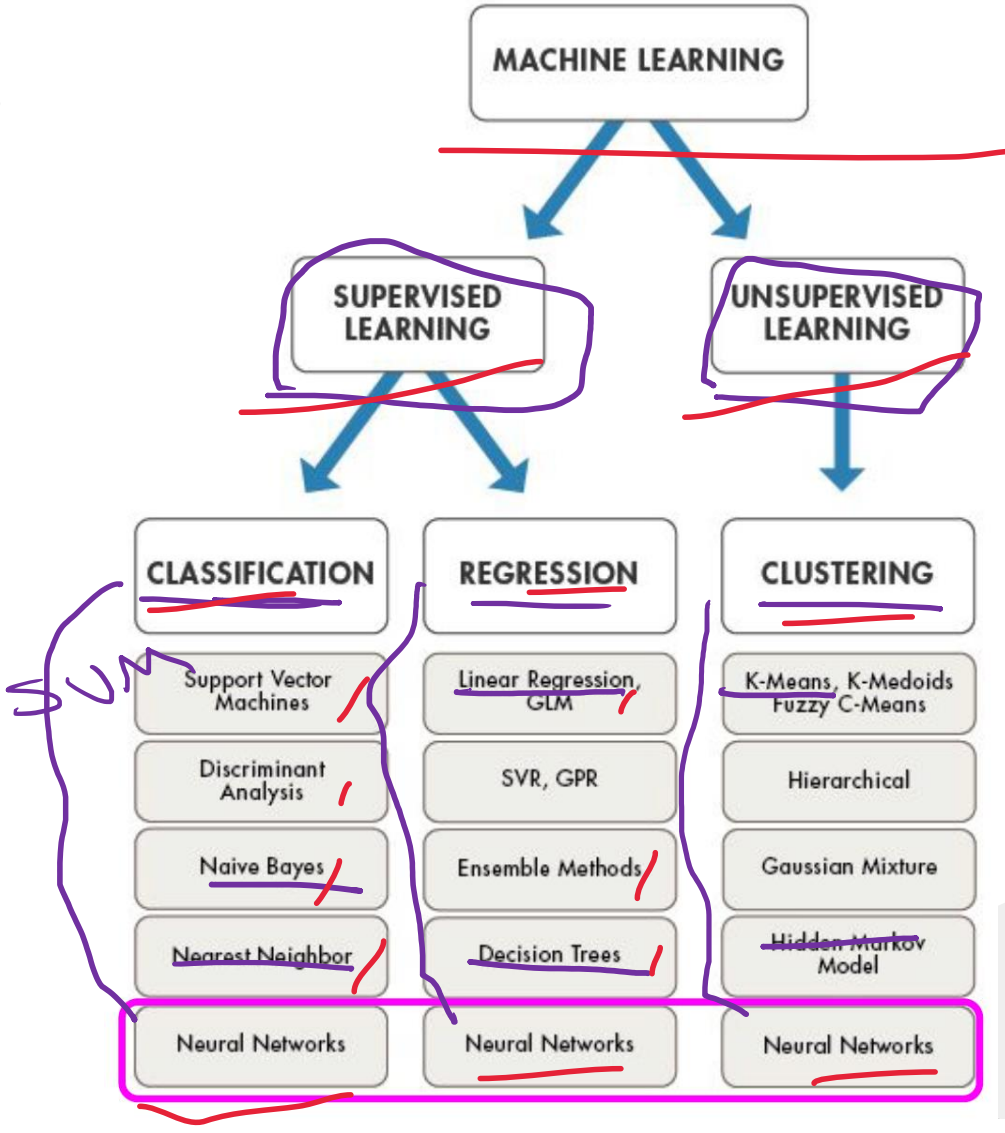


Sistema Especialista

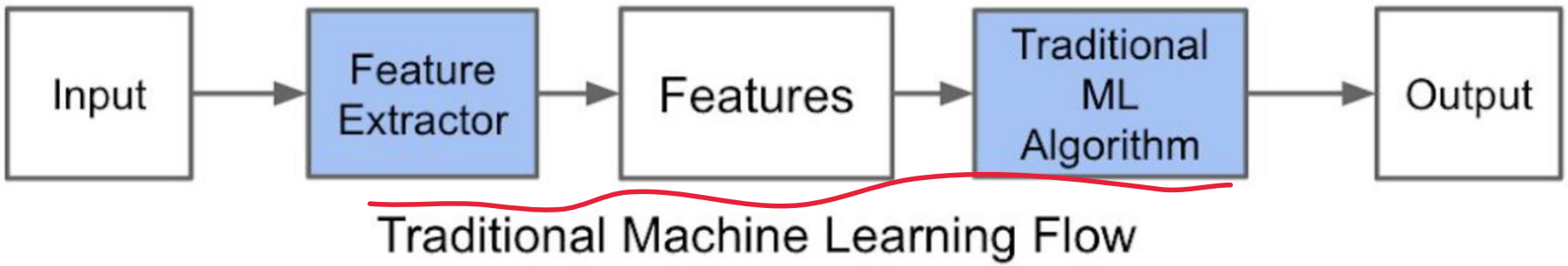


Aprendizado de Máquina

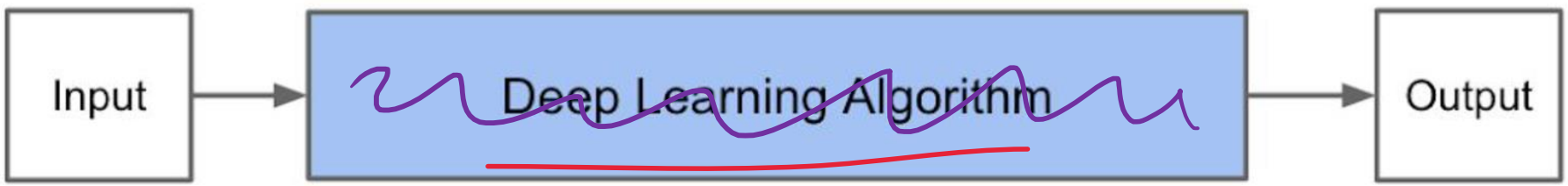
O que é Machine Learning?



Machine Learning vs Deep Learning

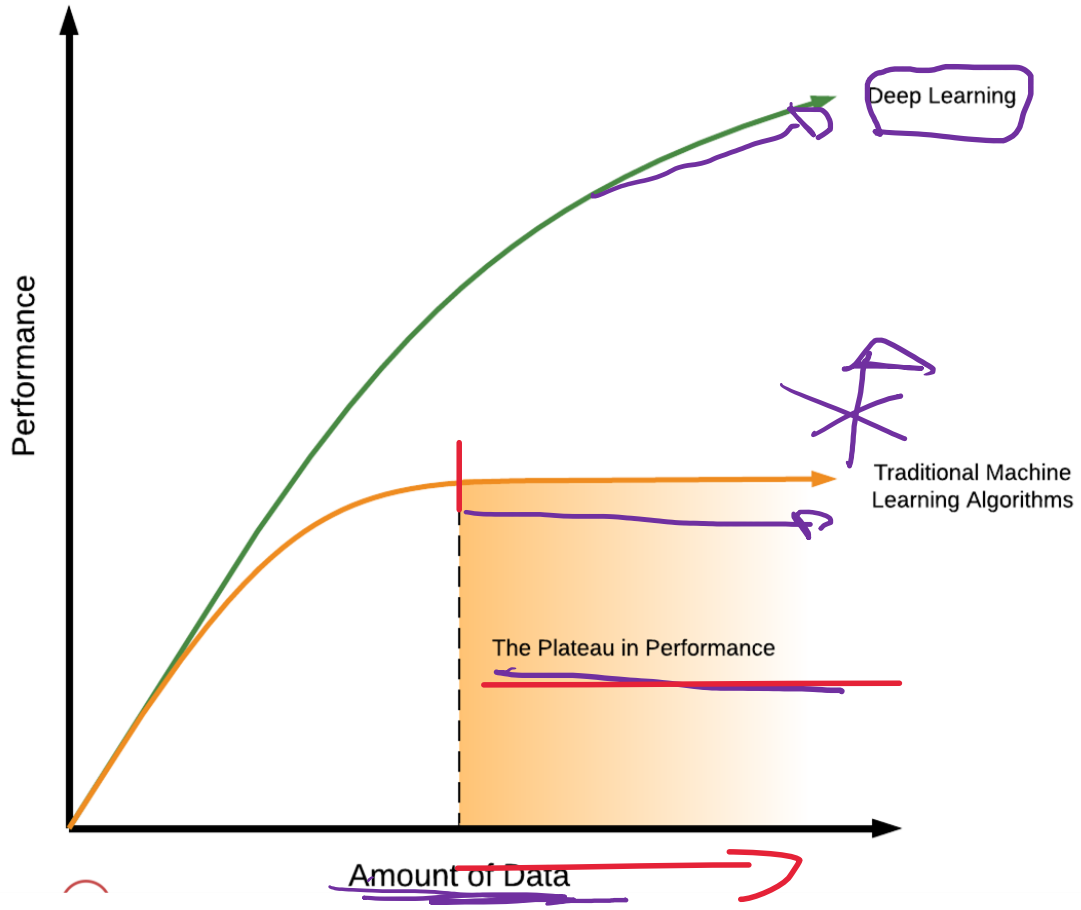


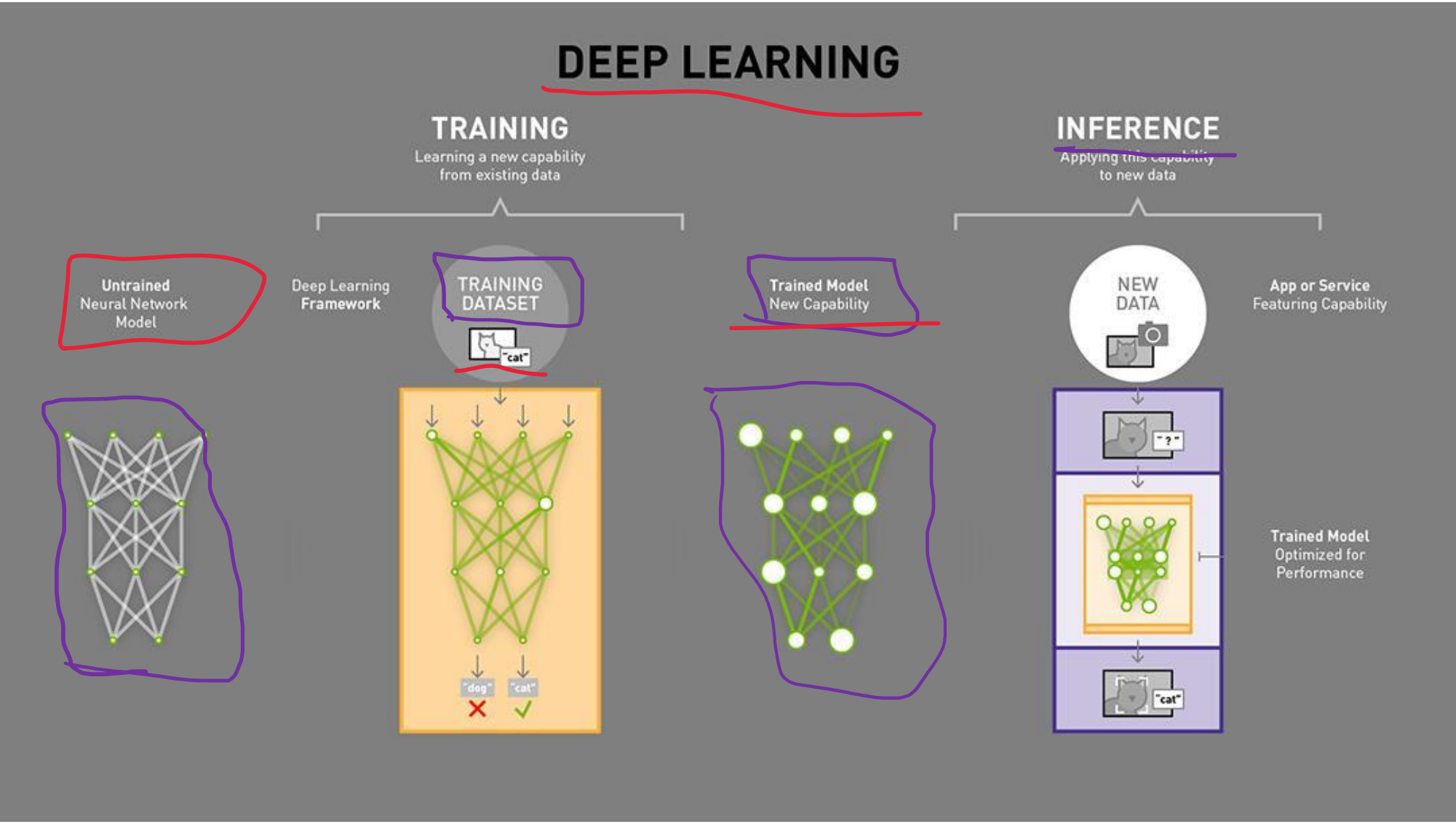
Traditional Machine Learning Flow



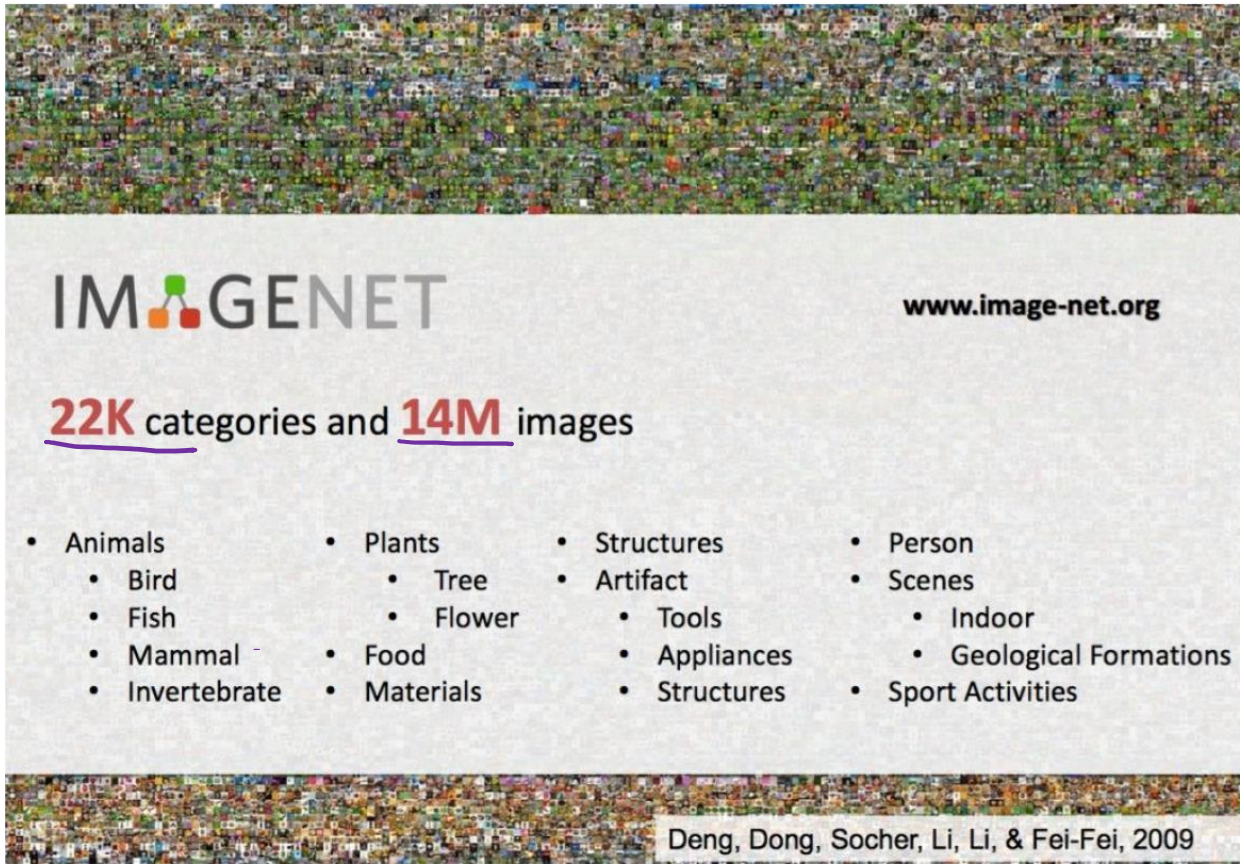
Deep Learning Flow

Machine Learning vs Deep Learning

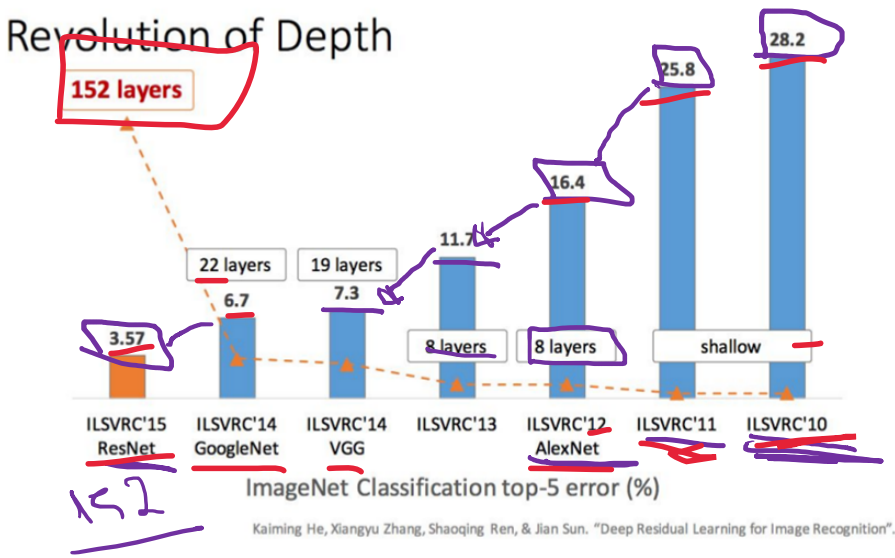



DEEP LEARNING
TRAINING
Learning a new capability from existing data
INFERENCE
Applying this capability to new data
Untrained Neural Network Model
Deep Learning Framework
TRAINING DATASET
"cat"
"dog"
"cat"
Trained Model New Capability
NEW DATA
"?"
"cat"
App or Service Featuring Capability
Trained Model Optimized for Performance

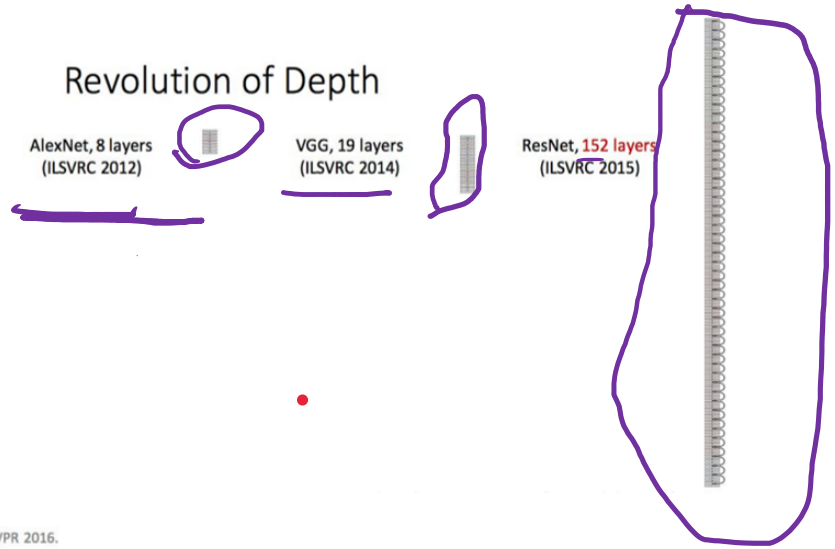
ImageNET (desde 2010)



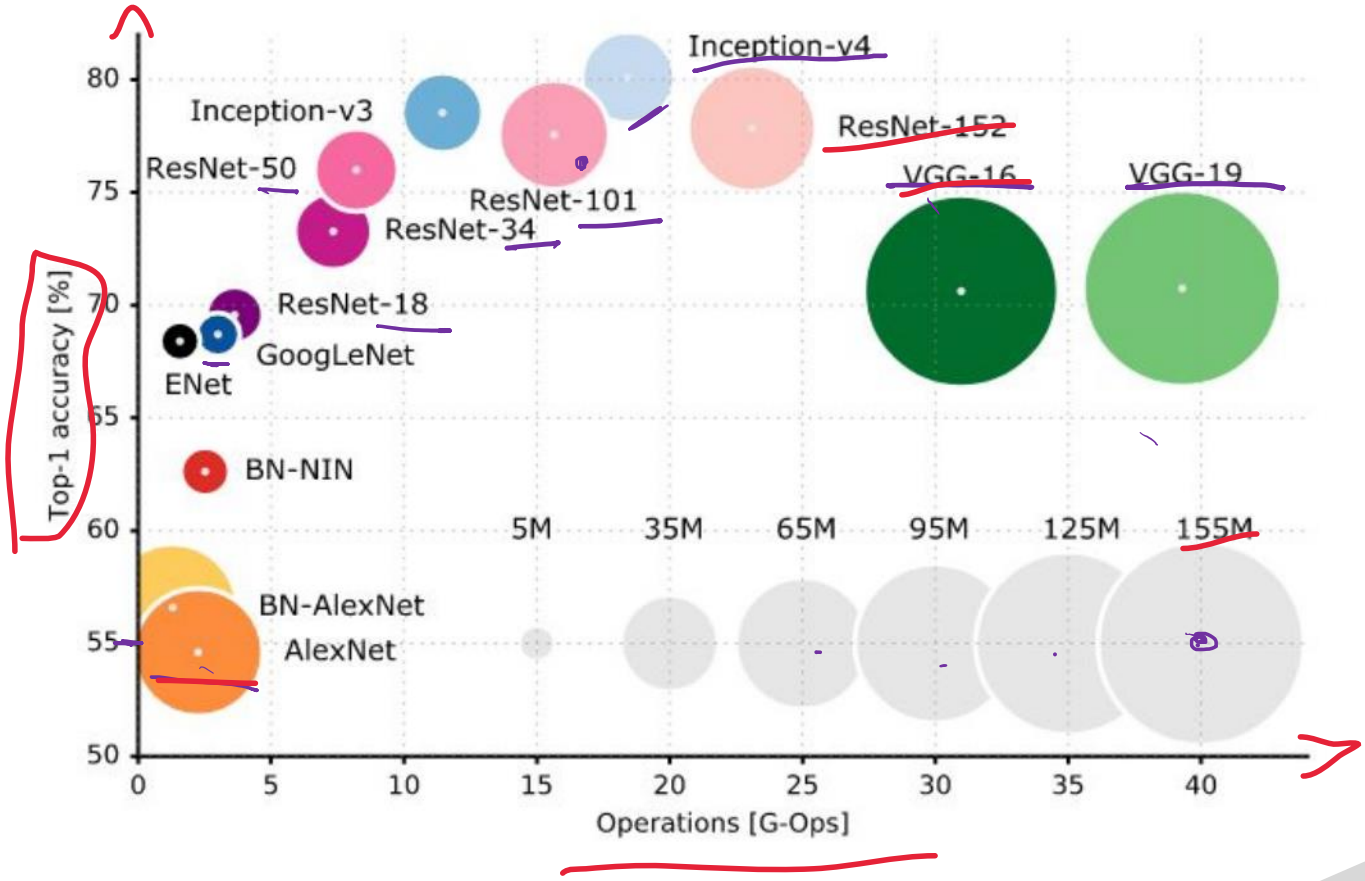
Redes cada vez mais profundas
ResNet, ILSVRC 2015, MSResearch



Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". CVPR 2016.



Desempenho das arquiteturas recentes



Principal dificuldade de Deep Learning?

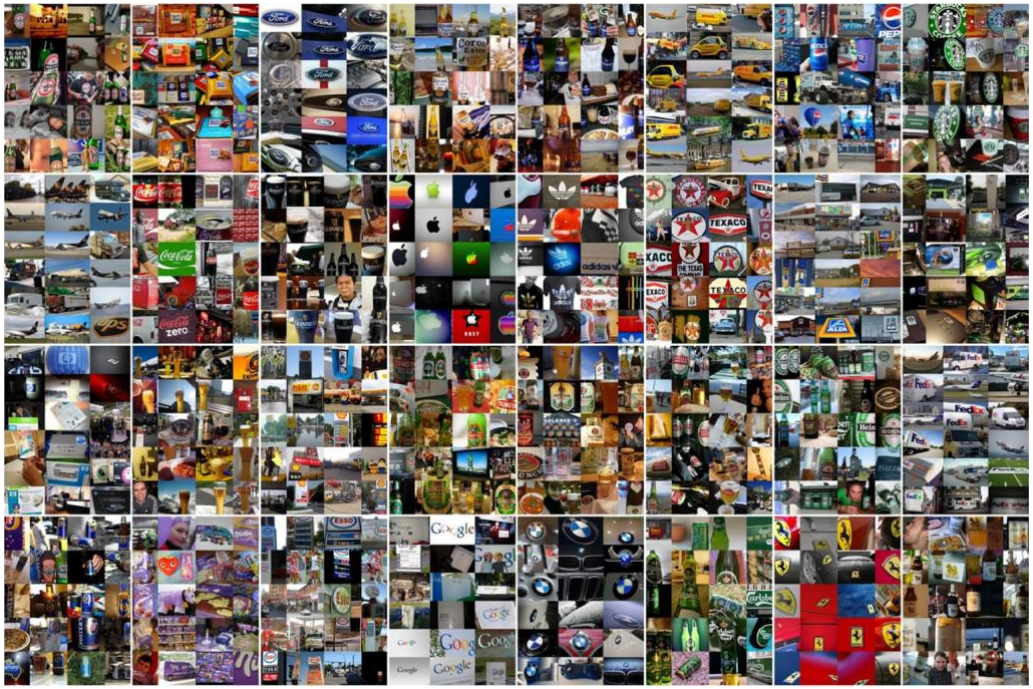
Alto custo do dado anotado (supervisionado)

https://www.remotasks.com/

It's estimated that 1.2 trillion photos will be taken in 2017. Even if each photo only took someone 1 second to organize, tag and annotate, it would still take over 38,000 years to classify them all!

https://www.kaggle.com/c/imagenet-object-detection-challenge

Principal dificuldade de Deep Learning?

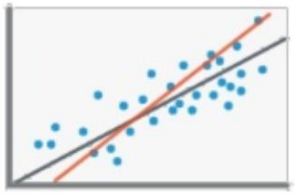


A taxa de acerto de um classificador supervisionado depende muito dos dados usados em seu treinamento.

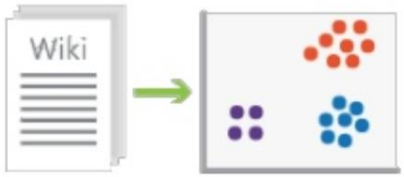
Escassez de boas bases de dados anotados em imagens médicas é um dos principais gargalos da área.

Tipos de Problemas

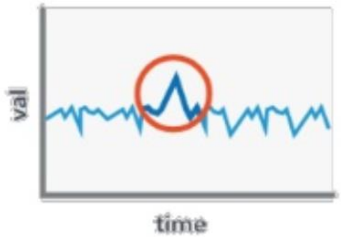
Classification
(supervised – predictive)



Regression
(supervised – predictive)



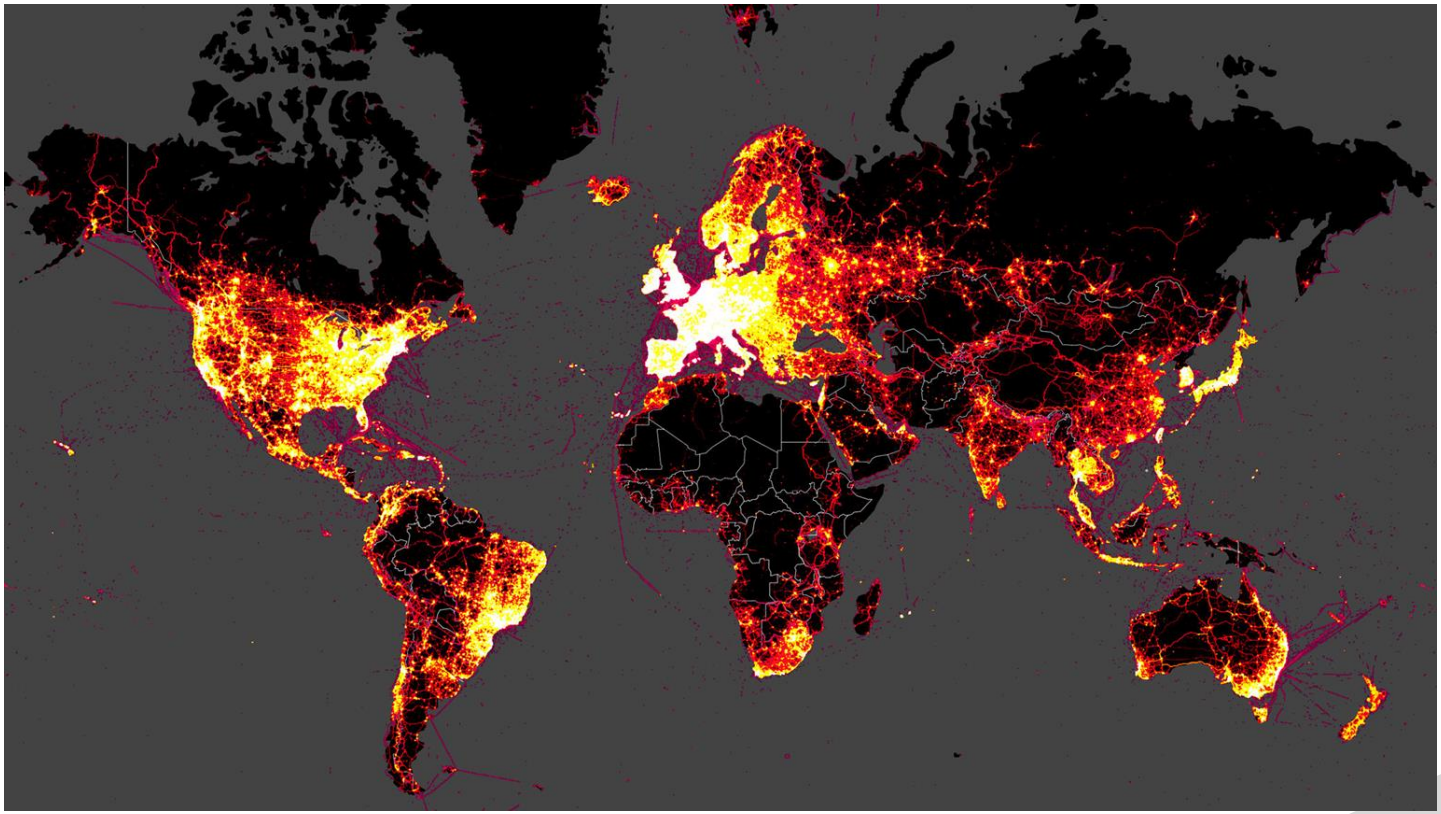
Clustering
(unsupervised – descriptive)



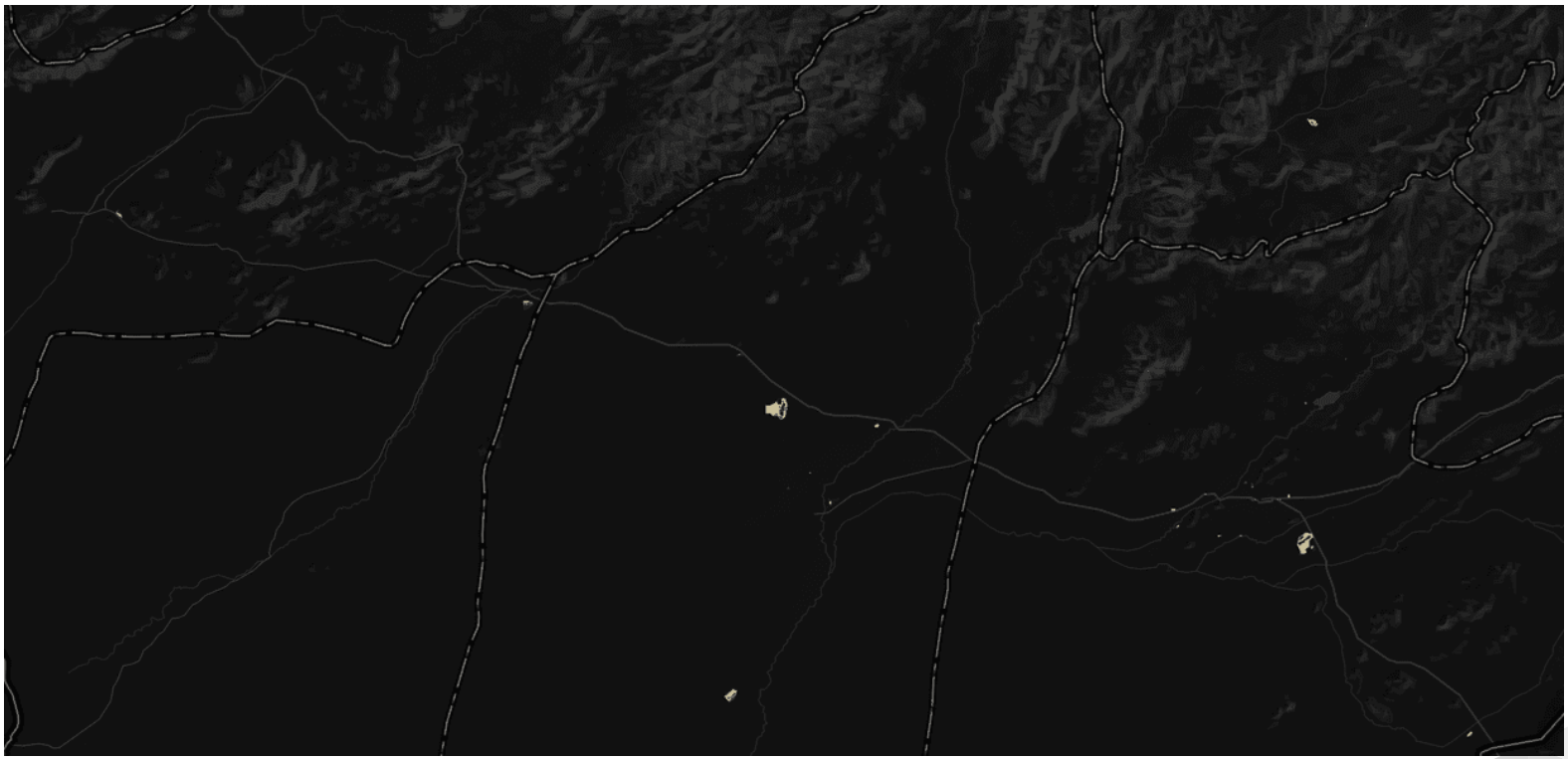
Anomaly Detection
(unsupervised – descriptive)

Como as máquinas aprendem?

Importância dos Dados



Importância dos Dados



Área remota do Afeganistão

Importância dos Dados

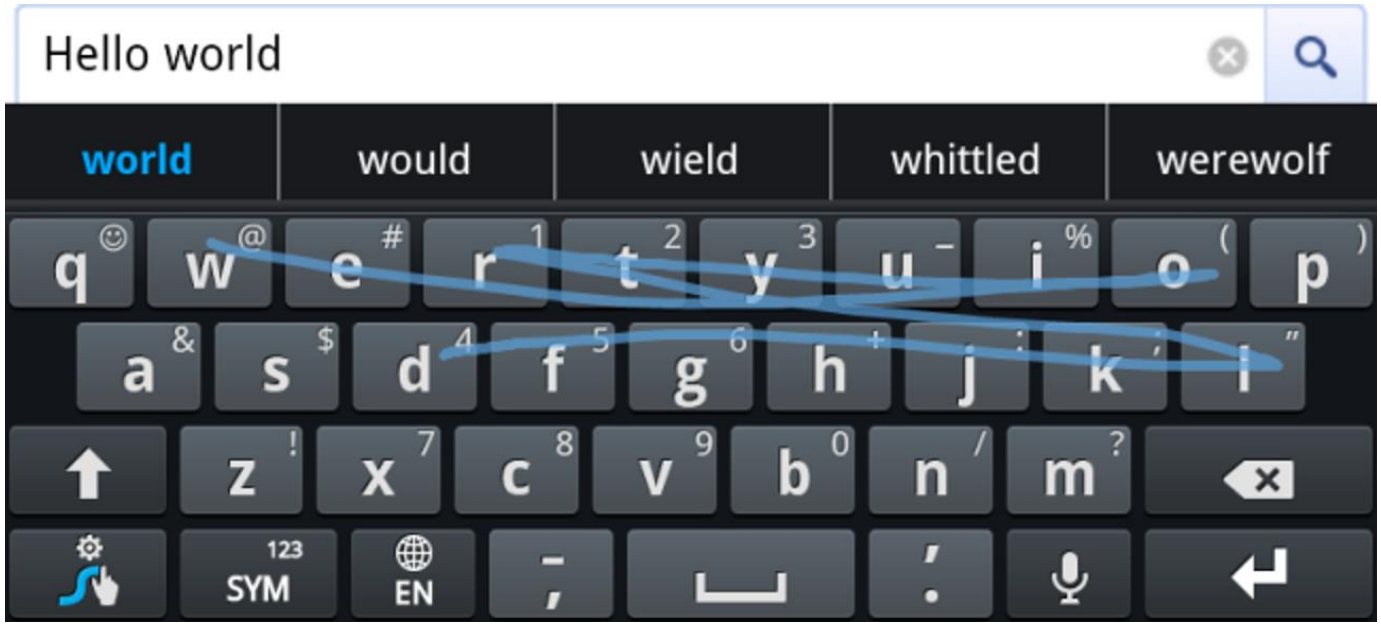
Heatmap de atividades do Strava em base militar secreta

AI no dia-a-dia

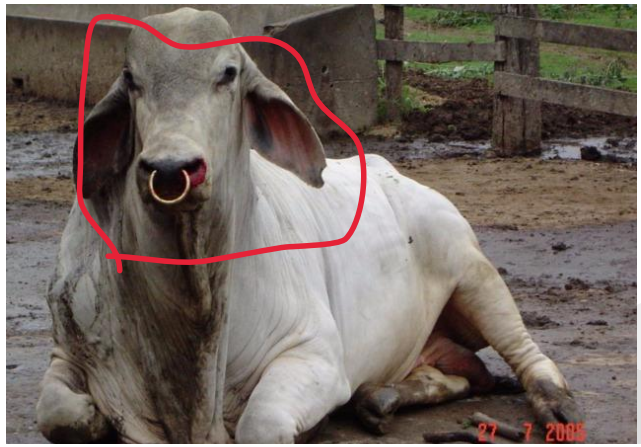
Motorola C115

AI no dia-a-dia

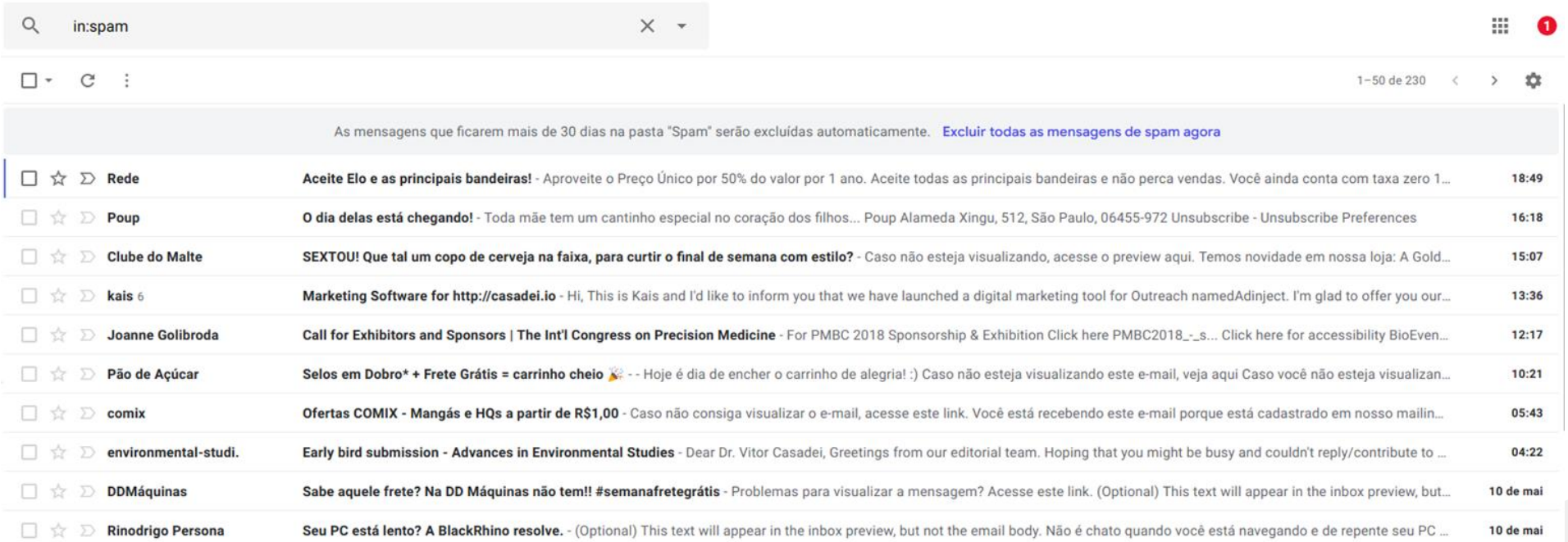
Motorola C115



AI no dia-a-dia

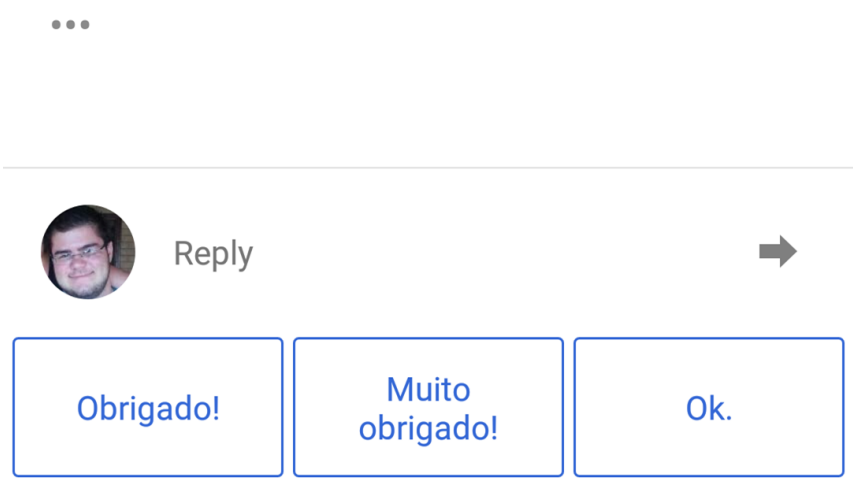


AI no dia-a-dia

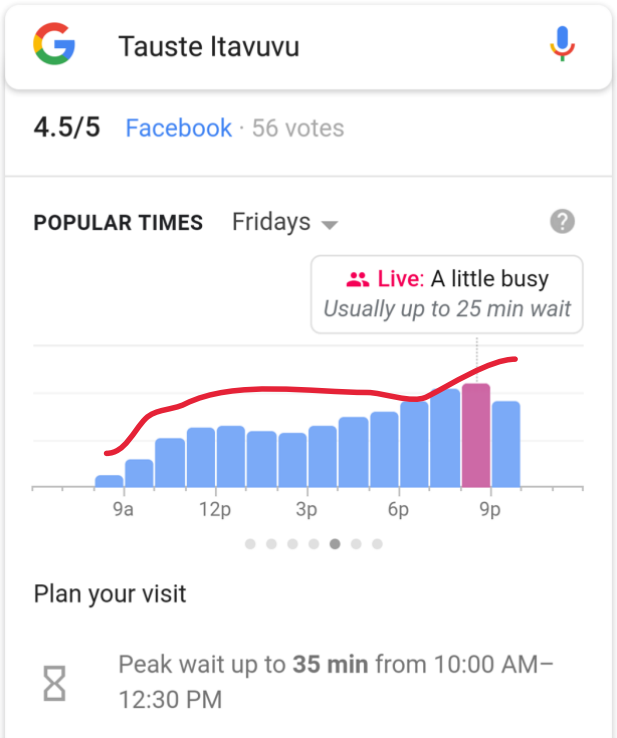


AI no dia-a-dia

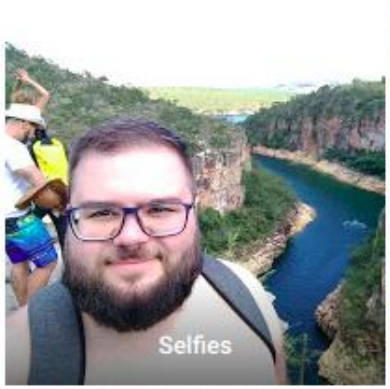
AI no dia-a-dia



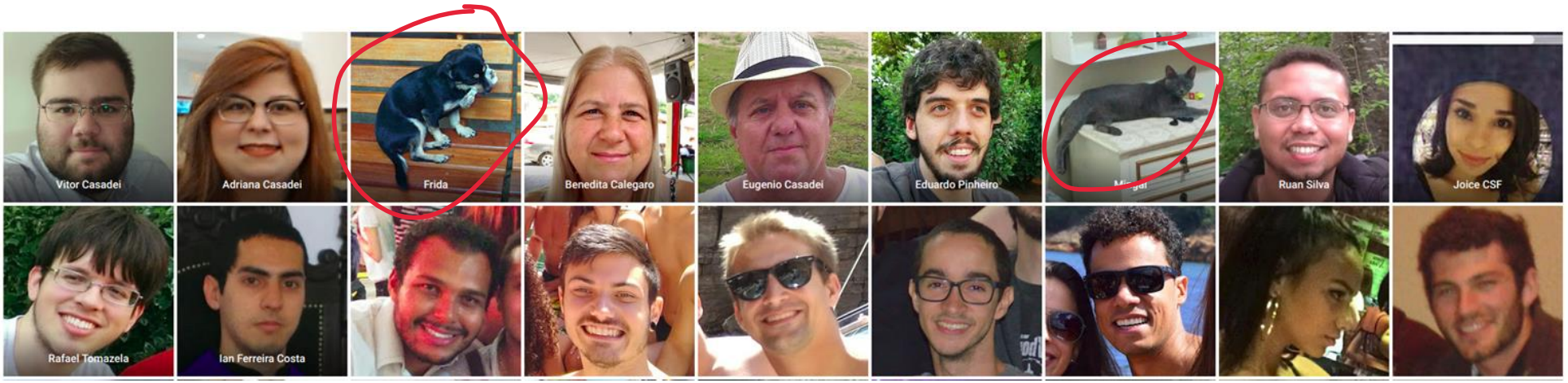
AI no dia-a-dia



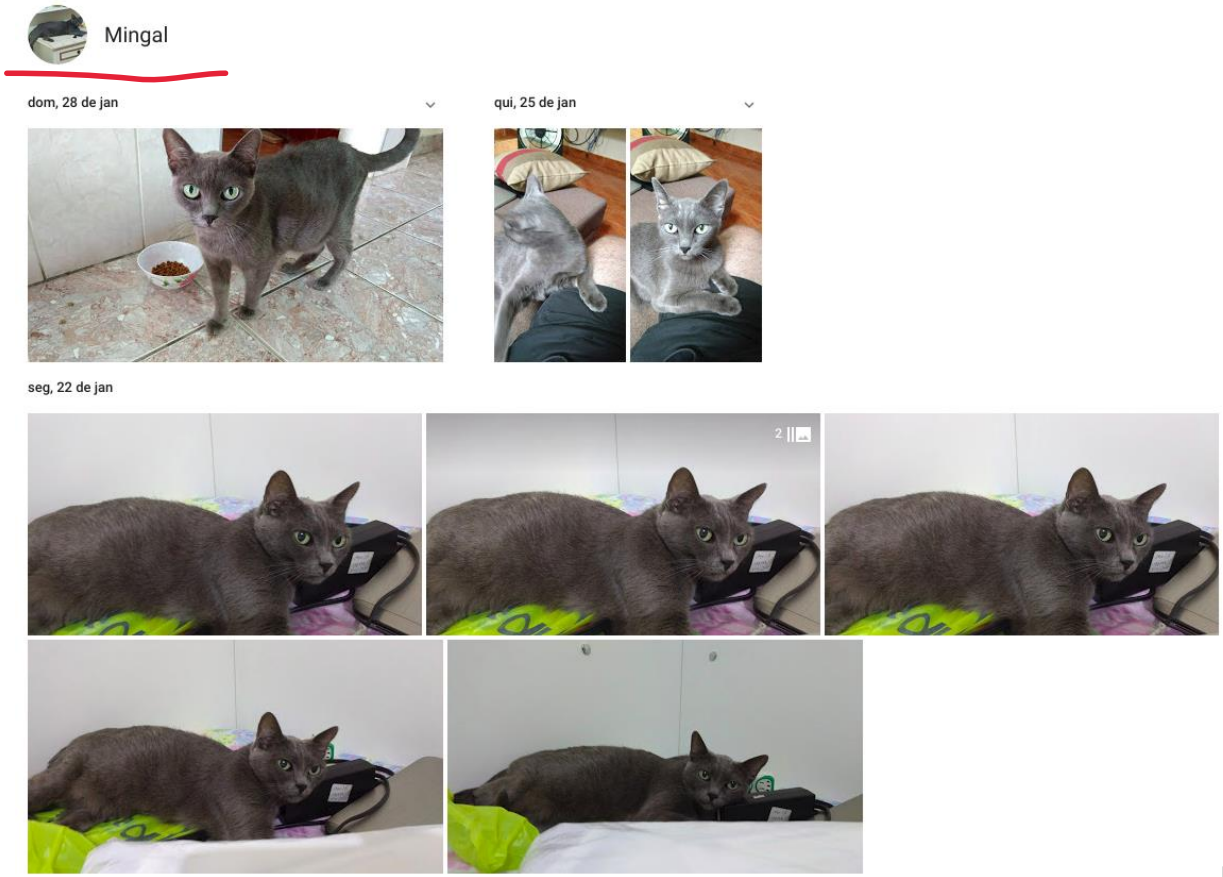
AI no dia-a-dia



AI no dia-a-dia



AI no dia-a-dia



AI no dia-a-dia

Frida

sáb, 30 de dez de 2017

dom, 19 de nov de 2017

qui, 6 de jul de 2017

AI no dia-a-dia