

Modelos Preditivos Conexionistas

Vitor Casadei

Especialização em Dados - 2022.02

Outline

- 1 Introdução
- 2 Aprendizado
- 3 Regressão Linear
- 4 Regressão Logística
- 5 Overfitting e Underfitting
- 6 Redes Neurais
- 7 Machine Learning e Deep Learning

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)

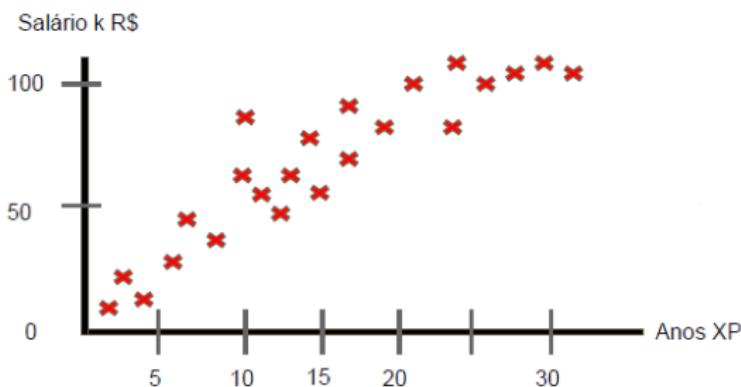
Well posed Learning Problem: A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

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

Predição de Salário (Anual)



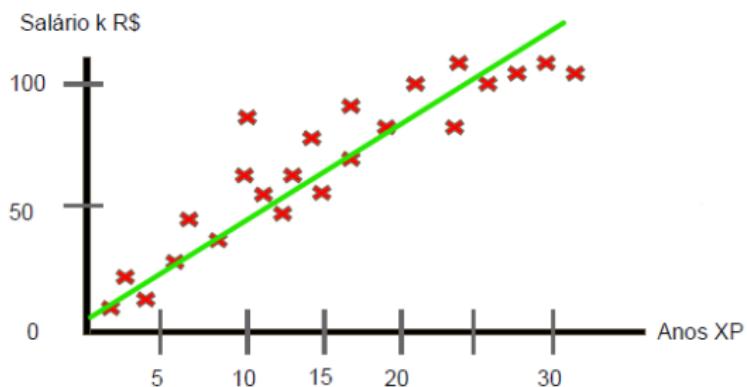
Aprendizado com base nas
"respostas certas"

Régressão: Predição de
valores contínuos

Figure: Exemplo de Aprendizado Supervisionado

Aprendizado Supervisionado

Predição de Salário (Anual)



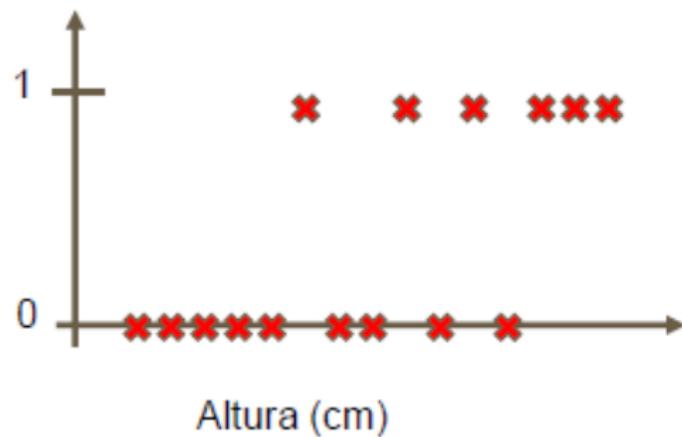
Aprendizado com base nas
"respostas certas"

Régressão: Predição de
valores contínuos

Figure: Exemplo de Aprendizado Supervisionado

Aprendizado Supervisionado

Pessoa Adulto

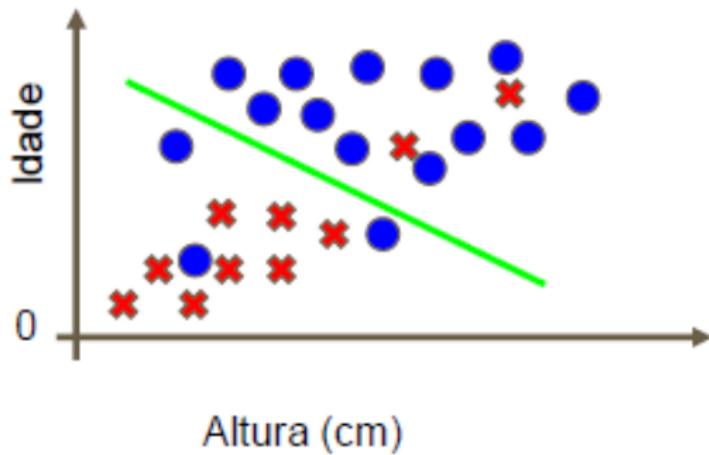


Classificação: Predição de valores discretos

Figure: Exemplo de Aprendizado Supervisionado

Aprendizado Supervisionado

Pessoa Adulto



Classificação: Predição de valores discretos

Figure: Exemplo de Aprendizado Supervisionado

Aprendizado Não Supervisionado

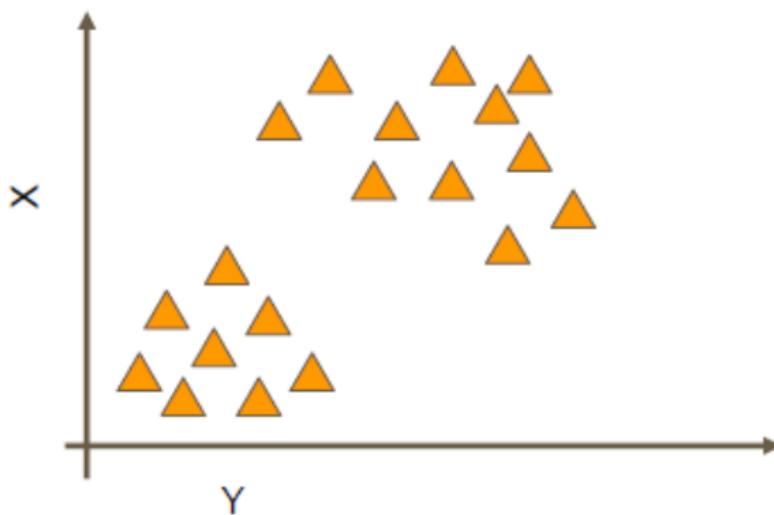


Figure: Exemplo de Aprendizado Não Supervisionado

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!"



Manoel

Retired man, 70 years.

"Technology is not just for the younger"



Lucia

A worker' retired, 60 years.

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

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.

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 keeps 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 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.

Regressão Linear - Uma variável

Predição de Salário (Anual)

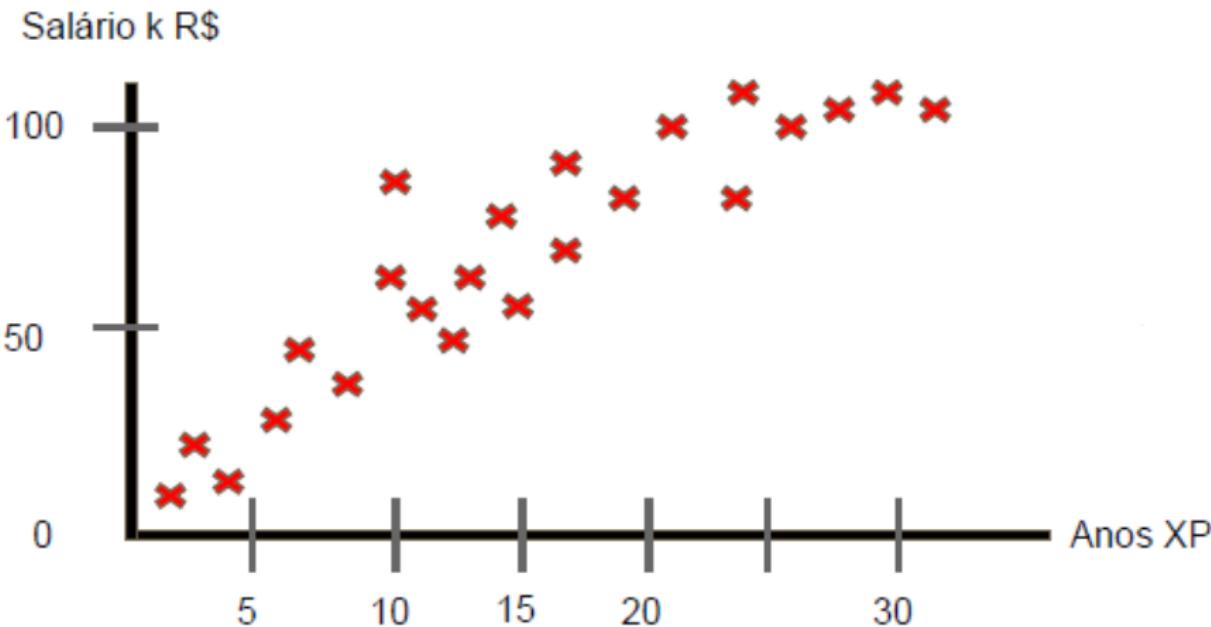


Figure: Exemplo de Aprendizado Não Supervisionado

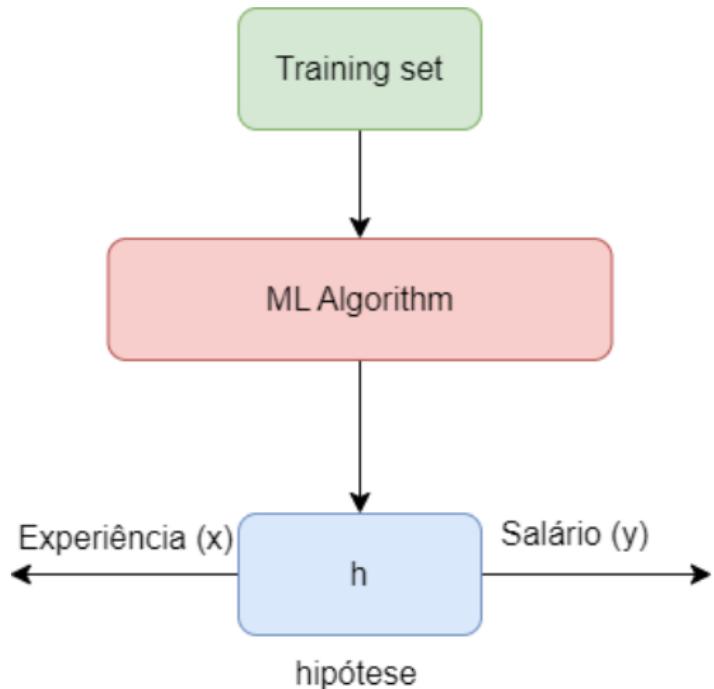
Regressão Linear - Uma variável

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
...	...

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

Regressão Linear - Função de Custo

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

$$Hipótese : h(x) = 40_0 + 0_1 * X$$

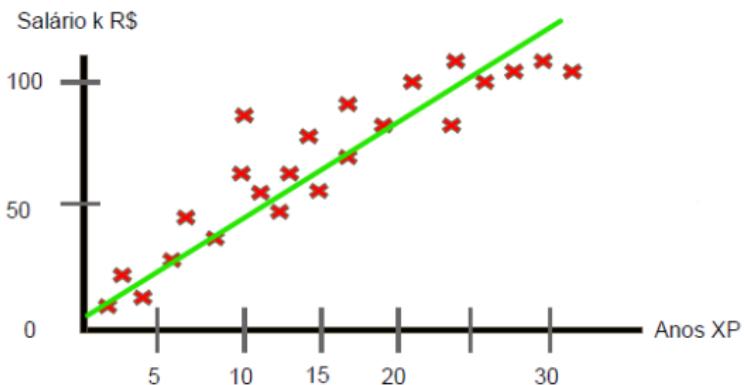
$$\theta_0 = 40 | \theta_1 = 0$$

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

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

Regressão Linear - Função de Custo

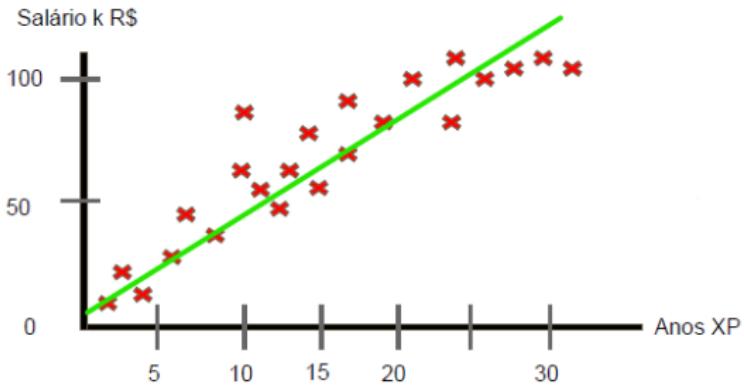
Predição de Salário (Anual)



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

Regressão Linear - Função de Custo

Predição de Salário (Anual)



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

$$\text{Hipótese : } h(x) = \theta_0 + \theta_1 * X$$

$$\text{Minimize : } \theta_0 \theta_1$$

Mean Square Error Function

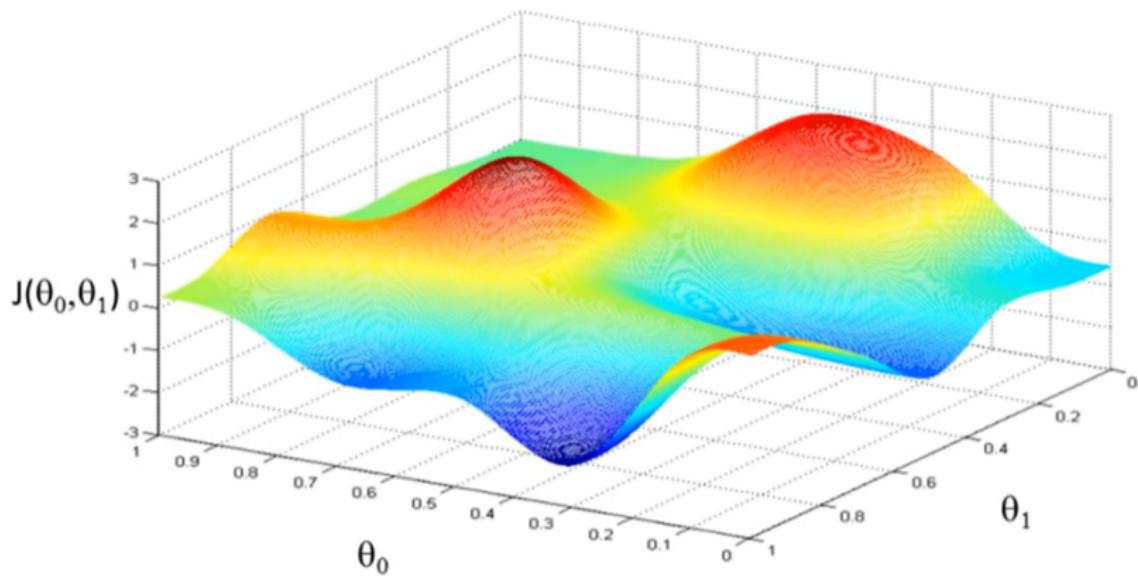
Objetivo

- Minimizar θ_0 e θ_1

Como Fazer:

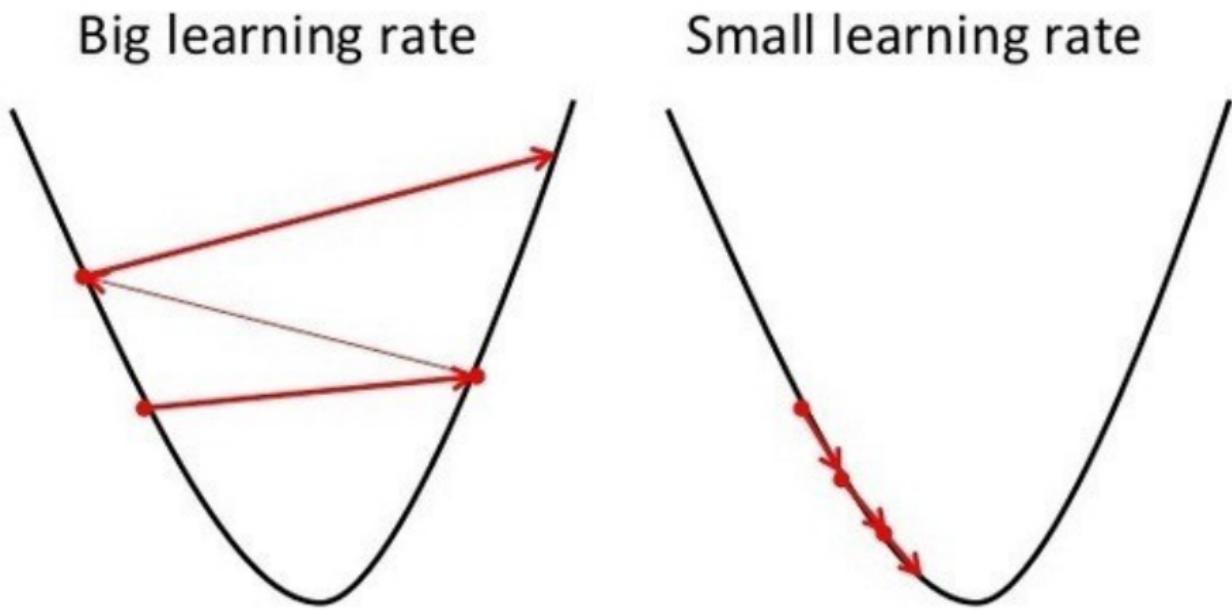
- Começar com um θ_0 e θ_1 qualquer
- Mudar os valores de θ_0 e θ_1 para reduzir o erro, até que se chegue a um mínimo

Regressão Linear - Gradiente Descendente



Regressão Linear - Gradiente Descendente

Regressão Linear - Gradiente Descendente



Regressão Logística

Classificação

- y é um valor discreto

Exemplos:

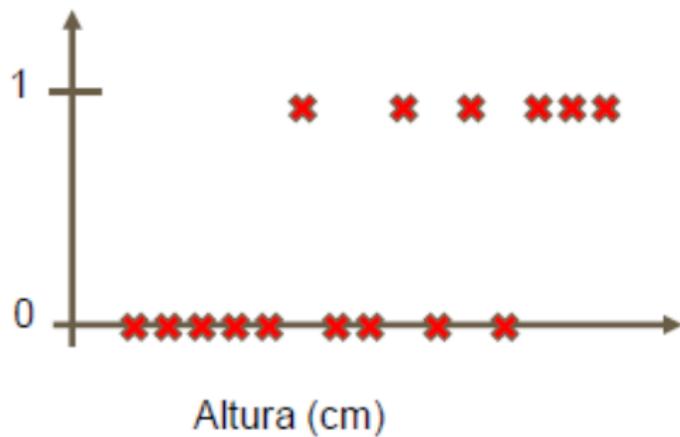
- Email: Spam/Not Spam
- Online Transactions
- Tumor

Y

- is 0
- is 1

Regressão Logística

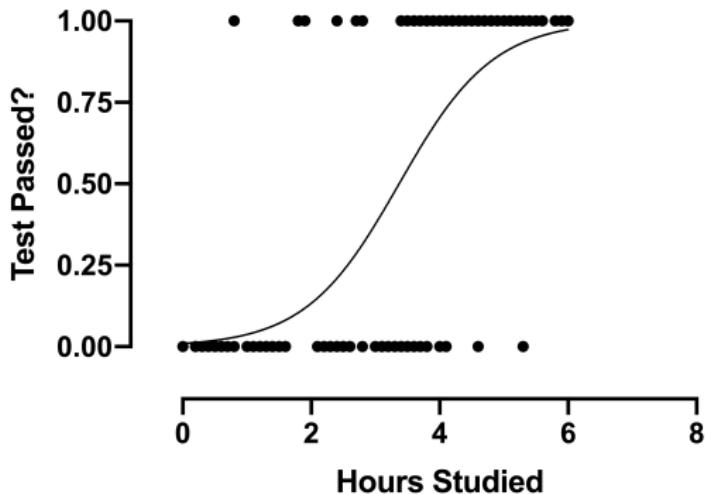
Pessoa Adulto



Classificação: Predição de valores discretos

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

Regressão Logística



Objetivo

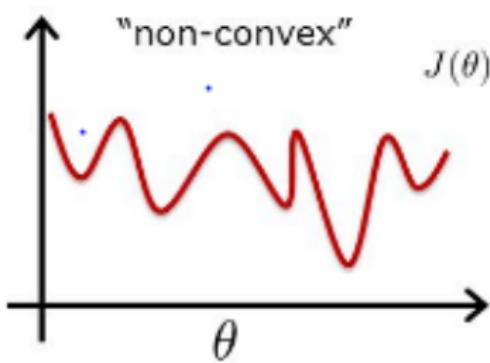
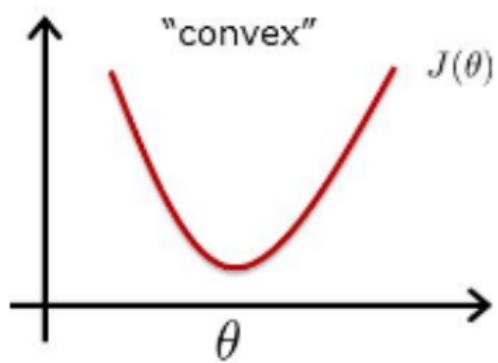
- $0 \leq h(x) \leq 1$
- $h(x) = g(\theta^T X)$
- Sigmoid Function

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

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

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

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

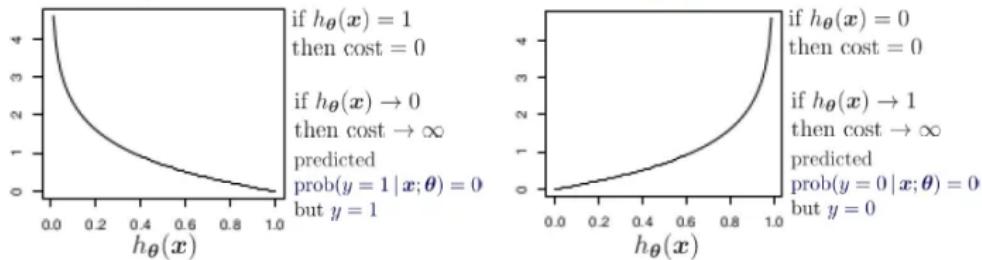
$$\text{Cost}(h_\theta(x), y) = -\log(1 - h_\theta(x)) \quad \text{if } y = 0$$

Regressão Logística - Função de Custo

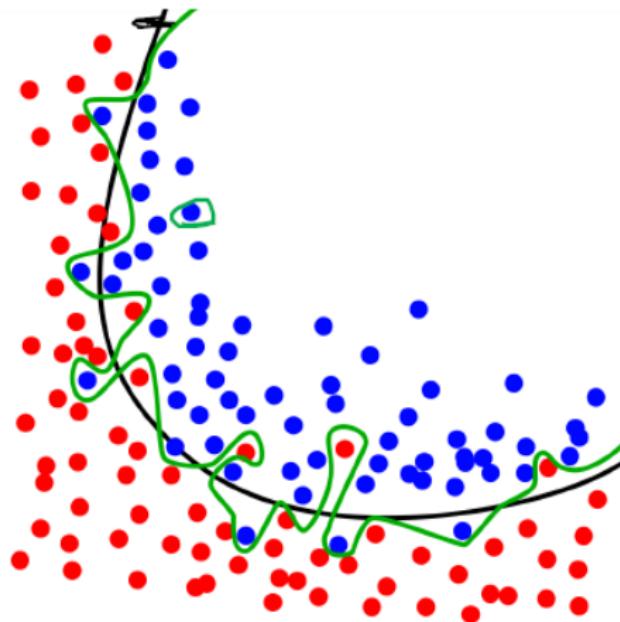
Cost Function for Logistic Regression

"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."

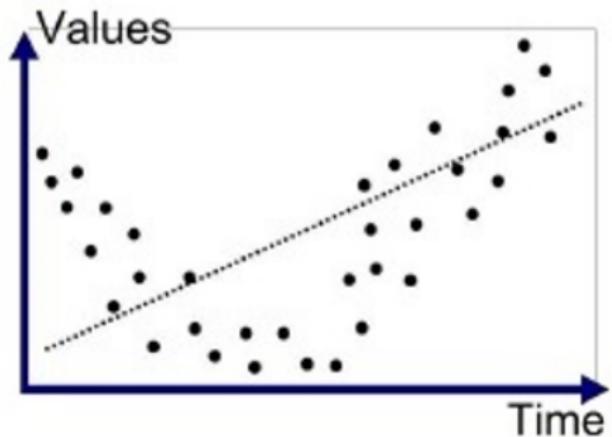
$$\text{cost}(h_{\theta}(\mathbf{x}), y) = \begin{cases} -\log(h_{\theta}(\mathbf{x})) & \text{if } y = 1 \\ -\log(1 - h_{\theta}(\mathbf{x})) & \text{if } y = 0 \end{cases}$$



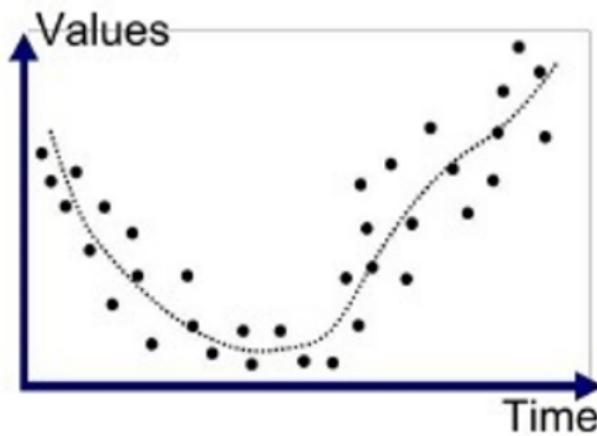
Overfitting e Underfitting



Overfitting e Underfitting

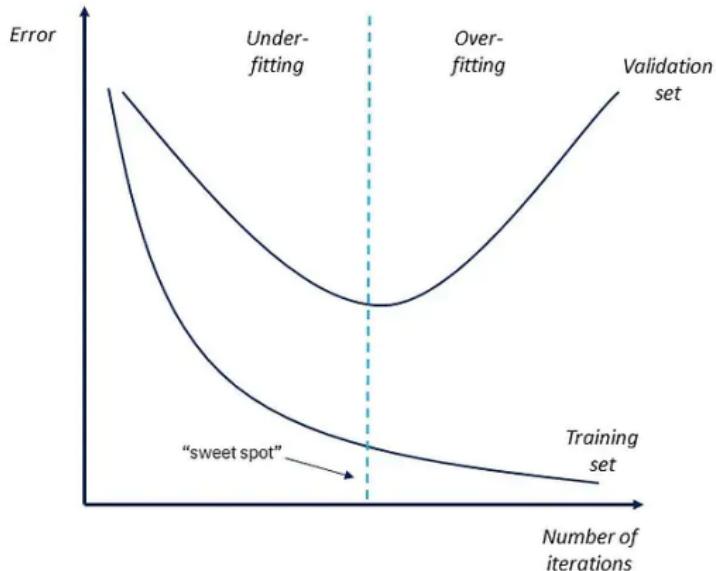


Underfitted

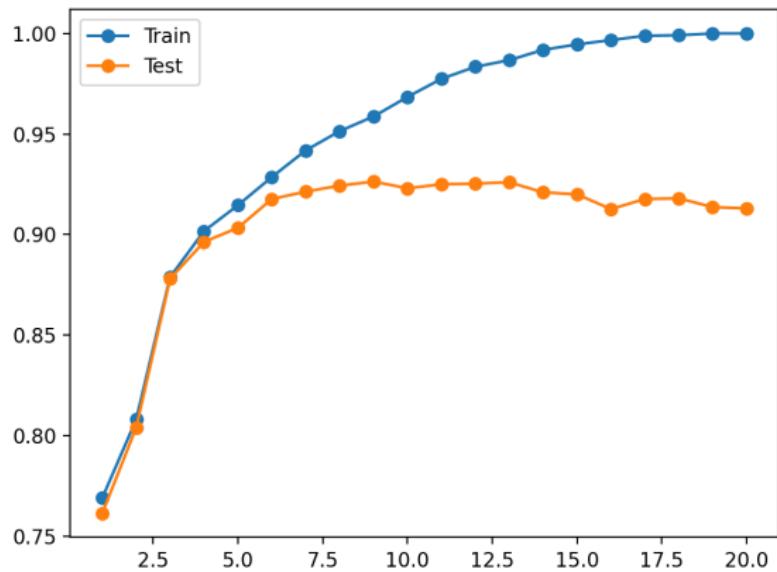


Good Fit/R robust

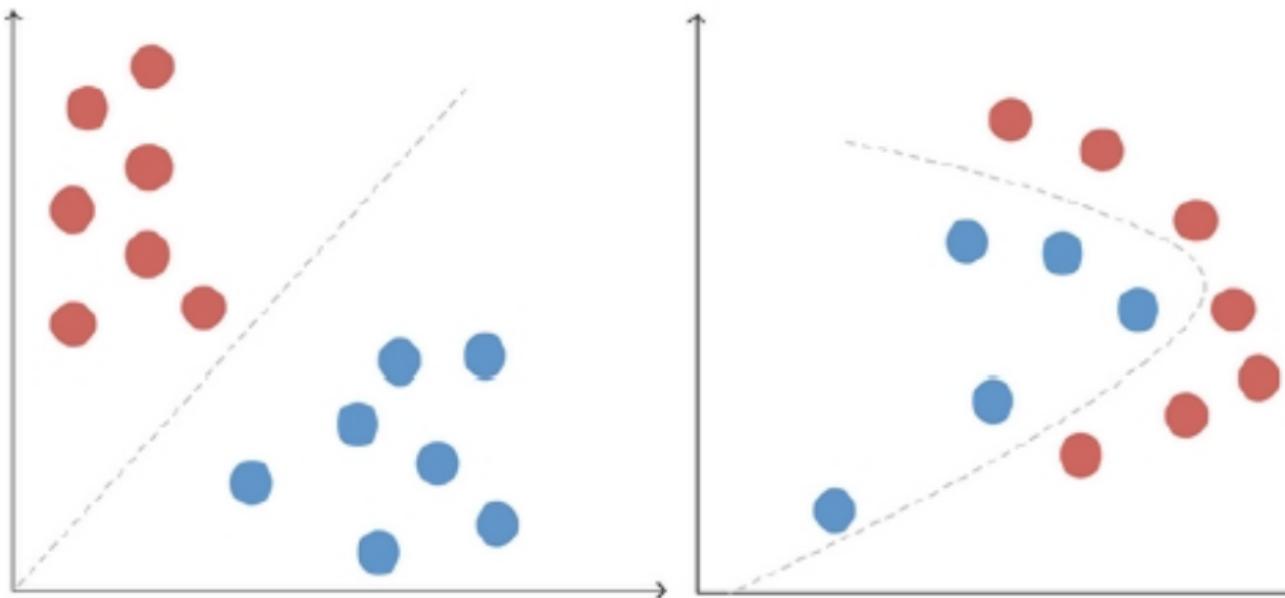
Overfitting e Underfitting



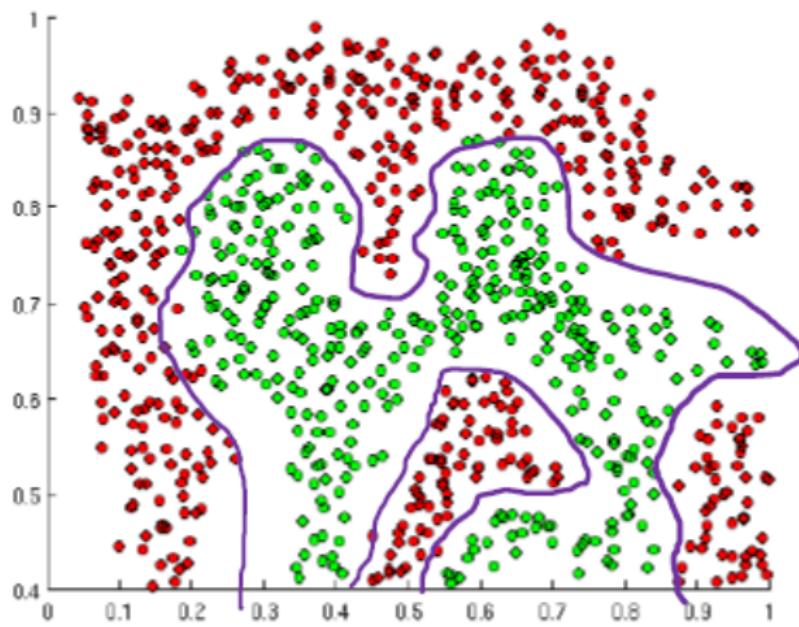
Overfitting



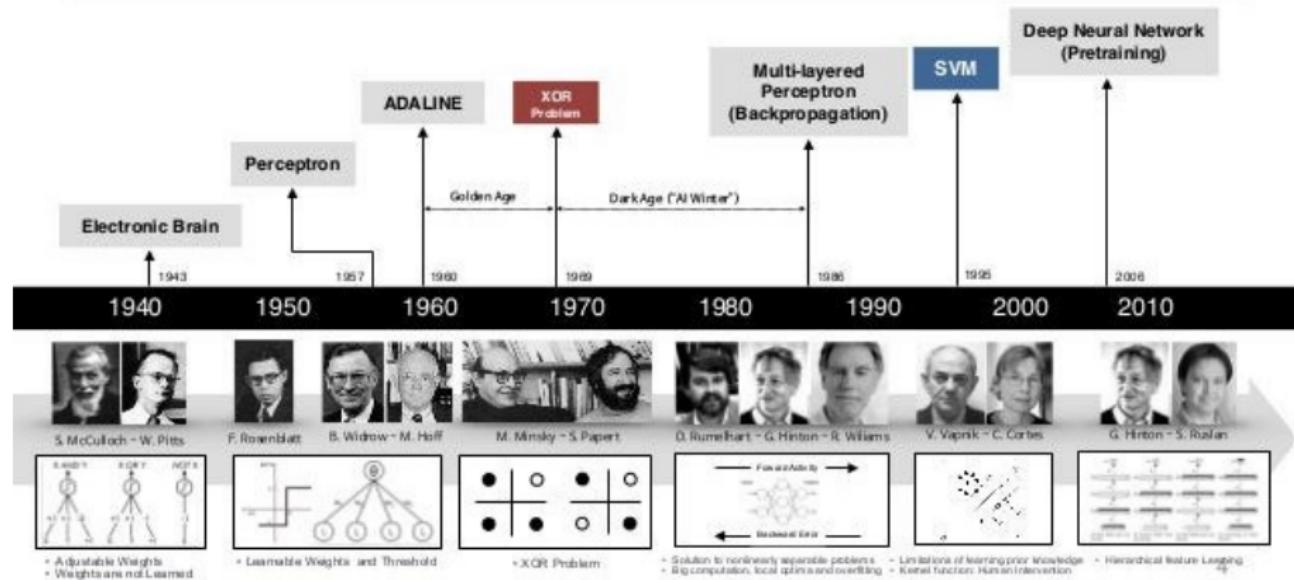
Redes Neurais - Hipótese não-linear



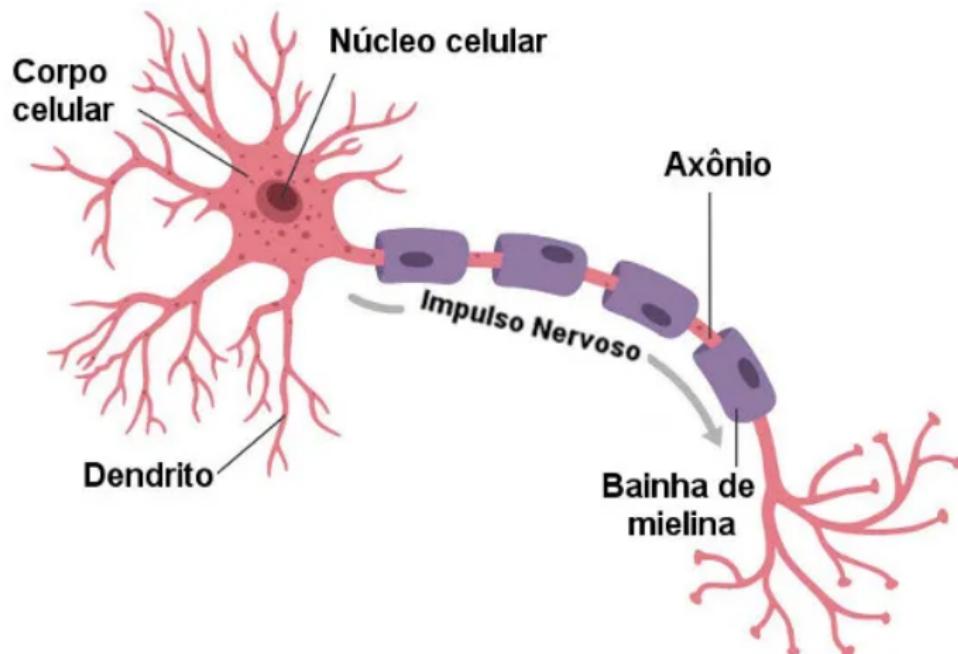
Redes Neurais - Hipótese não-linear



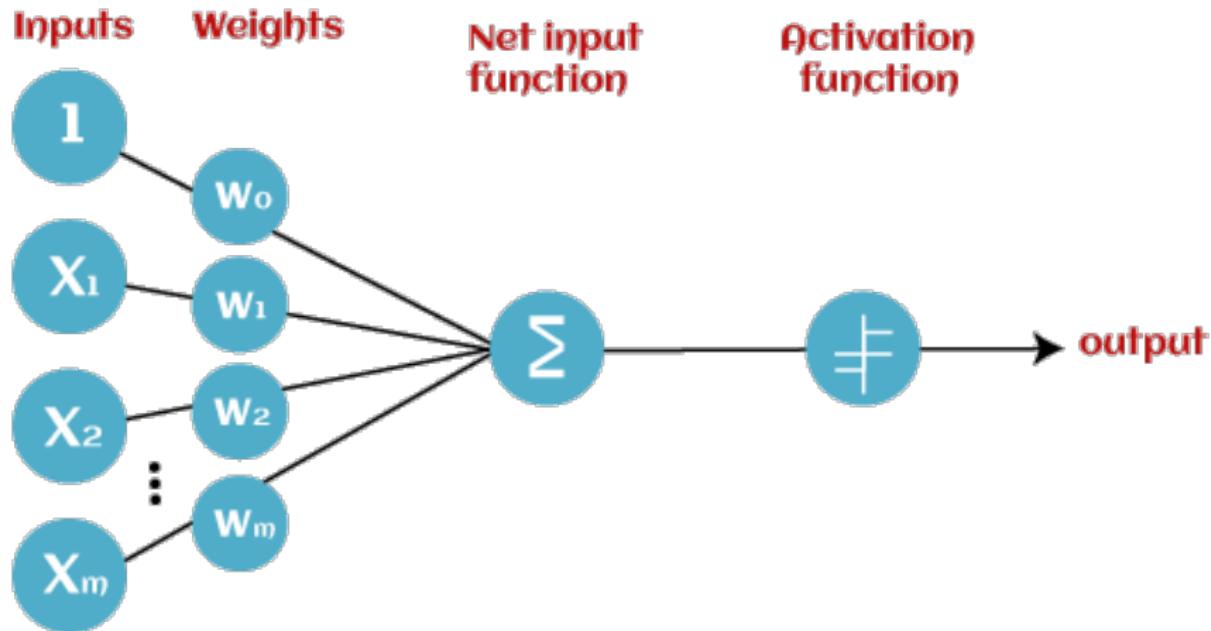
Redes Neurais - História



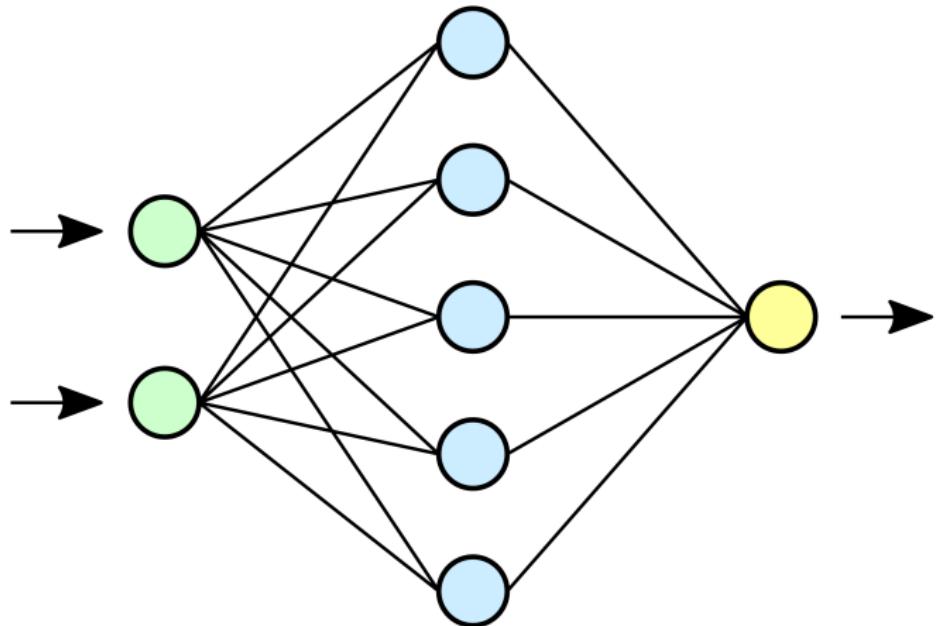
Redes Neurais - Neurônio



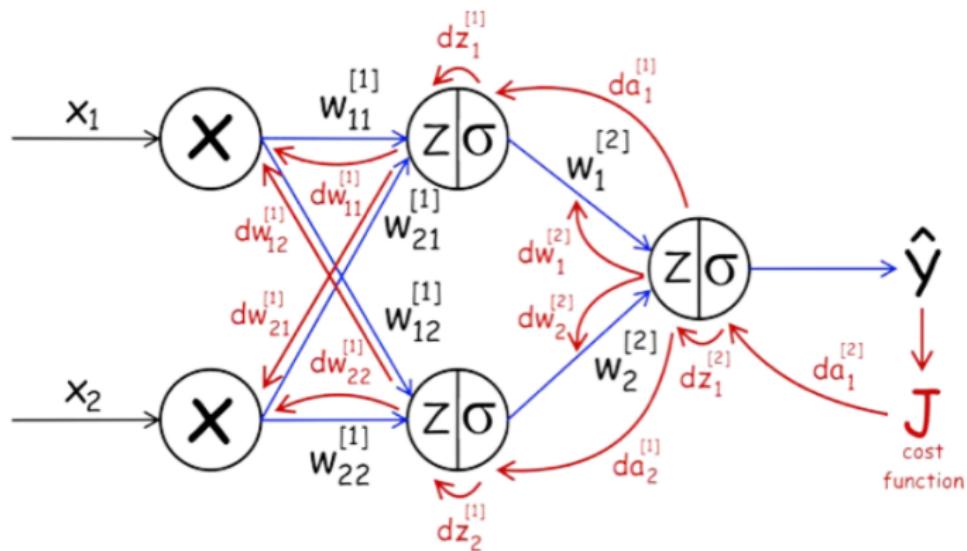
Redes Neurais - Perceptron



Rede Neural



Redes Neurais - Backpropagation



Backpropagation

Redes Neurais - Training

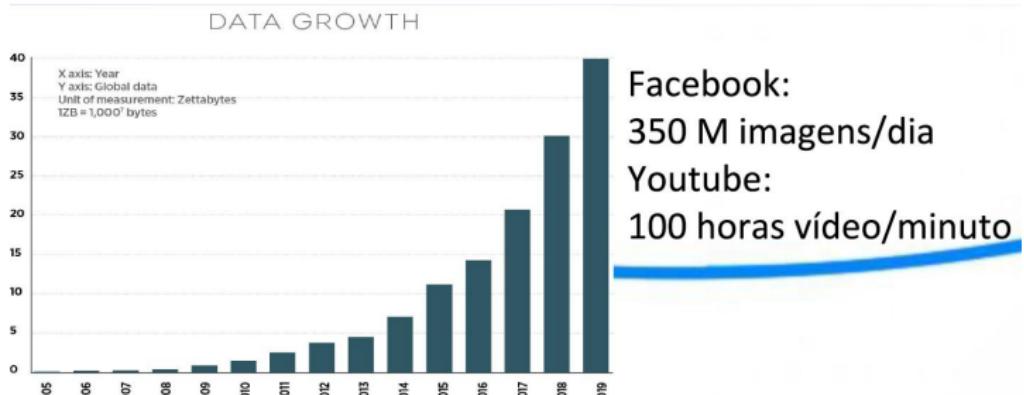
Redes Neurais - Training

Redes Neurais - Training

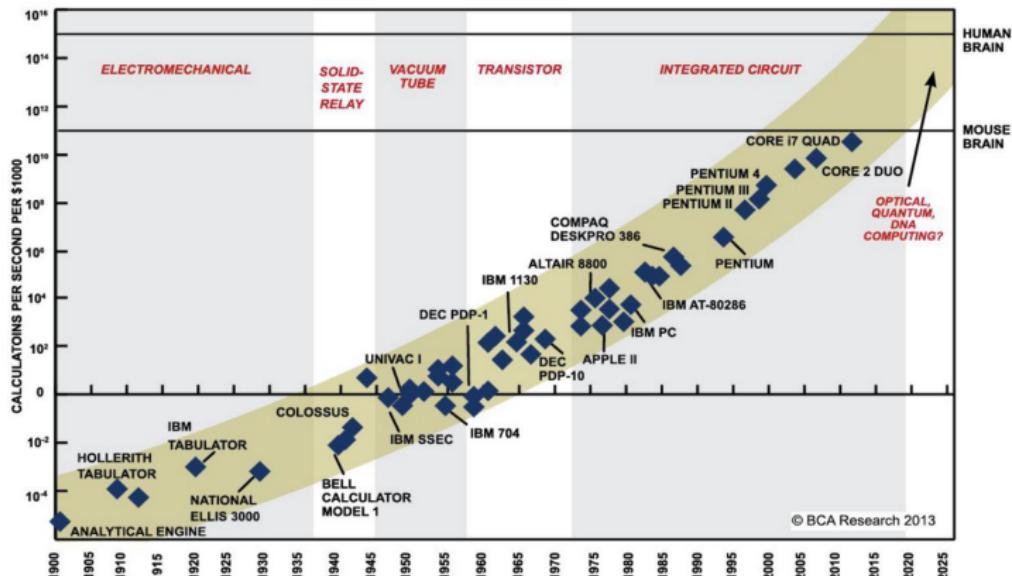


Por que Machine Learning está cada vez mais importante?

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



Por que Machine Learning está cada vez mais importante?

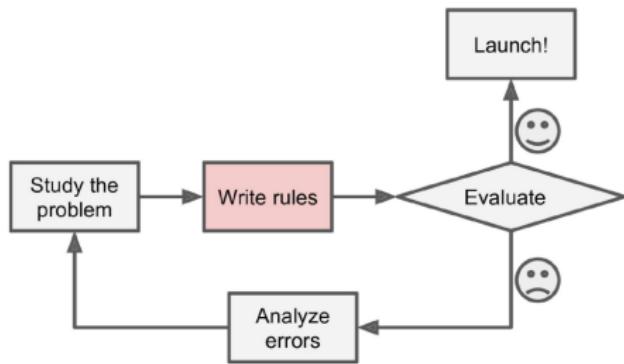


SOURCE: RAY KURZWEIL, "THE SINGULARITY IS NEAR: WHEN HUMANS TRANSCEND BIOLOGY", P.67, THE VIKING PRESS, 2006. DATAPoints BETWEEN 2000 AND 2012 REPRESENT BCA ESTIMATES.

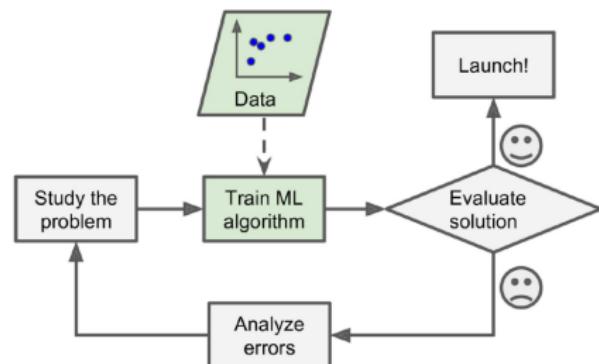


O que é Machine Learning?

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

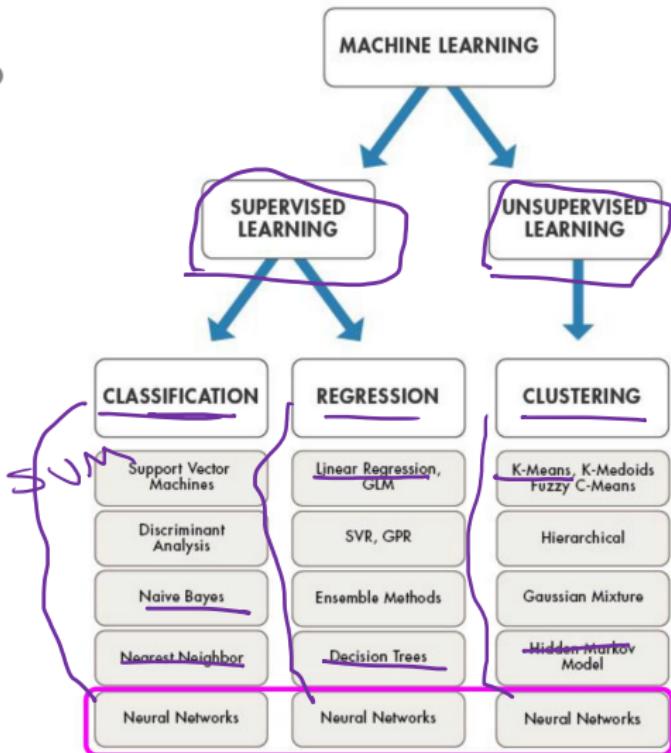


Sistema Especialista

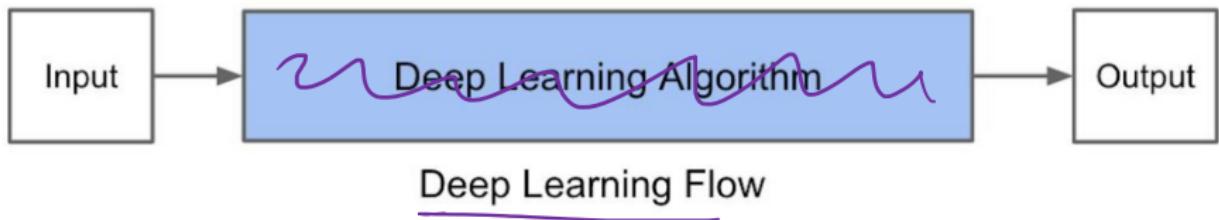
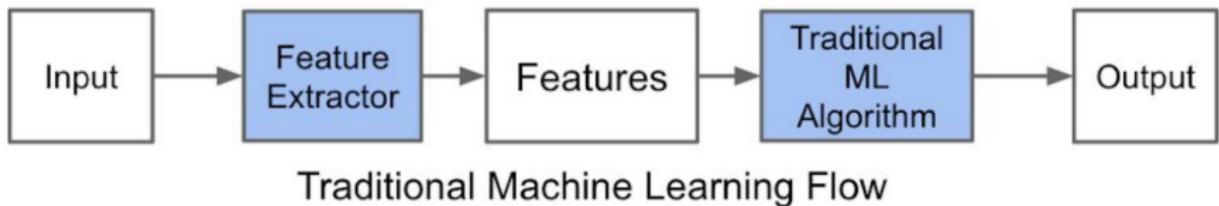


Aprendizado de Máquina

O que é Machine Learning?

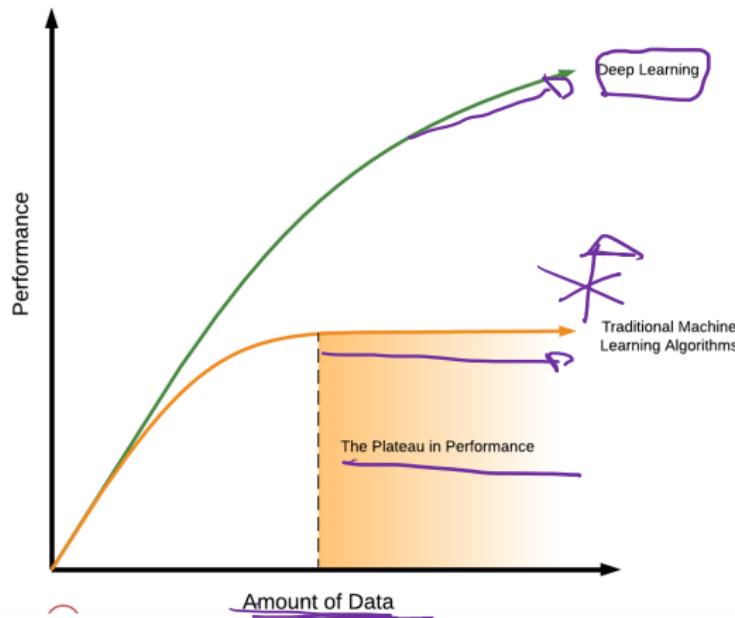


Machine Learning vs Deep Learning

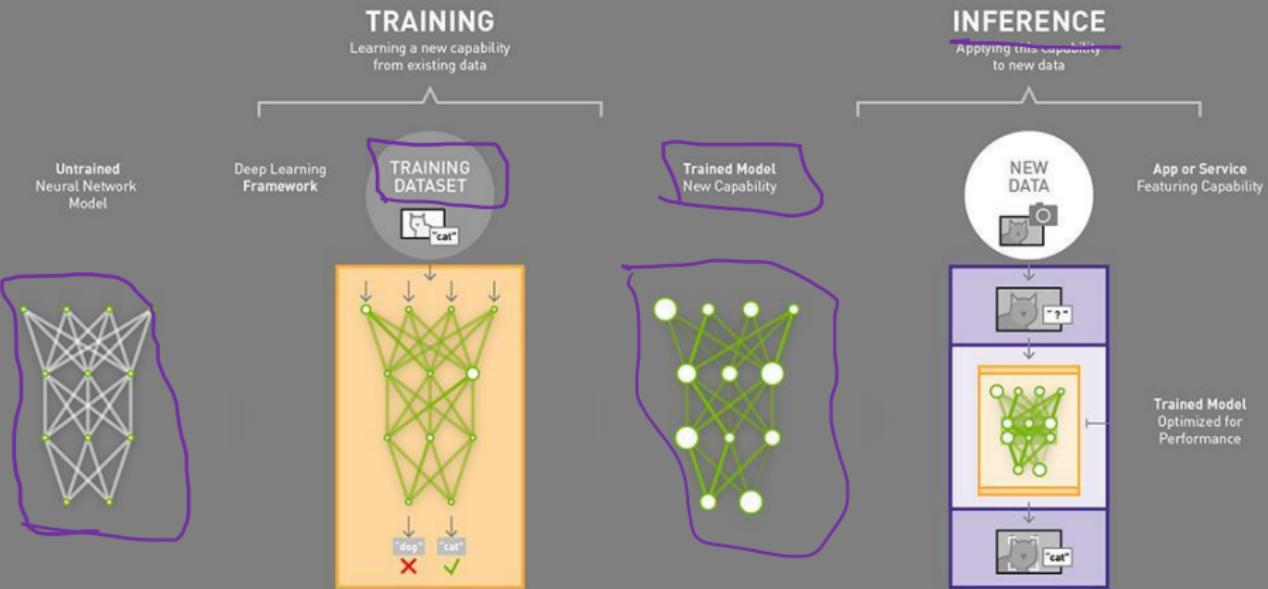




Machine Learning vs Deep Learning



DEEP LEARNING





ImageNET (desde 2010)



IM_{AG}ENET

www.image-net.org

22K categories and 14M images

- Animals
- Plants
- Structures
- Person
- Bird
- Tree
- Artifact
- Scenes
- Fish
- Flower
- Tools
- Indoor
- Mammal
- Food
- Appliances
- Geological Formations
- Invertebrate
- Materials
- Structures
- Sport Activities

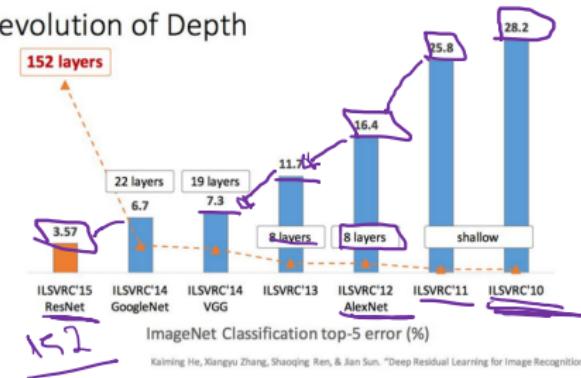


Deng, Dong, Socher, Li, Li, & Fei-Fei, 2009

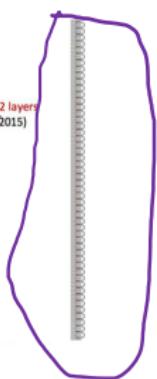
Redes cada vez mais profundas

ResNet, ILSVRC 2015, MSResearch

Revolution of Depth



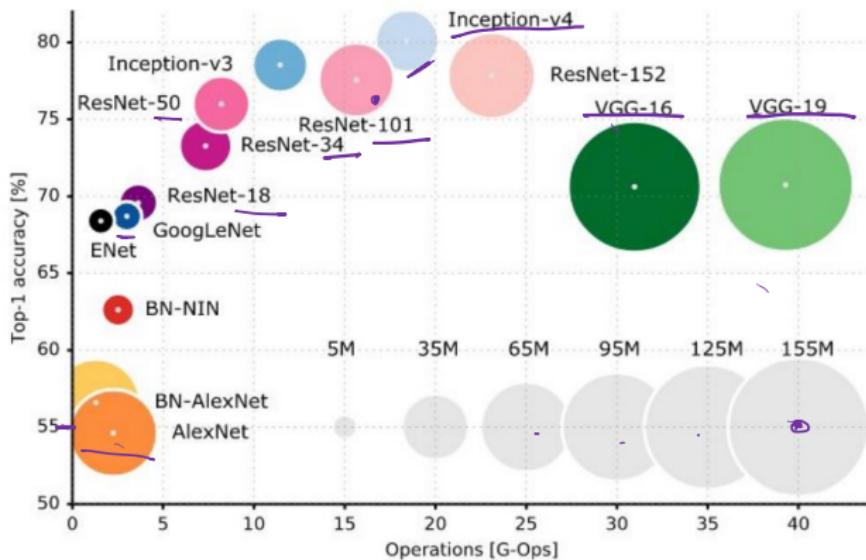
Revolution of Depth



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)

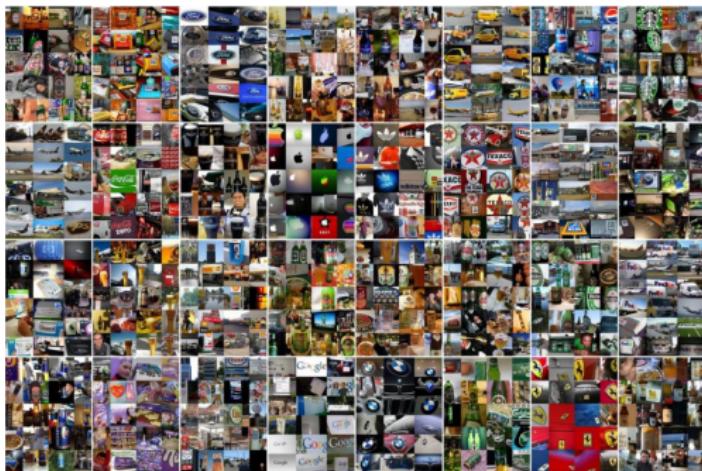
<https://www.remotasks.com/>

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.



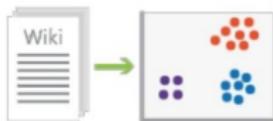
Tipos de Problemas



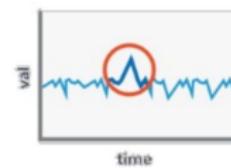
Classification
(supervised – predictive)



Regression
(supervised – predictive)



Clustering
(unsupervised – descriptive)



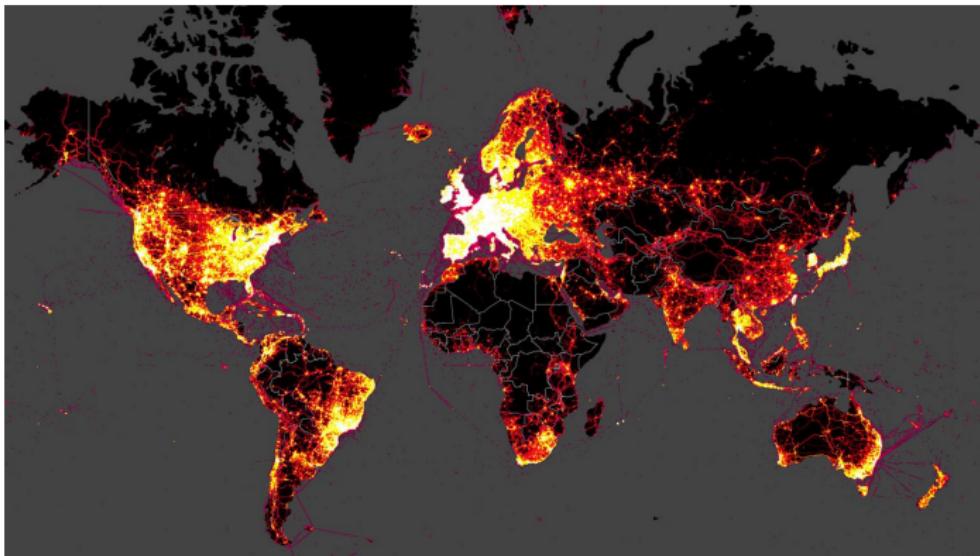
Anomaly Detection
(unsupervised – descriptive)



Como as máquinas aprendem?

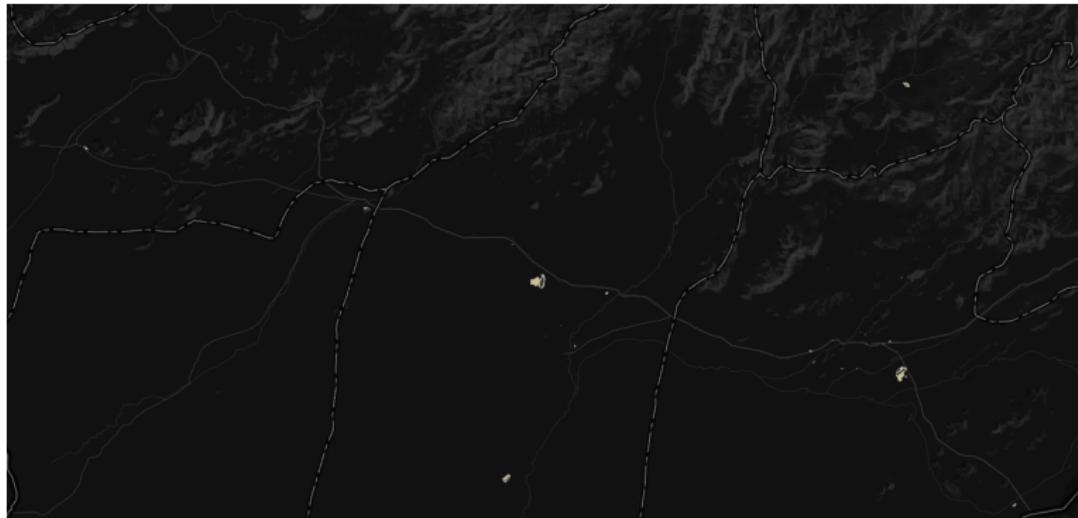


Importância dos Dados





Importância dos Dados



Área remota do Afeganistão



Importância dos Dados



Heatmap de atividades do Strava em base militar secreta

Al no dia-a-dia



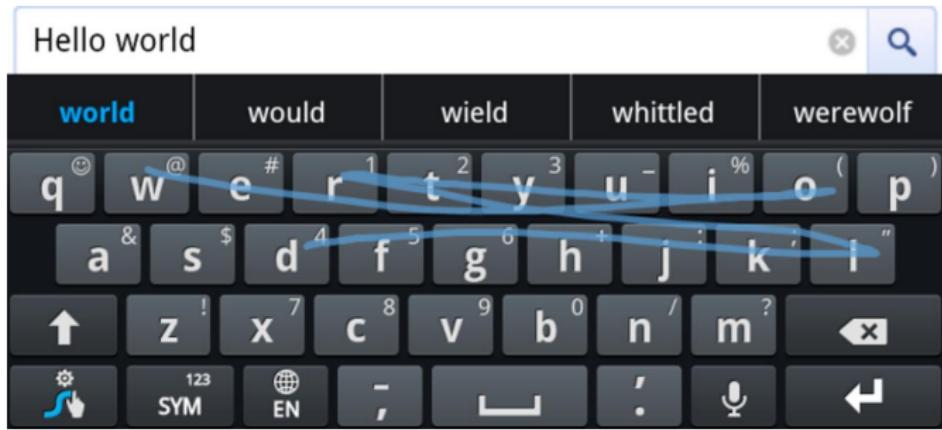
Motorola C115

AI no dia-a-dia



Motorola C115

Hello world



A virtual QWERTY keyboard is displayed over a white background. The keyboard is dark grey with white text. It includes a numeric keypad, a symbol key, and various function keys like 123, SYM, EN, and punctuation. Above the keyboard, a text input field shows the phrase "Hello world". The word "world" is highlighted in blue, indicating it is the current selection or being typed. A blue line with arrows shows the path of a finger as it types each letter of "world" sequentially from left to right. The entire interface is enclosed in a light blue border.

AI no dia-a-dia



AI no dia-a-dia

in:spam

As mensagens que ficarem mais de 30 dias na pasta "Spam" serão excluídas automaticamente. [Excluir todas as mensagens de spam agora](#)

<input type="checkbox"/>	<input type="star"/>	<input type="circle"/>	Rede
Aceite Elo e as principais bandeiras! - Aproveite o Preço Único por 50% do valor por 1 ano. Aceite todas as principais bandeiras e não perca vendas. Você ainda conta com taxa zero 1...			
18:49			
<input type="checkbox"/>	<input type="star"/>	<input type="circle"/>	Poup
O dia delas está chegando! - Toda mãe tem um cantinho especial no coração dos filhos... Poup Alameda Xingu, 512, São Paulo, 06455-972 Unsubscribe - Unsubscribe Preferences			
16:18			
<input type="checkbox"/>	<input type="star"/>	<input type="circle"/>	Clube do Malte
SEXTOU! Que tal um copo de cerveja na faixa, para curtir o final de semana com estilo? - Caso não esteja visualizando, acesse o preview aqui. Temos novidade em nossa loja: A Gold...			
15:07			
<input type="checkbox"/>	<input type="star"/>	<input type="circle"/>	kais 6
Marketing Software for http://casadei.io - Hi, This is Kais and I'd like to inform you that we have launched a digital marketing tool for Outreach namedAdinject. I'm glad to offer you our...			
13:36			
<input type="checkbox"/>	<input type="star"/>	<input type="circle"/>	Joanne Golbroda
Call for Exhibitors and Sponsors The Int'l Congress on Precision Medicine - For PMBC 2018 Sponsorship & Exhibition Click here PMBC2018._s... Click here for accessibility BioEven...			
12:17			
<input type="checkbox"/>	<input type="star"/>	<input type="circle"/>	Pão de Açúcar
Selos em Dobro* + Frete Grátis = carrinho cheio 🎉 - Hoje é dia de encher o carrinho de alegria! :) Caso não esteja visualizando este e-mail, veja aqui Caso você não esteja visualizan...			
10:21			
<input type="checkbox"/>	<input type="star"/>	<input type="circle"/>	comix
Ofertas COMIX - Mangás e HQs a partir de R\$1,00 - Caso não consiga visualizar o e-mail, acesse este link. Você está recebendo este e-mail porque está cadastrado em nosso mailin...			
05:43			
<input type="checkbox"/>	<input type="star"/>	<input type="circle"/>	environmental-studi.
Early bird submission - Advances in Environmental Studies - Dear Dr. Vitor Casadei, Greetings from our editorial team. Hoping that you might be busy and couldn't reply/contribute to ...			
04:22			
<input type="checkbox"/>	<input type="star"/>	<input type="circle"/>	DDMáquinas
Sabe aquele frete? Na DD Máquinas não tem!! #semanafretegrátis - Problemas para visualizar a mensagem? Acesse este link. (Optional) This text will appear in the inbox preview, but...			
10 de mai			
<input type="checkbox"/>	<input type="star"/>	<input type="circle"/>	Rinodrigo Persona
Seu PC está lento? A BlackRhino resolve. - (Optional) This text will appear in the inbox preview, but not the email body. Não é chato quando você está navegando e de repente seu PC ...			
10 de mai			

AI no dia-a-dia

Principal

Social 1 novo Twitter

Promoções 3 novos TudoAzul, Spotify, Kaggle Team

Atualizações 2 novos Videostream Team, Alemão Sem...

Fóruns 1 novo Infowrede via Possor

⭐ Isaac, eu 2 tradução Juramentada italiano e com as apostilas de Haia - Olá, eu recebi sim. Só estou esperando uma confirmação sobre uma informação de um dos documentos para de fato soli... 10:34
 Mai...

⭐ Isaac Salomon - Apo. tradução Juramentada italiano e com as apostilas de Haia - Vitor, bom dia. Podemos lhe ajudar com a tradução Juramentada italiano e com as apostilas de Haia. RESUMO PRAZO E C... 9 de mai
 Mai...

⭐ Gleyce, eu 2 Tradução juramentada - Olá, seguem os documentos para orçamento de tradução e apostilamento para italiano. 8 de mai
 Vitor Casadei T...  Domingos_Obit... Mai...



AI no dia-a-dia



Reply

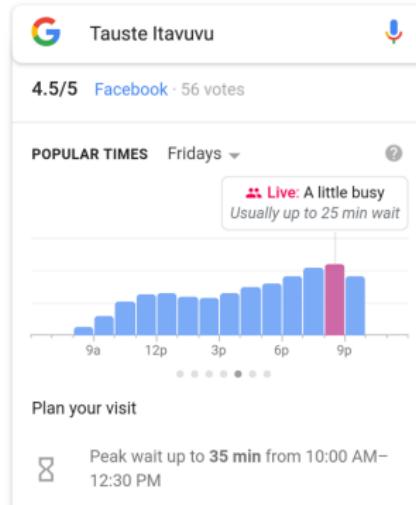


Obrigado!

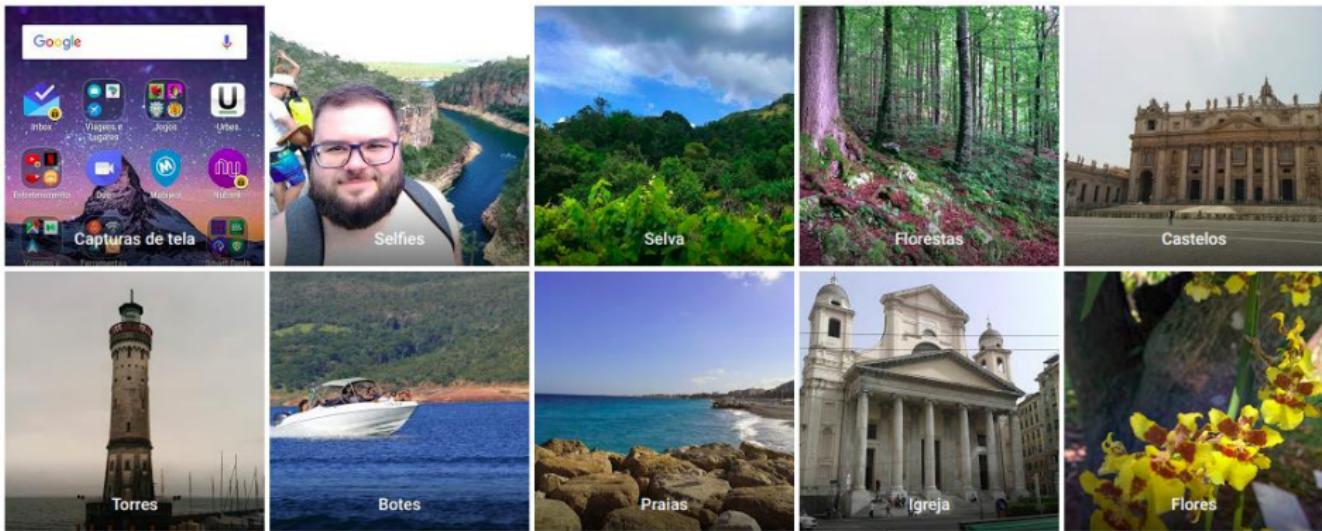
Muito
obrigado!

Ok.

AI no dia-a-dia



AI no dia-a-dia



AI no dia-a-dia



Al no dia-a-dia



Mingal

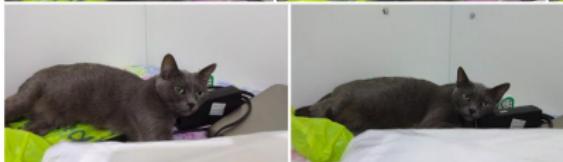
dom, 28 de jan



qui, 25 de jan



seg, 22 de jan



AI no dia-a-dia



Frida

sáb, 30 de dez de 2017



dom, 19 de nov de 2017



qui, 6 de jul de 2017



Al no dia-a-dia



X

Good Lens review!



Potcake dog



Black and Tan
Terrier



Miniature
Pinscher