

MBA
USP
ESALQ

Outros modelos de Machine Learning III

João F. Serrajordia R. de Mello

Você vai precisar de...

Preparativos

- Abrir o R
- Importar as bibliotecas
- Algo para fazer suas anotações



Agenda

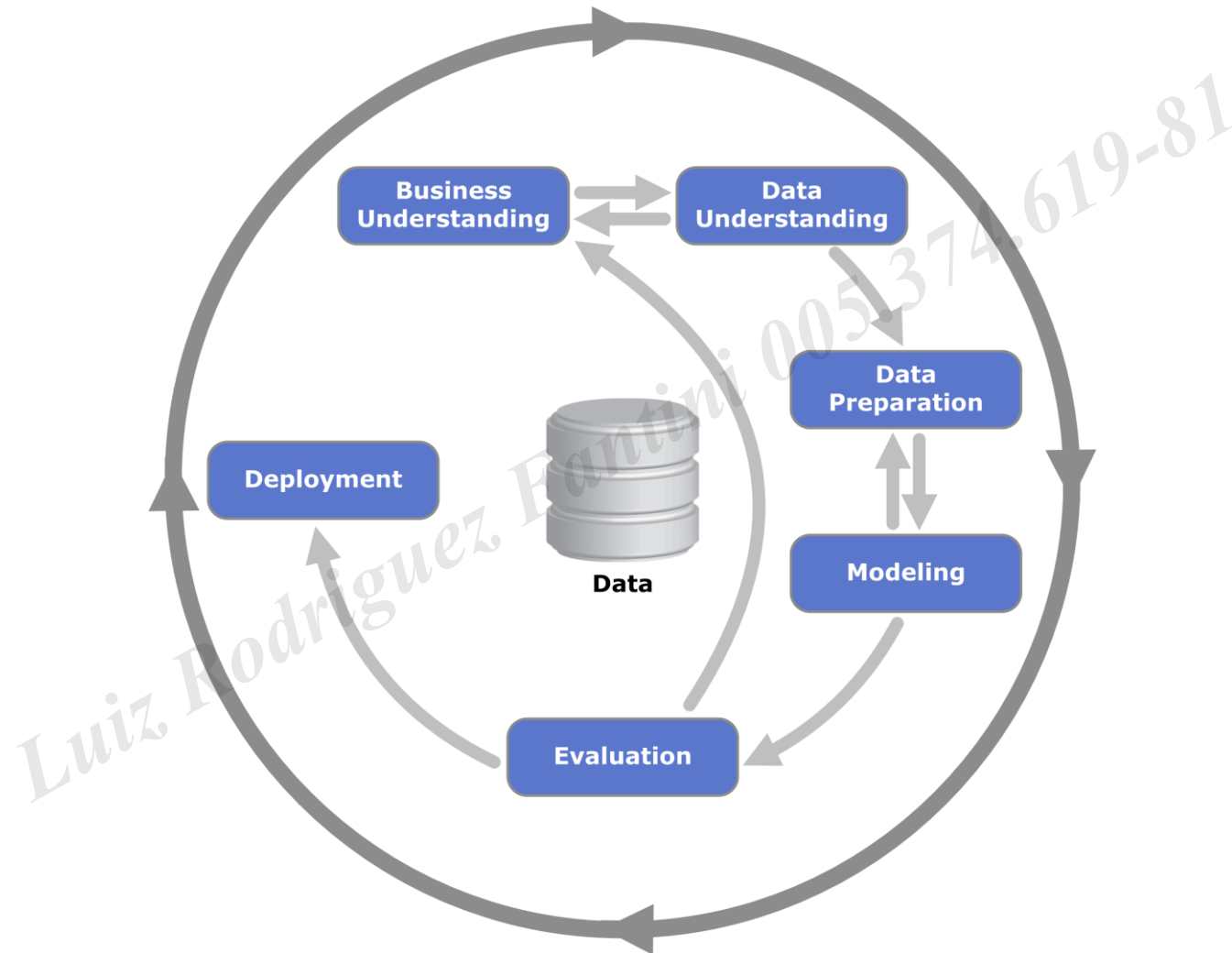
Revisão

Histórico

Ideias básicas

Usos

CRISP-DM



Fonte: <https://www.the-modeling-agency.com/crisp-dm.pdf>



Ensemble

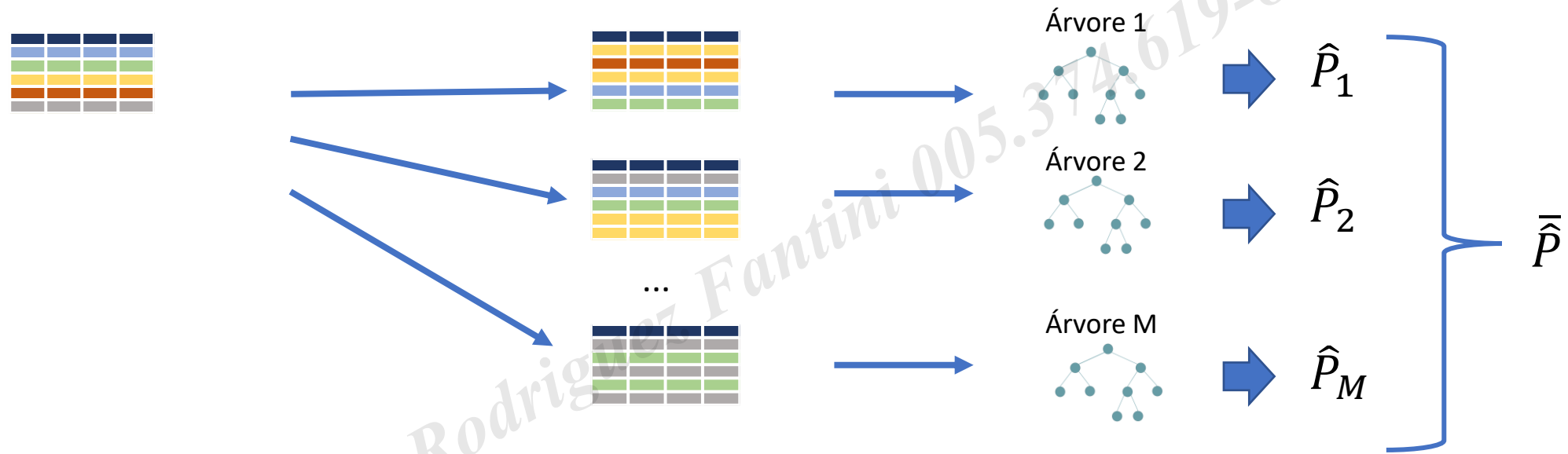
Um ensemble é qualquer mistura de modelos já existentes.
Os principais tipos são:

Bagging

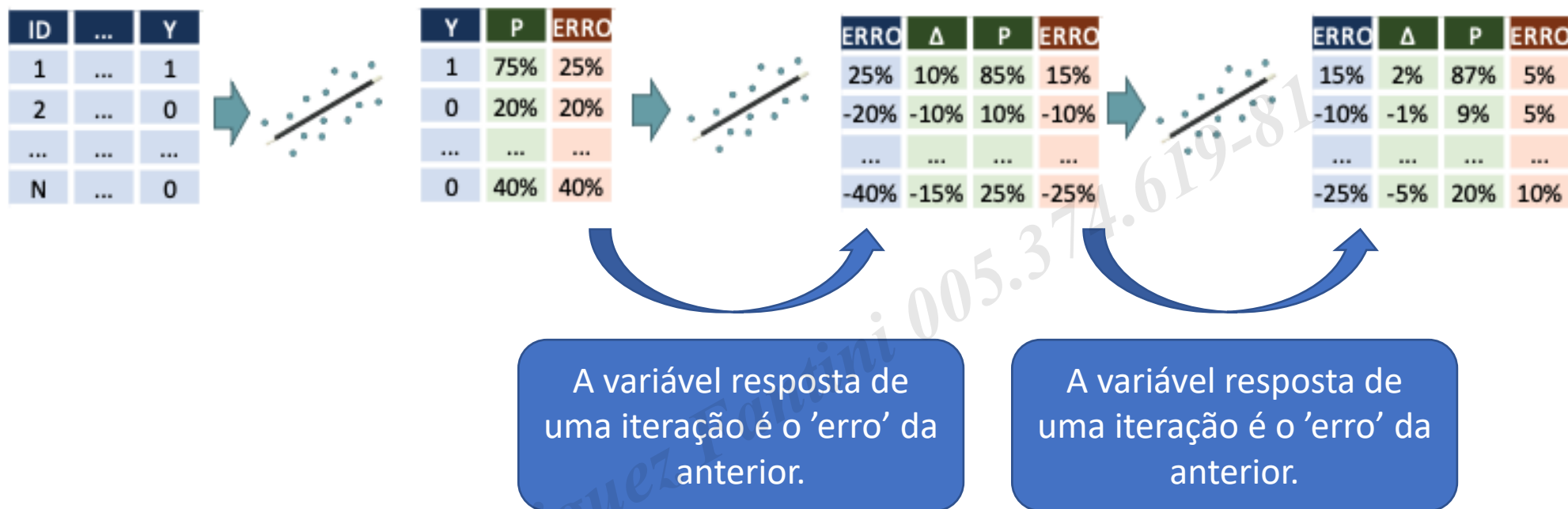
Boosting

Stacking

Bootstrap – aggregation (bagging)



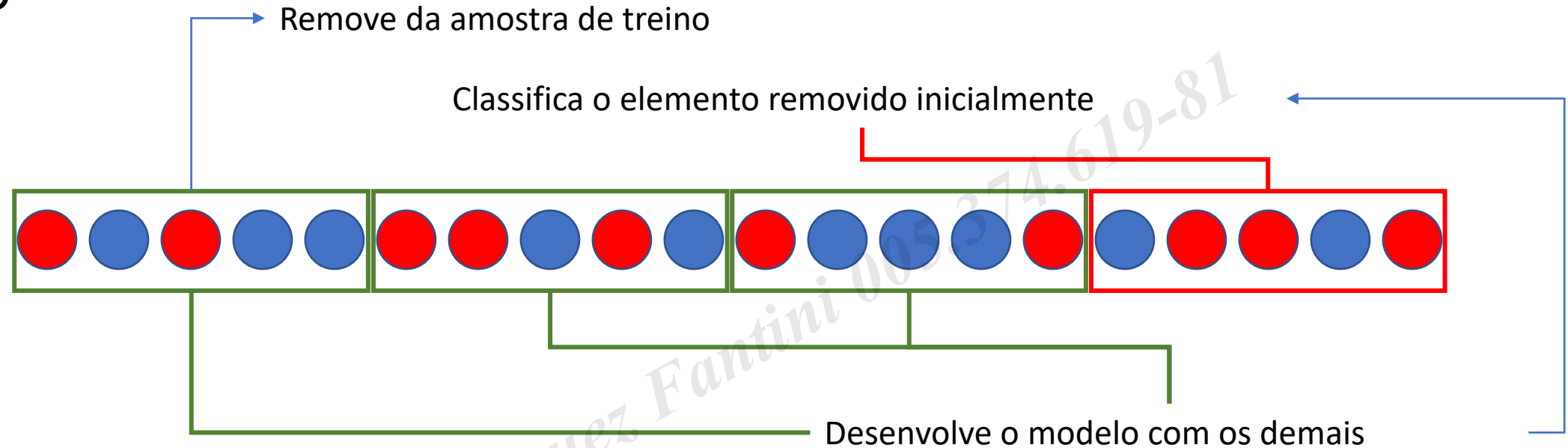
O bagging com árvores é o famoso *Random Forest*



Boosting

- Os métodos de *boosting* são modelos sequenciais que tentam melhorar o erro do modelo anterior

K-fold



- Dividimos a base em k sub-amostras
- Para cada sub-amostra:
 - Removemos a sub-amostra como validação
 - Treinamos o modelo com as observações restantes
 - Utilizamos este modelo para classificar a sub-amostra removida
 - Avaliamos a métrica de desempenho do modelo
- Calculamos a média das métricas de desempenho do modelo

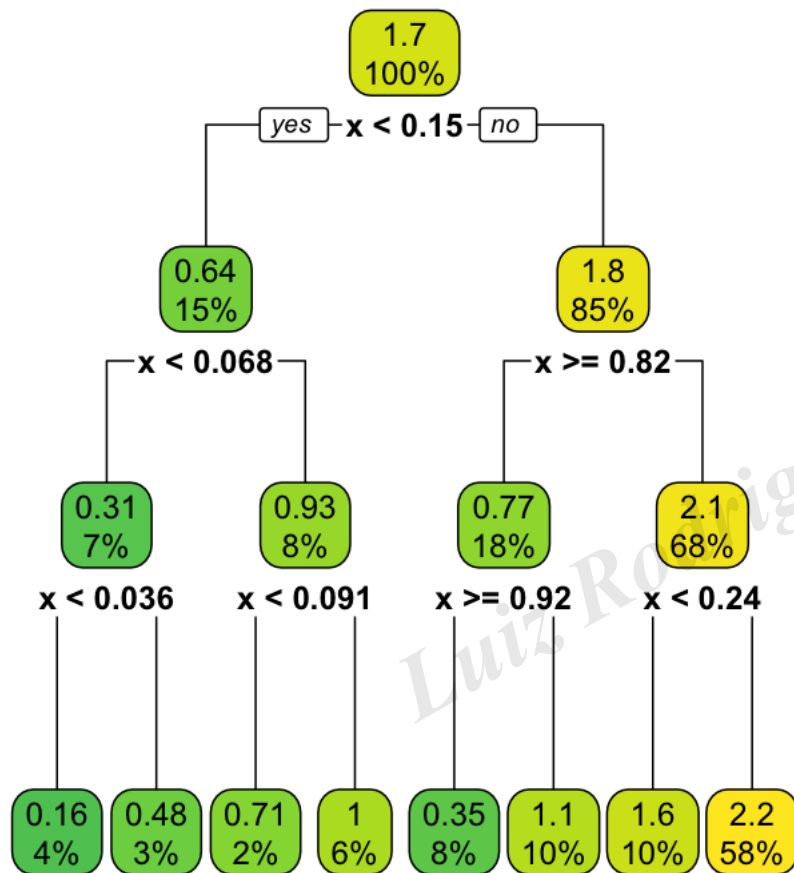
Árvores de regressão

São muito semelhantes a árvores de classificação

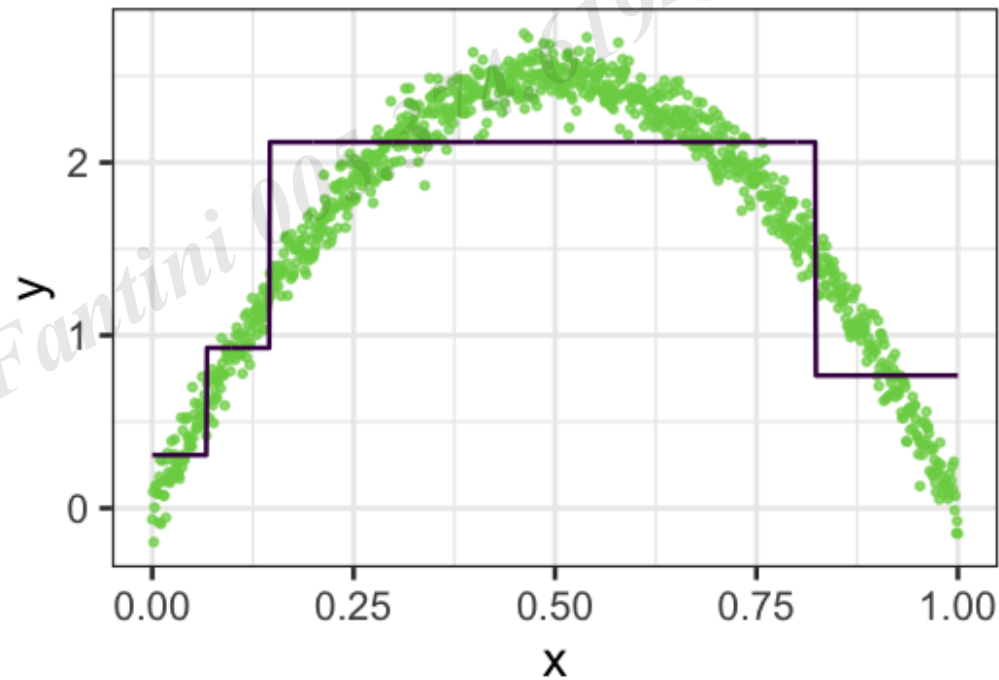
O que muda é o critério de impureza

$$SQE = \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

Árvores de regressão



Valores observados vs esperados



Dado: — Esperado — Observado

Problemas de preditivos e de classificação



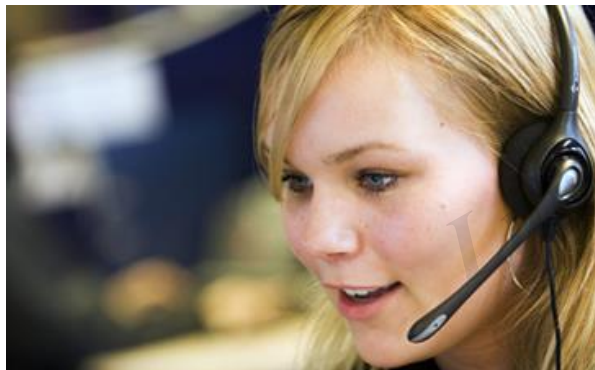
Qual a eficácia de uma vacina?



O cliente vai pagar o empréstimo?



Quanto de petróleo tem no poço?



O cliente vai comprar meu produto?



O que a pessoa está fazendo?



Quão ecológico esse veículo é?

Classificação

Luiz Rodriguez Fantini 005.374.619-81

Classificação dos algoritmos



Supervisionados

- Regressão
- GLM
- GLMM
- Support vector machines
- Naive Bayes
- K-nearest neighbors
- Redes Neurais
- Decision Trees



Não supervisionados

- K-Means
- Métodos hierárquicos
- Mistura Gaussiana
- DBScan
- Mini-Batch-K-Means

Estamos aqui!

Classificação dos algoritmos



Resposta contínua

- Regressão
- GLM
- GLMM
- Support vector machines
- K-nearest neighbors
- Redes Neurais
- Regression Trees



Resposta discreta

- Regressão logística
- Classification trees
- Redes Neurais
- GLM
- GLMM

Estamos aqui!

Classificação dos algoritmos

Métodos Machinelânicos

- Árvores de decisão
- Bagging
- Boosting
- K-NN
- Redes Neurais
- Support Vector Machines

Métodos Machinelânico- estatísticos

- Regressão
- GLM
- GLMM
- ANOVA

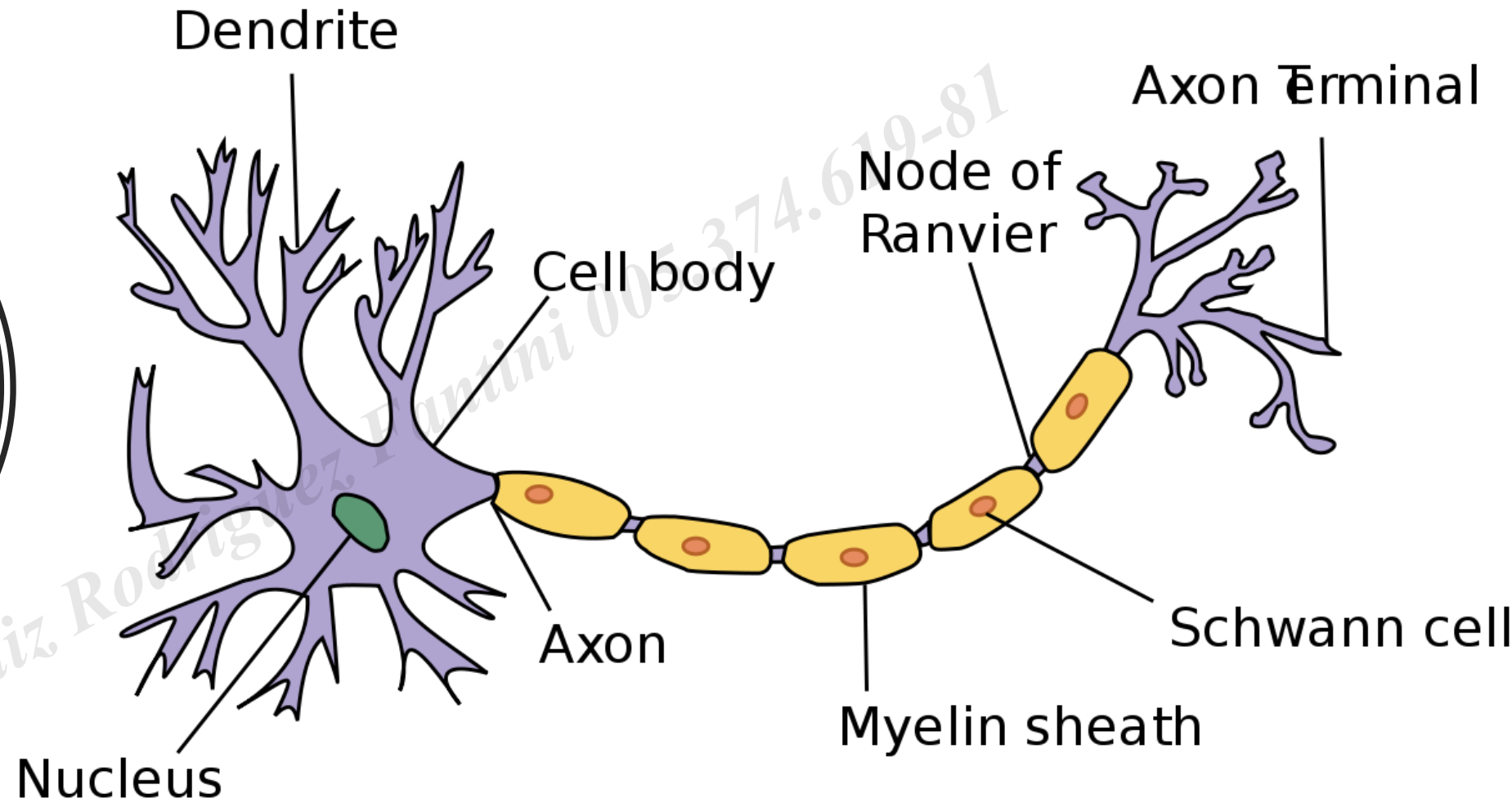
Estamos aqui!



Redes Neurais Artificiais

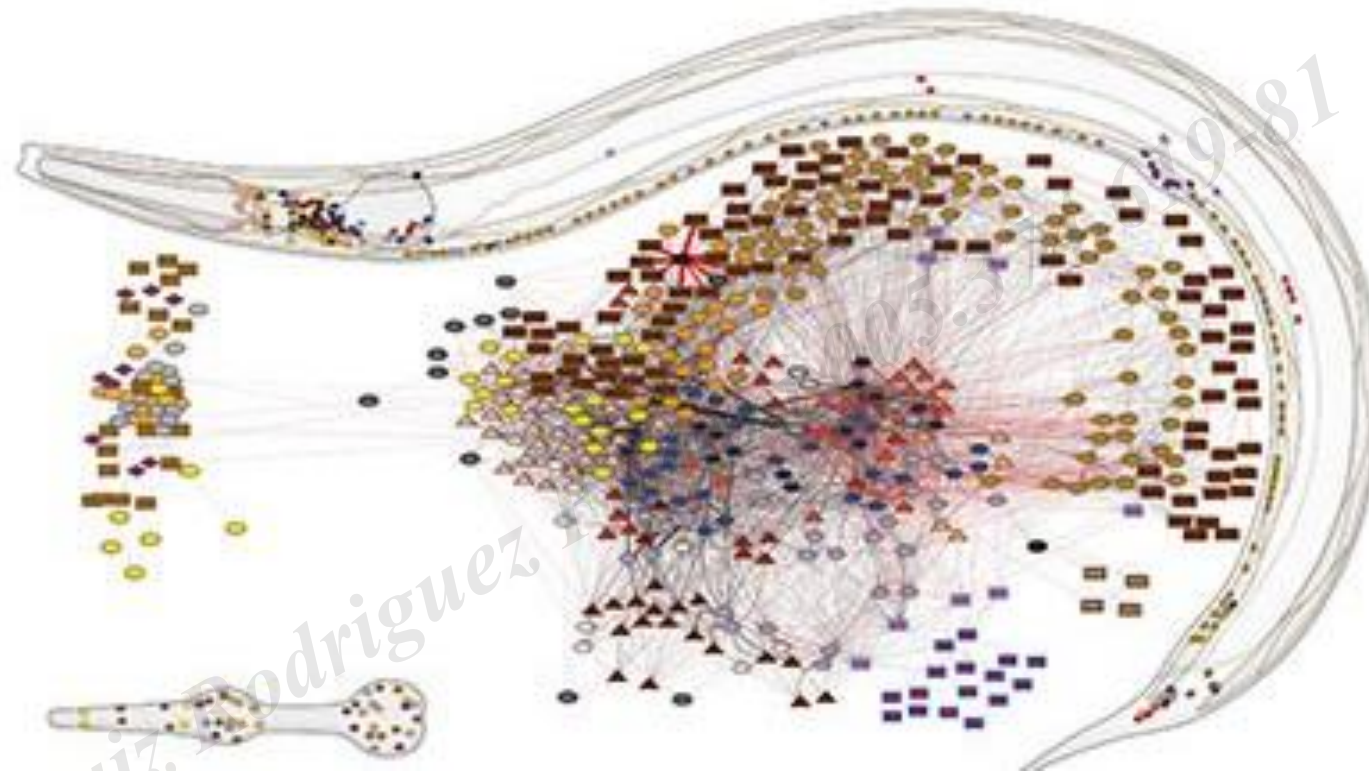
Iliz Rodriguez Portini 005.374.079-81

Metáfora



<https://en.wikipedia.org/wiki/Myelin>

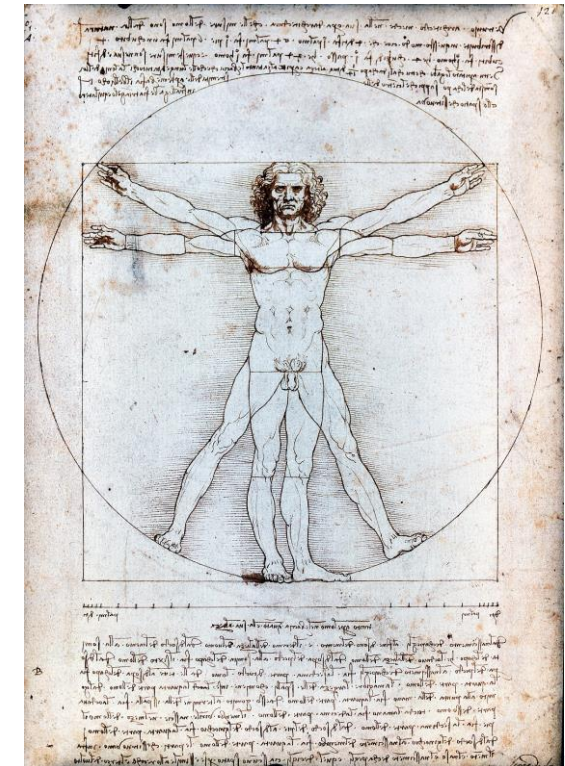
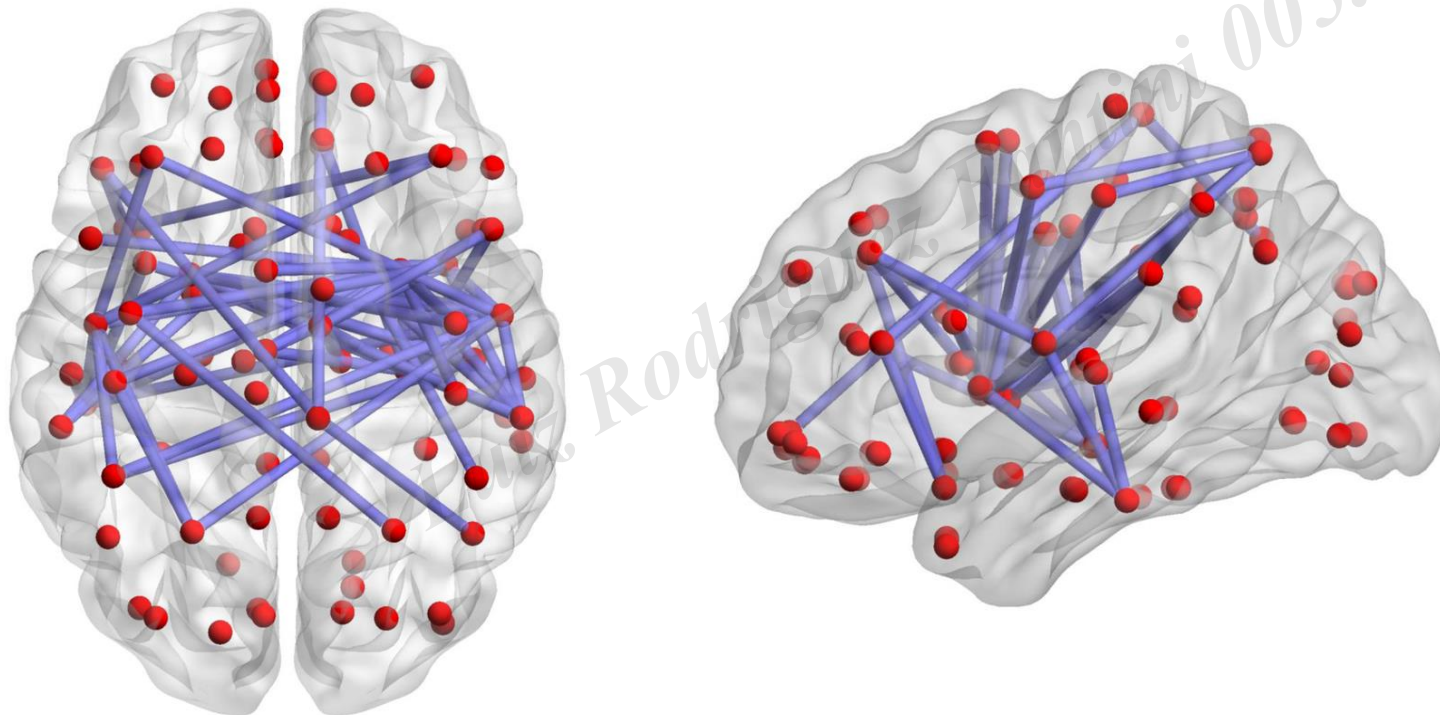
Exemplo biológico



- *Nematelminto*: 302 neurônios

Rede Neural Humana

- *Homo sapiens*: 100.000.000.000 de neurônios

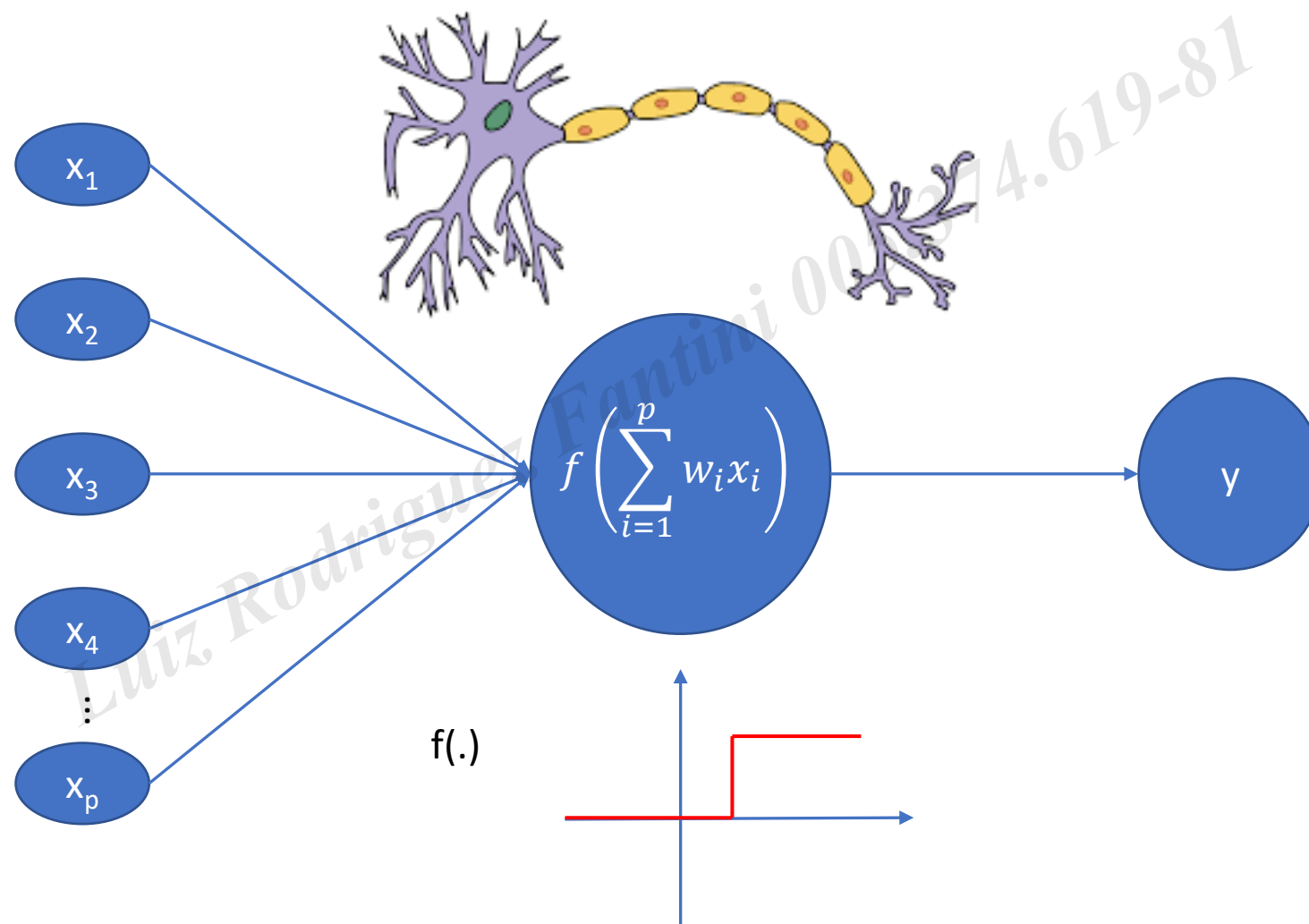


Onde vivem?



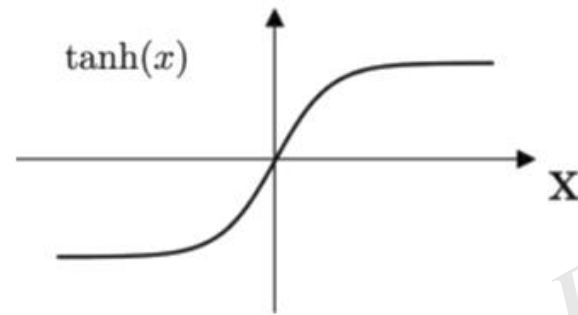
Redes Neurais Artificiais têm tido muito sucesso em problemas com dados pouco estruturados como imagens, áudios, textos e vídeos.

Neurônio de McCulloch-Pitts

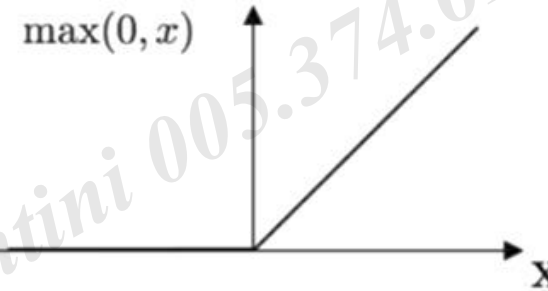


Funções de ativação

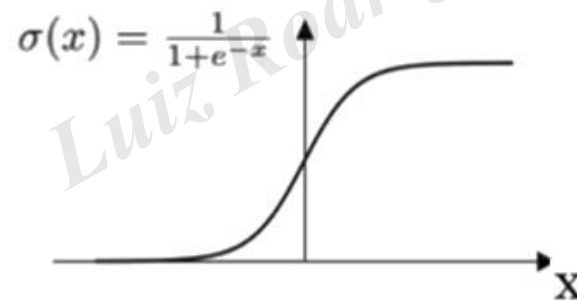
Tanh



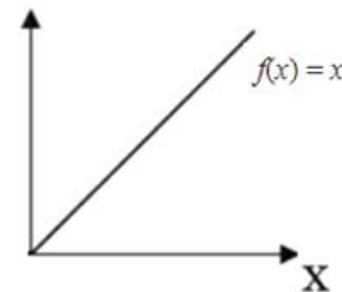
ReLU



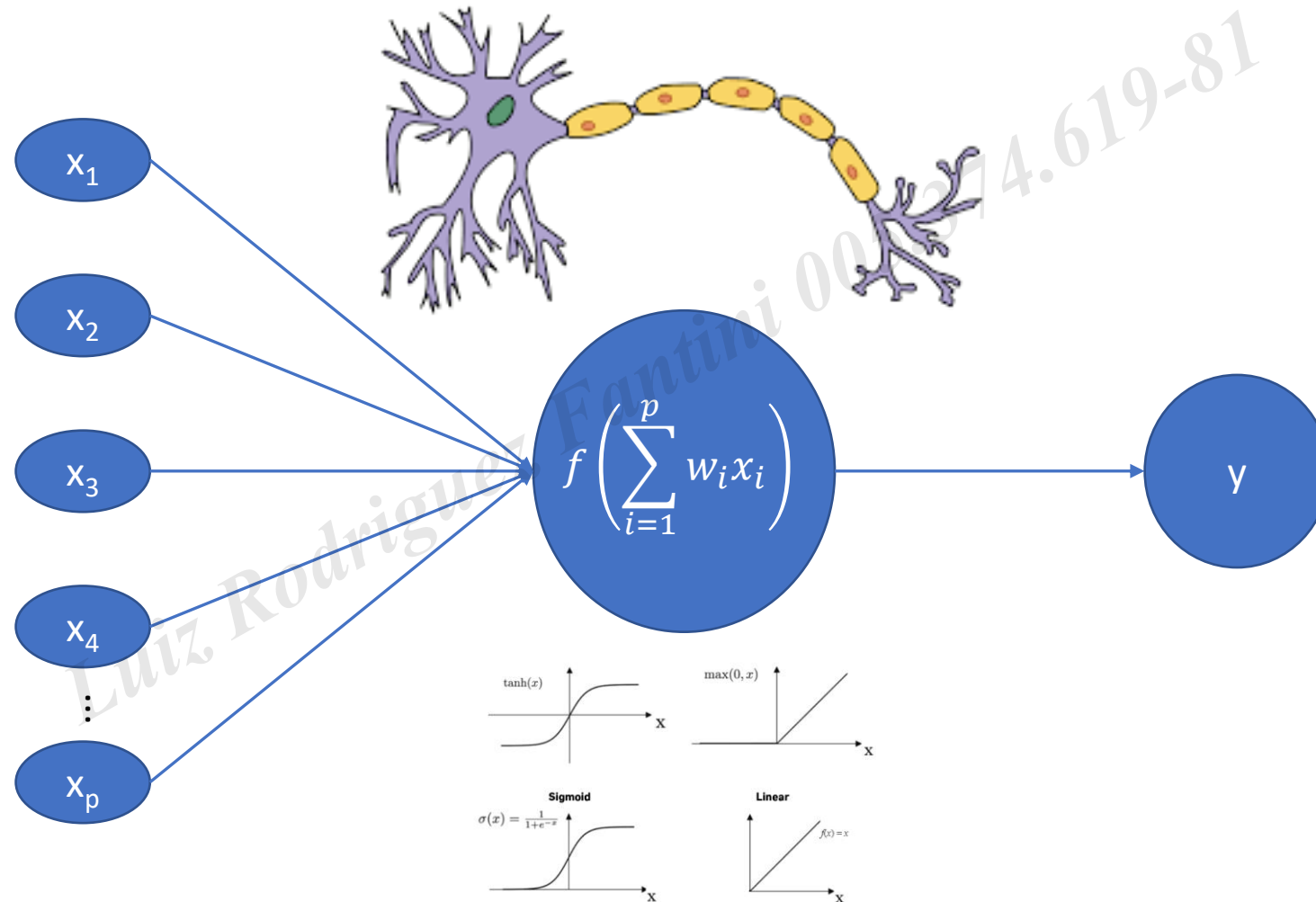
Sigmoid



Linear



Perceptron



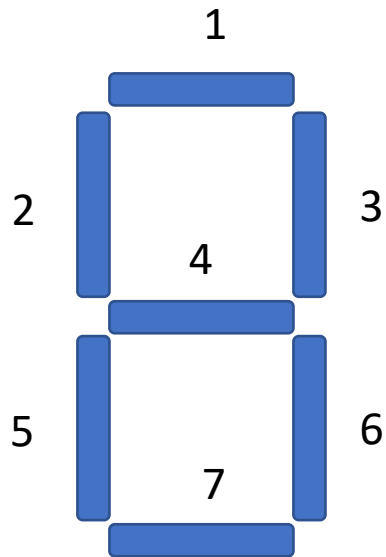
OCR – Optical Character Recognition



Vamos pensar em uma versão bem simples do problema. Dígitos de um rádio relógio antigo possuem uma estrutura bem simples.

Luiz Rodriguez. Fone: 005.374.619-81

OCR – Optical Character Recognition



Há 7 regiões básicas, que podem estar ativas ou inativas, e definem um dígito.

Por exemplo, se somente as regiões 1, 3 e 6 estão ativadas, temos o número 7.

Luiz Rodriguez Fantini 005.274.619-81

Perceptron de Rosenblatt

- O Perceptron de Rosenblatt (~1950-1960) tem essa ideia, só que com propósito mais geral
- foi construído para fazer OCR (optical character recognition)
- Para isso, mapeia regiões de uma imagem “ativadas” e “não ativadas”
- Cada unidade é um neurônio de McCulloch-Pitt

FIG. 1 — Organization of a biological brain. (Red areas indicate active cells, responding to the letter X.)

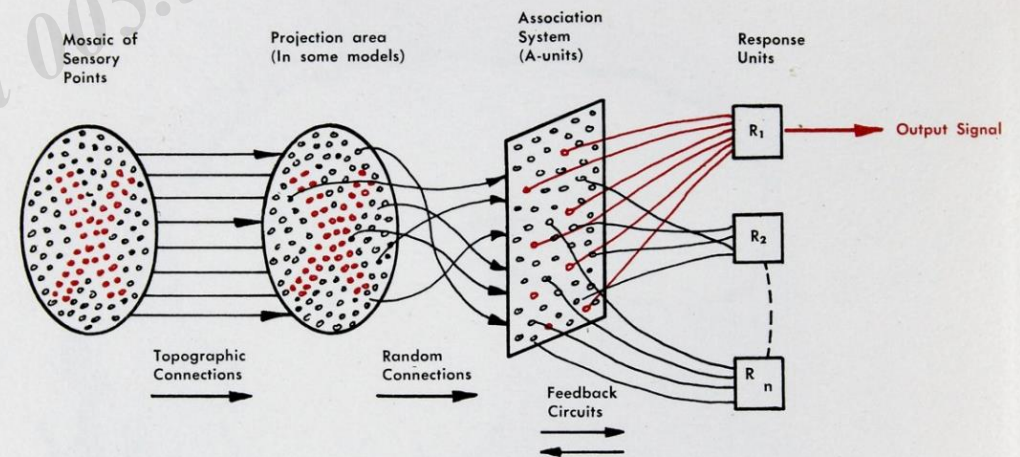
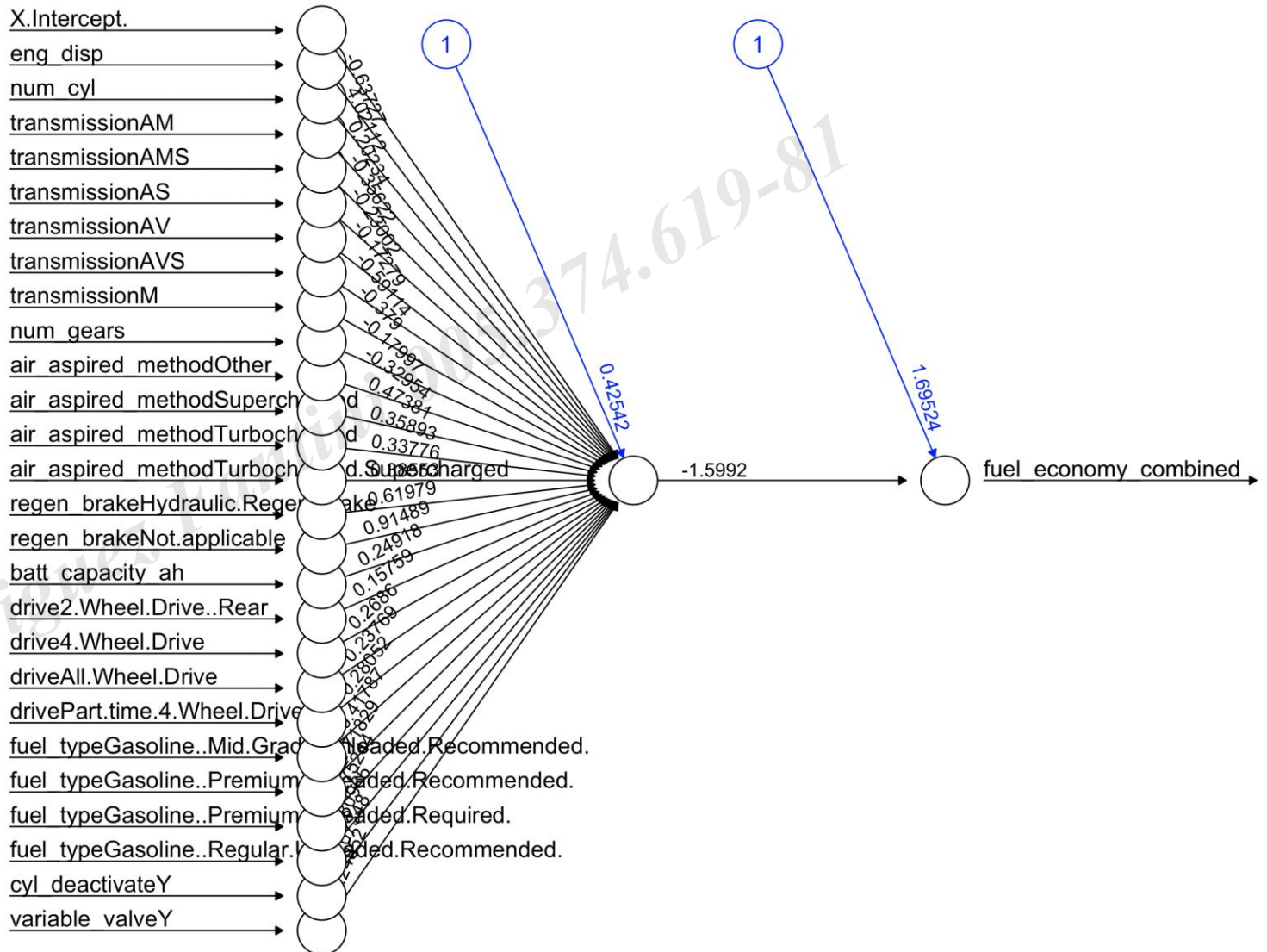


FIG. 2 — Organization of a perceptron.

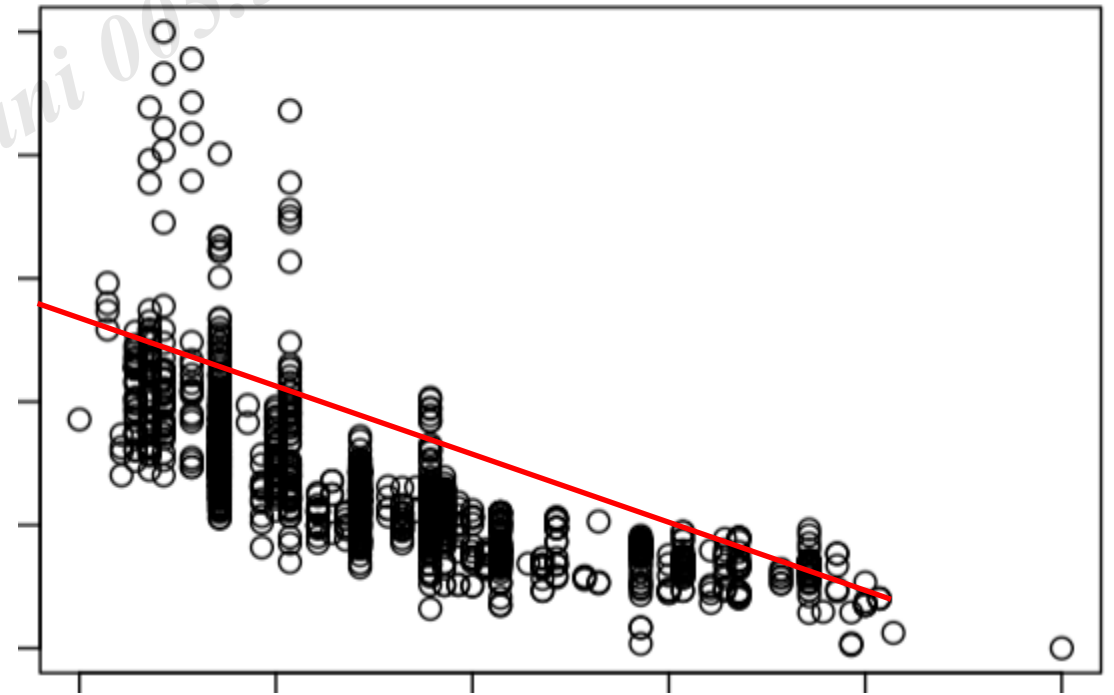
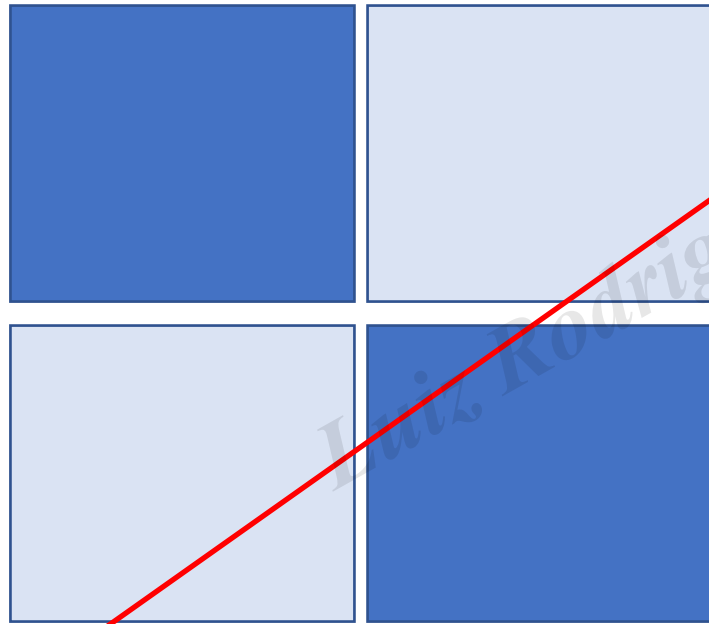
Perceptron Linear

- Possui a mesma estrutura de uma regressão linear com a função de ativação indicada.



Limitações do perceptron linear

- O perceptron linear só captura padrões lineares

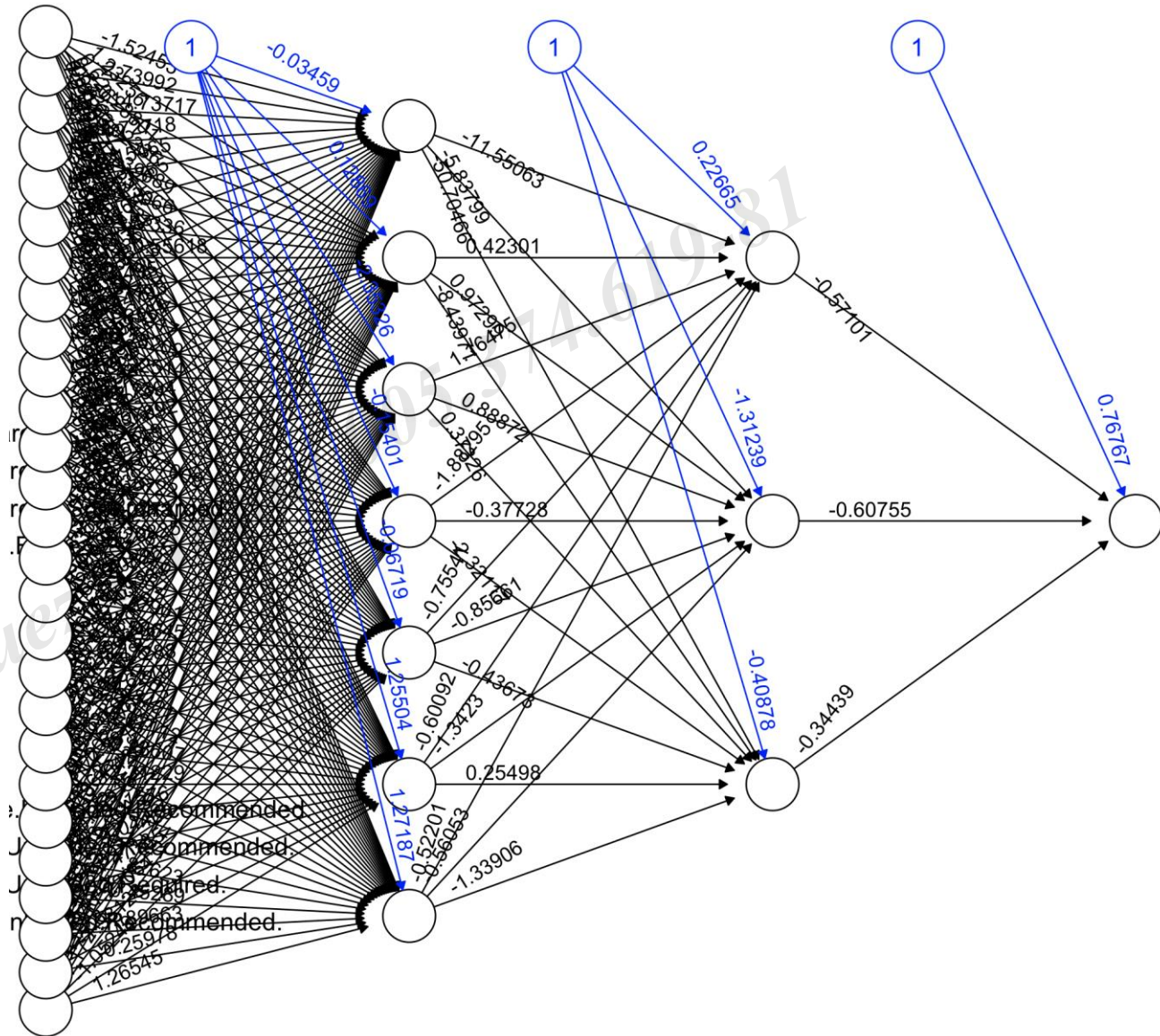


Luiz Rodriguez Fantini 005.374.619-81



Perceptron multi-camada

- Possui camadas “escondidas” intermediárias
- Captura padrões não lineares
- Pode se aproveitar do processamento paralelo de GPUs
- Não é “interpretável” como a regressão



Luiz Rodriguez Fantini 005.374.619-81



Funções de perda

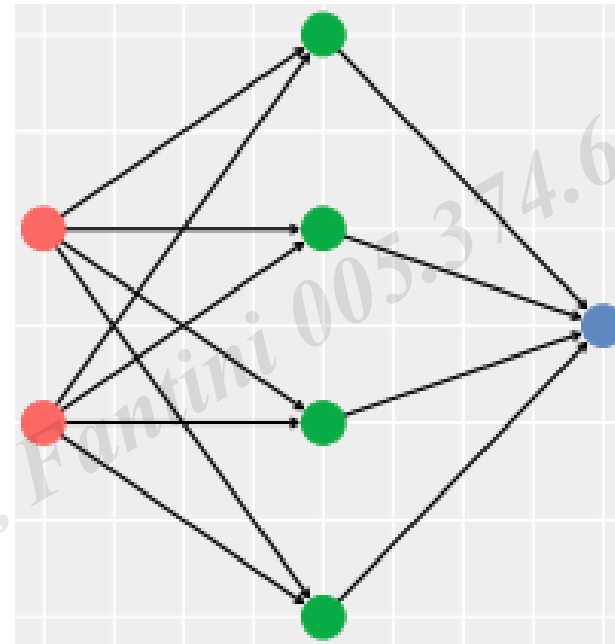
Variáveis Contínuas
SQE

$$SQE = \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

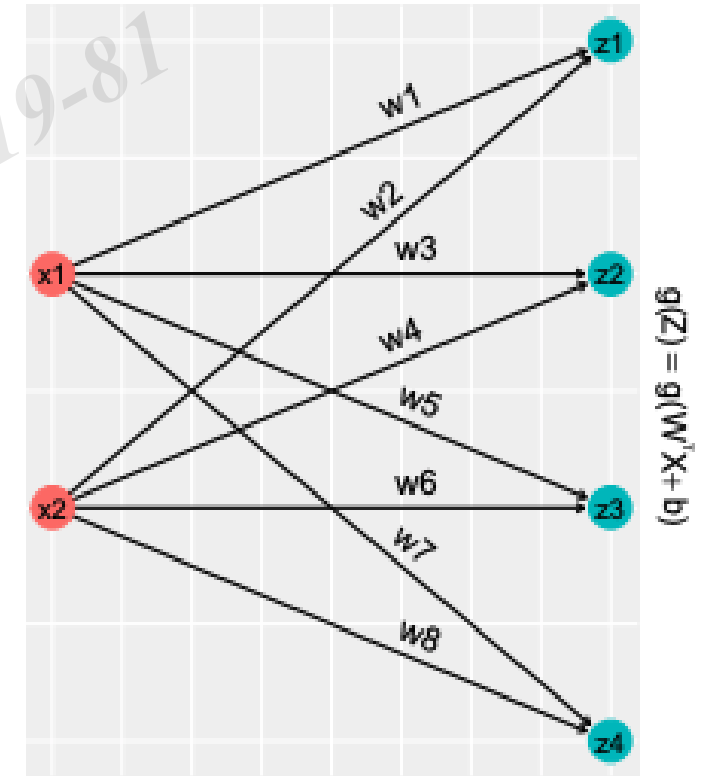
Variáveis binárias
Cross-Entropy

$$L = y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)$$

Redes Neurais Artificiais



layer ● Input ● Hidden ● Output



Deep learning with R - Abhijit Ghatak, ed. Springer, 2019

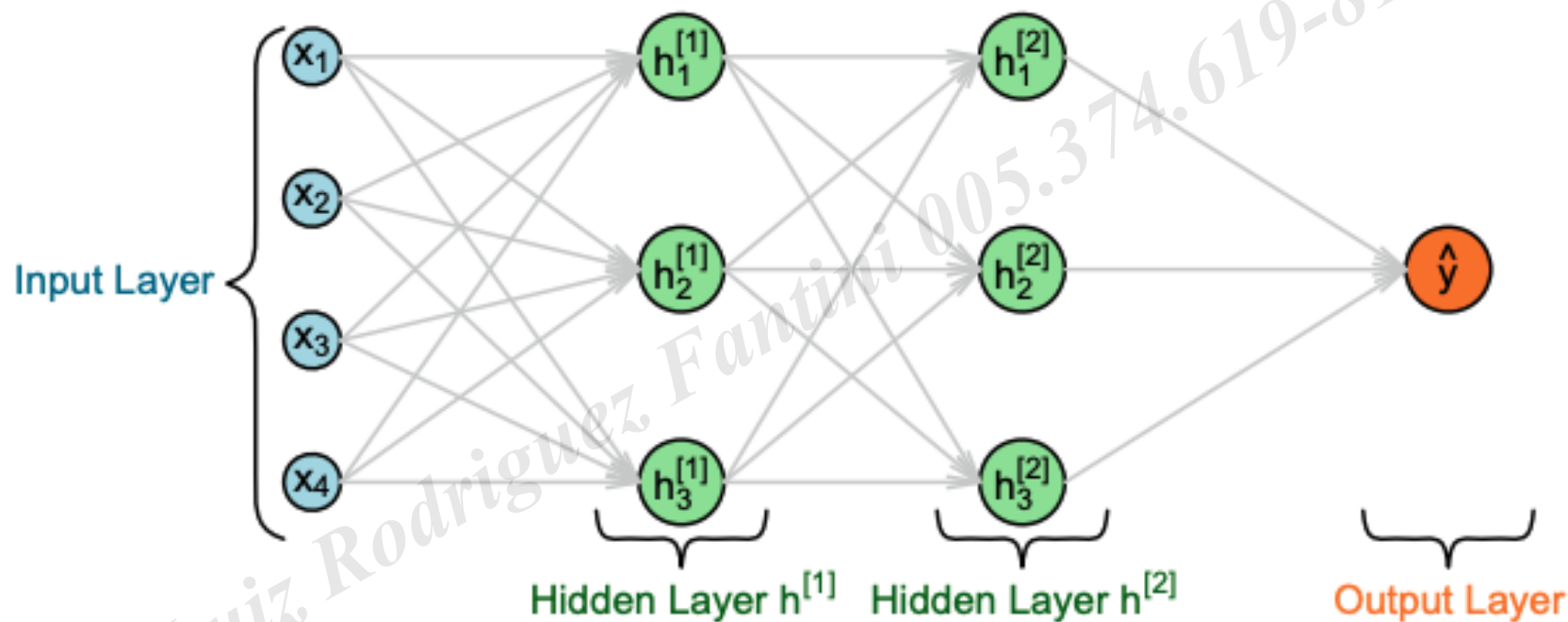
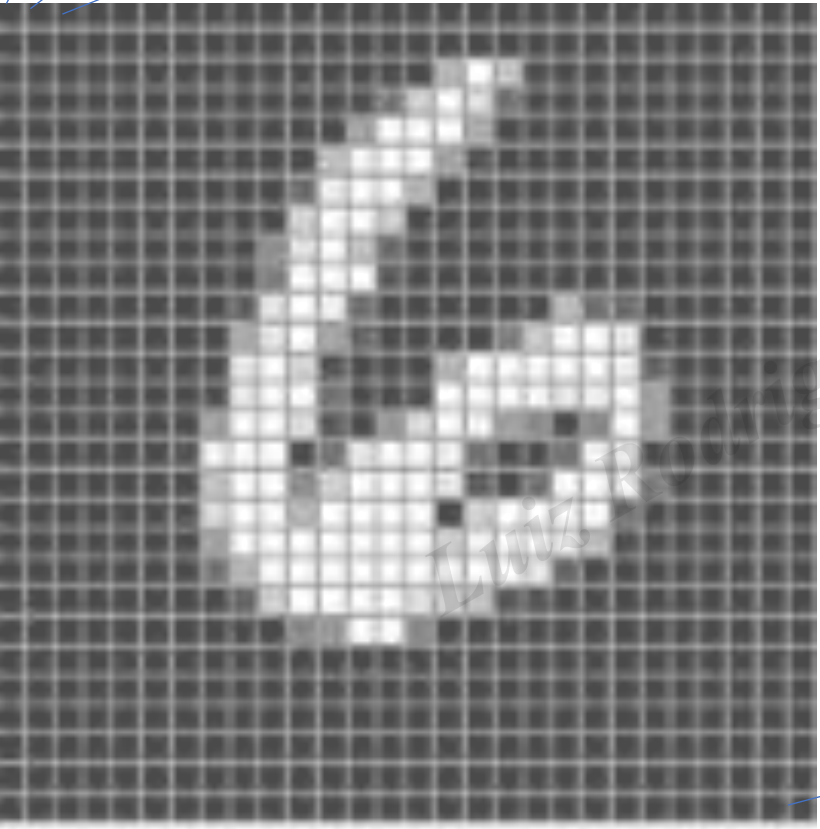


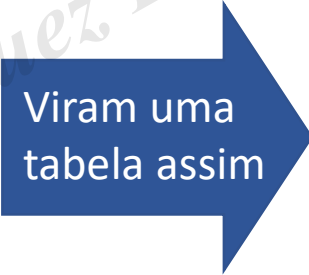
Fig. 2.3 A representation of a neural network with four input features, two hidden layers with three nodes each, and an output layer

Tratamento inicial dos dados

Pixel 1
Pixel 2
Pixel 3



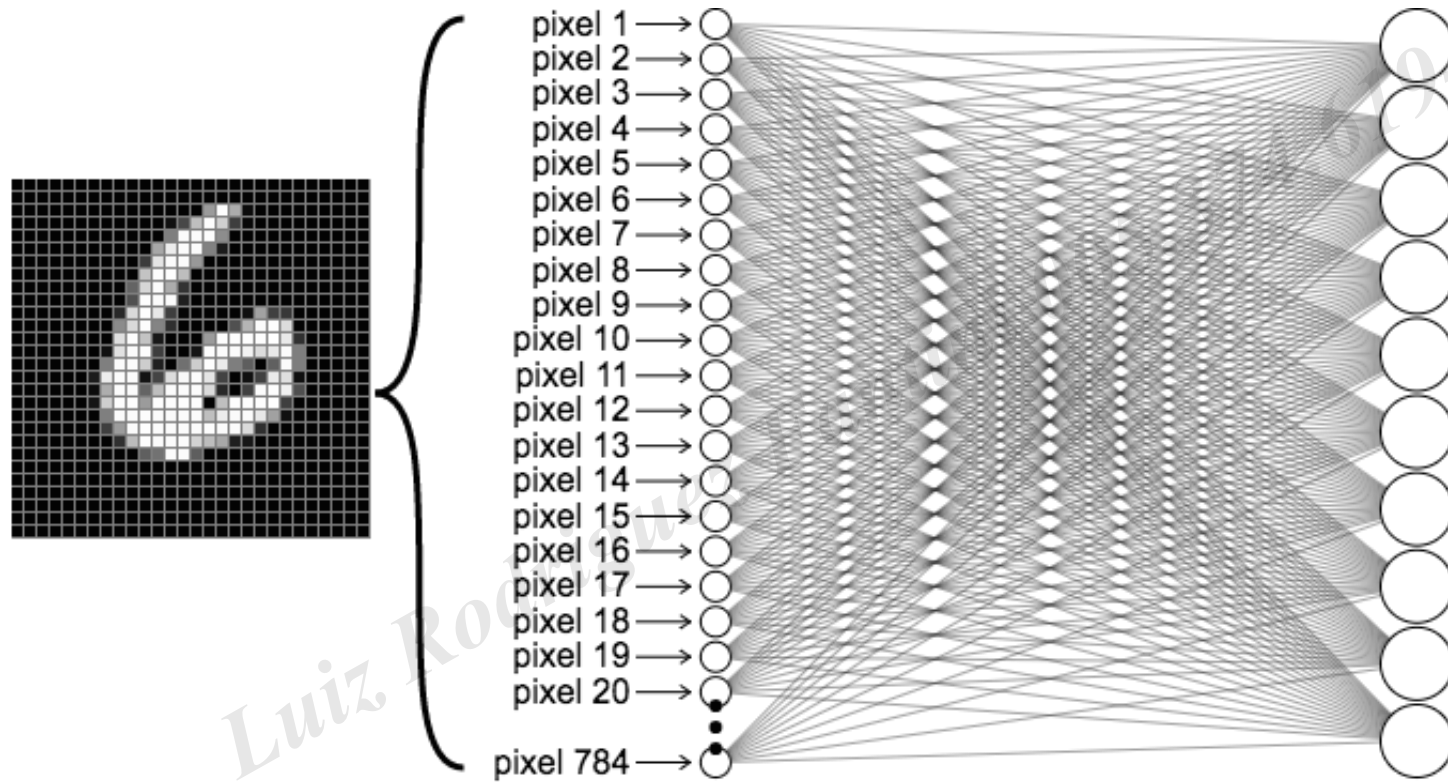
60.000 dessas



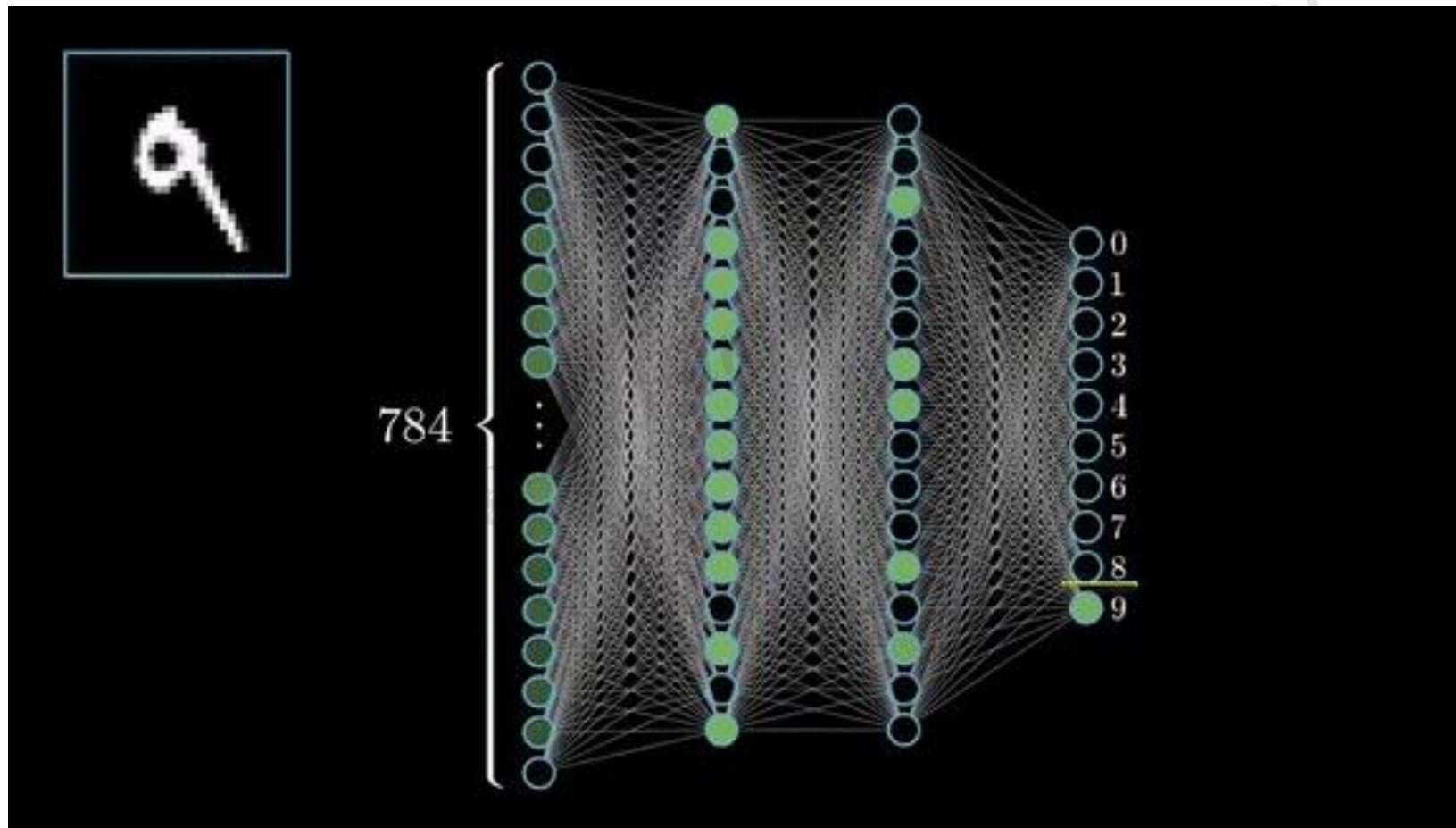
Pixel 784

	Pixel 1	Pixel 2	Pixel 3	⋮	Pixel 784	Rótulo
Imagem 1	0	0	0...		0	4
Imagem 2	0	0	0...		0	3
Imagem 3	0	0	0...		0	9
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Imagem 60.000	0	0	0...		0	5

Rede Neural MNIST



Com apenas uma camada já temos $784 \times 10 = 7.840$ parâmetros!



3blue1brown - <https://www.youtube.com/watch?v=aircAruvnKk>

Gradiente descendente

É o algoritmo mais popular para se treinar redes neurais artificiais por apresentar algumas características:

- Pode alterar as estimativas com pequenos subconjuntos de pontos a cada iteração (no limite 1 único ponto)
- Não depende de inverter matriz
- Funciona com uma base de dados bem grande
- Pode ser processado em paralelo com GPU
- Permite interromper o algoritmo a certo ponto e continuar mais tarde ou em outro problema semelhante (*transfer learning*)

Redes Gradiente Descendente

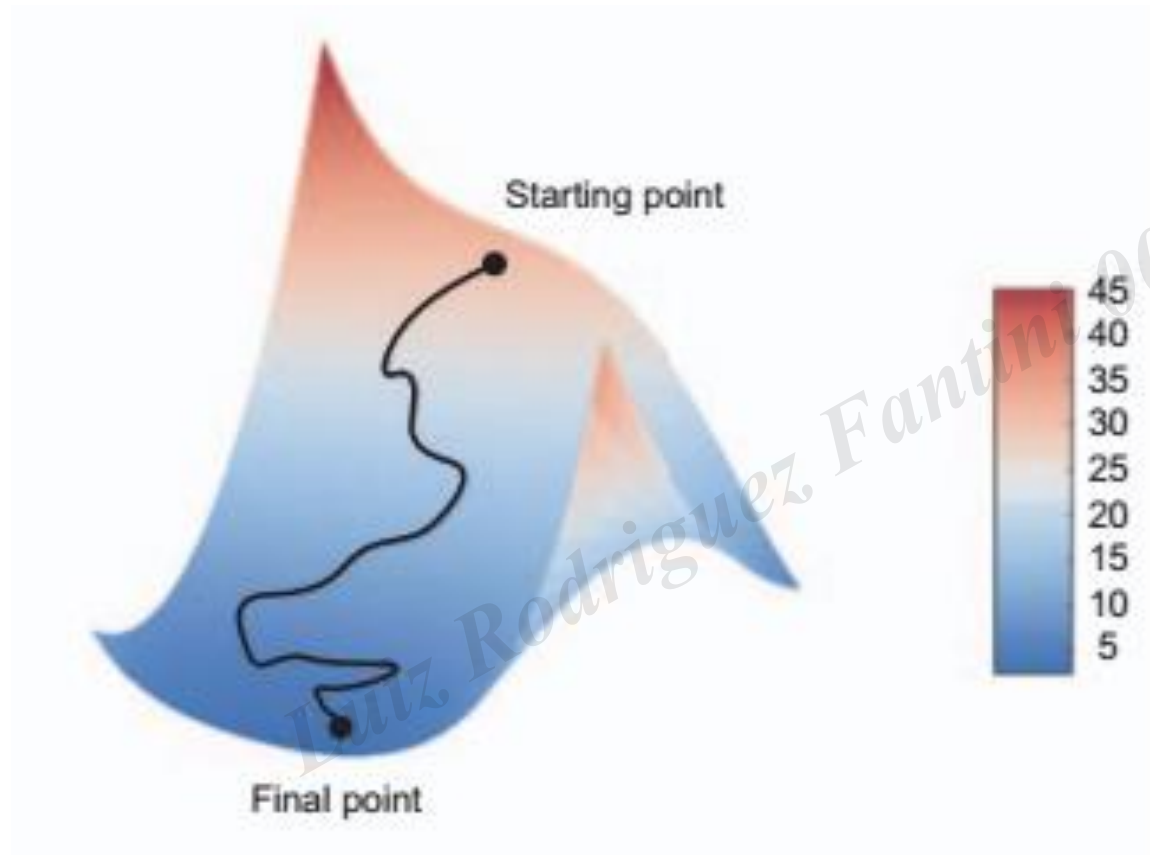


Figure 2.12 Gradient descent down a 2D loss surface (two learnable parameters)

Deep learning with python – François Chollet

Gradiente descendente

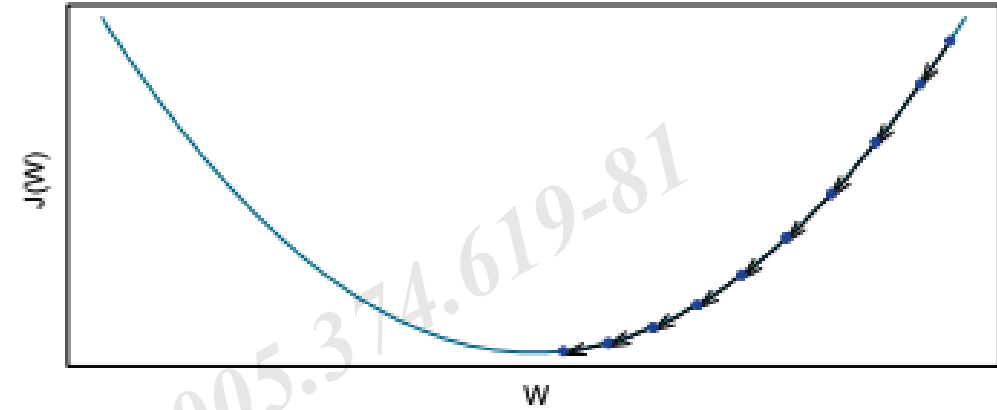


Fig. 1.4 Gradient descent: Rolling down to the minima by updating the weights by the gradient of the loss function

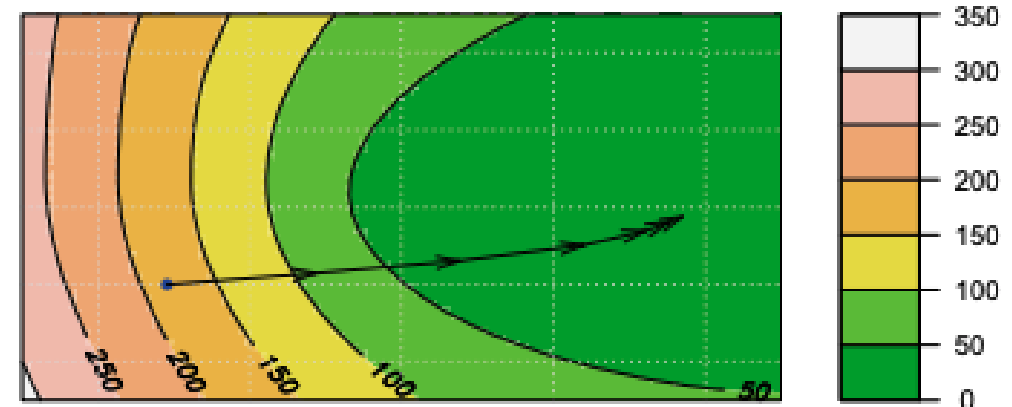


Fig. 1.5 A contour plot showing the cost contours of a sigmoid activation neural network and the cost minimization steps using the gradient descent optimization function

Previsão de consumo de veículo

- Tamanho do motor
- Combustível
- Número de cilindros
- Marca
- Potência
- Tração



Luiz Rodriguez Fantini 005.374.619-81





game

A contribuição da indústria do Game nas Redes Neurais

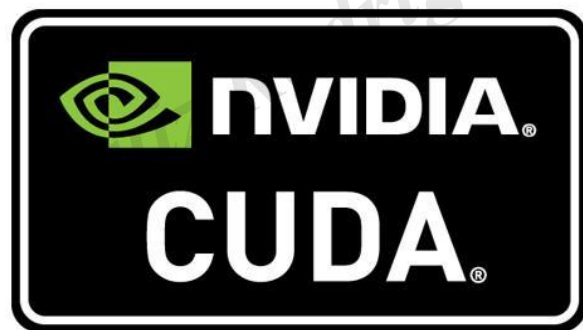
Processadores

- Distância entre transístores: 14 nm
- Fio de cabelo humano: 80.000 nm
- Diâmetro do átomo de ouro: 0,3 nm

A close-up, macro photograph of a GPU (Graphics Processing Unit) chip mounted on a printed circuit board (PCB). The chip is a dark, square integrated circuit with numerous gold-plated pins along its edges. The surrounding PCB is populated with various electronic components, including capacitors, resistors, and other smaller chips, all interconnected by a complex network of copper traces. The lighting is soft, highlighting the metallic surfaces and the intricate details of the microelectronics.

GPU

Processamento com GPU



AMD
ROCm

A close-up, artistic photograph of a circuit board. The board is dark blue with intricate, winding gold-colored traces. Numerous small, glowing blue circular components are scattered across the board, particularly concentrated in the lower-left and upper-right areas. The lighting is dramatic, highlighting the metallic textures and the vibrant blue glow.

TPU

Regularização L2

Variáveis Contínuas
SQE

$$SQE = \sum_{i=1}^N (y_i - \hat{y}_i)^2 + \lambda \sum \beta_i^2$$

Variáveis binárias
Cross-Entropy

$$L = \sum y_i \log(\hat{y}_i) + \lambda \sum \beta_i^2$$



Reconhecimento
de atividade
humana com
smartphone

Luiz Rodriguez Fantini 005.374.619-81



Conclusões

- Redes Neurais são a introdução ao Deep Learning (que é um ramo muito promissor)
- São poderosas e flexíveis
- Requerem poder computacional especial (GPU / TPU)
- São famosas em dados menos estruturados (ex: imagens, áudio)





“O mundo está em constante mudança, por isso não se apegue a nada.”

“Quando você evolui, o mundo evolui”

Sidarta Gautama



Por hoje é só ;)



[linkedin.com/in/joao-serrajordia](https://www.linkedin.com/in/joao-serrajordia)