

Data Science e Machine Learning aplicado a Economia

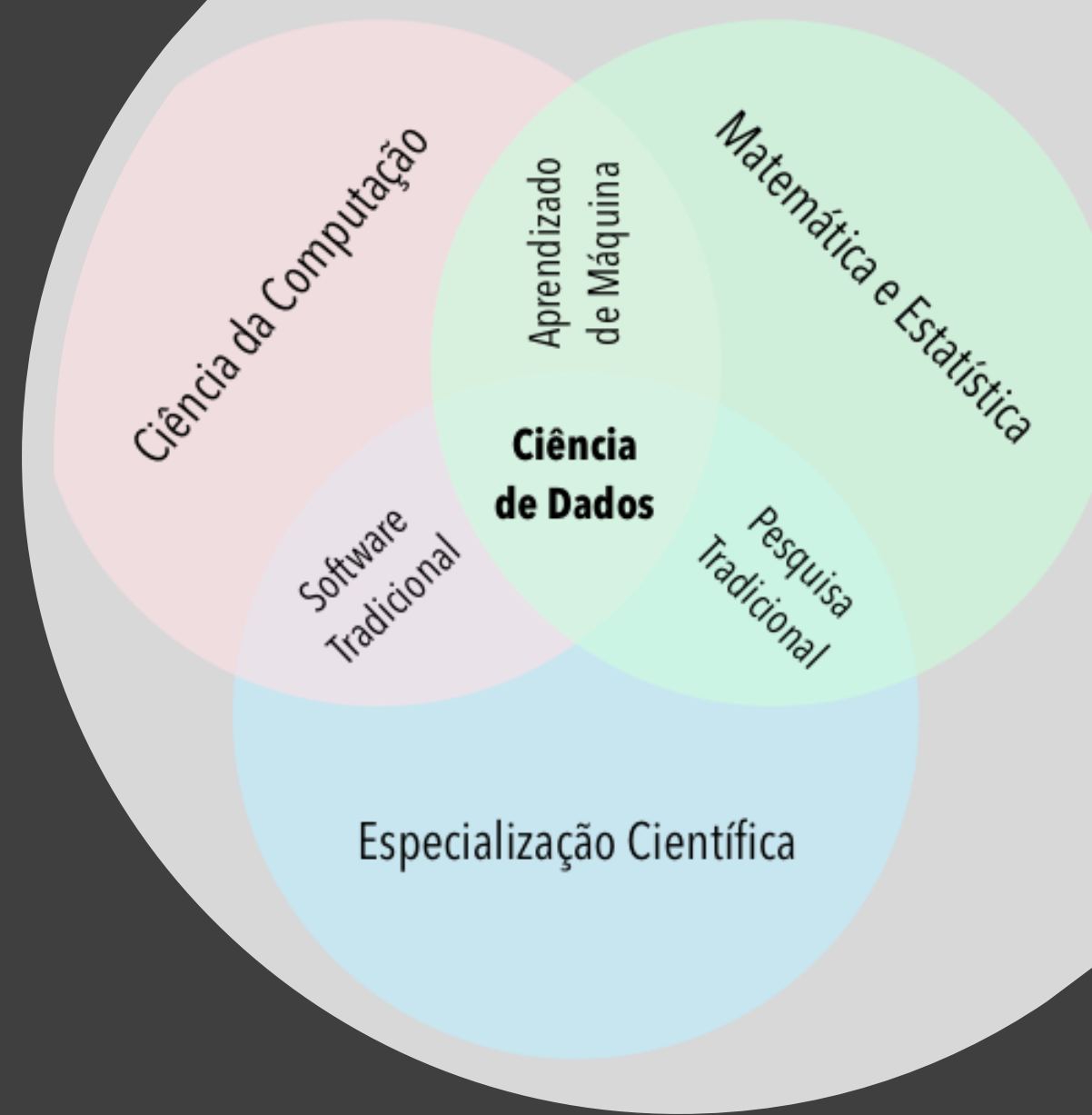
Vitor Bianchi Lanzetta – Consultor Cientista de Dados na 

Sumário

- I. Introdução a Ciência de Dados
- II. Redes Neurais Artificiais
- III. Redes Neurais usando R
- IV. Machine Learning vs Econometria

O que é Ciência de Dados

Campo multidisciplinar responsável por usar métodos científicos, processos e algoritmos para extrair conhecimento e insights de dados.



O que é Machine Learning

- Alan Turing
- Paradigma da Programação: Regra + Input > Resultado
- Machine Learning: Input + Resultado > Regra

Métodos

- **Regressão linear**, logística, naive bayes
- Clustering, análise de componentes principais (PCA)
- **Redes neurais** (Deep Learning), florestas aleatórias
- GBM (gradiente boosting machines)

Buzz Words

- Big Data, Deep Learning
- Supervised, Unsupervised Learning
- Feature Engineering

Principais aplicações

- Recomendadores (Amazon)
- Detecção de Spam (Gmail)
- Reconhecimento Facial (Facebook)



Principais Preocupações

Aplicações em larga escala, erros em larga escala

abc 15
ARIZONA

SELF-DRIVING UBER KILLS WOMAN

abc 15
ARIZONA

Ferramentas

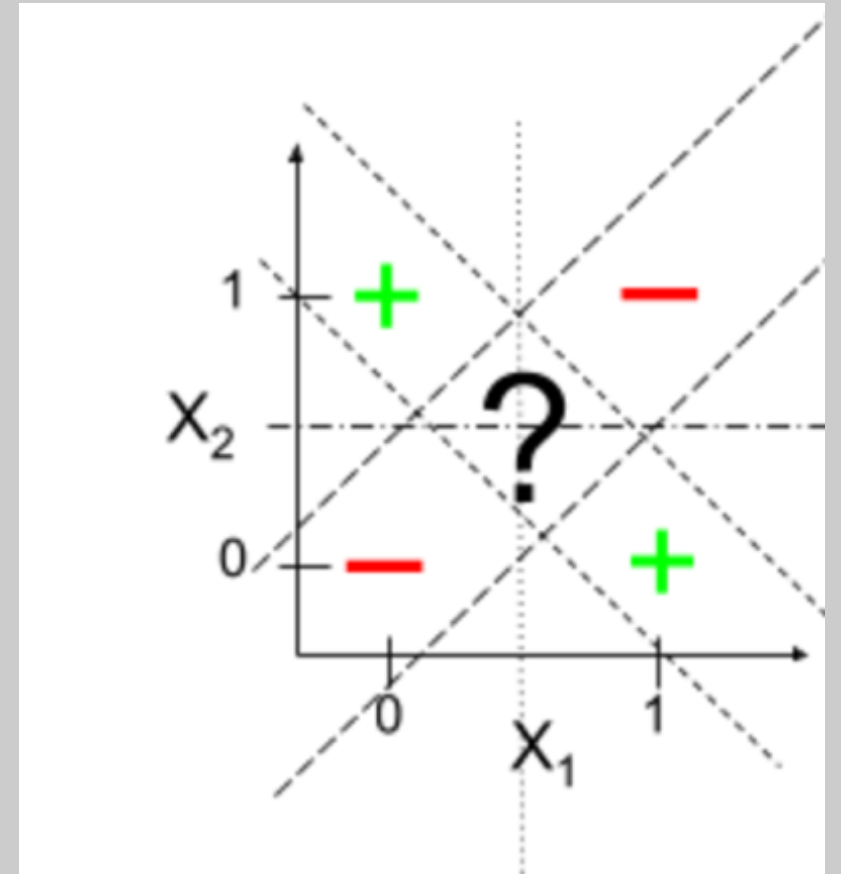
- R
- Rstudio
- Python
- Spark
- Haadop



Métodos: lineares e não lineares

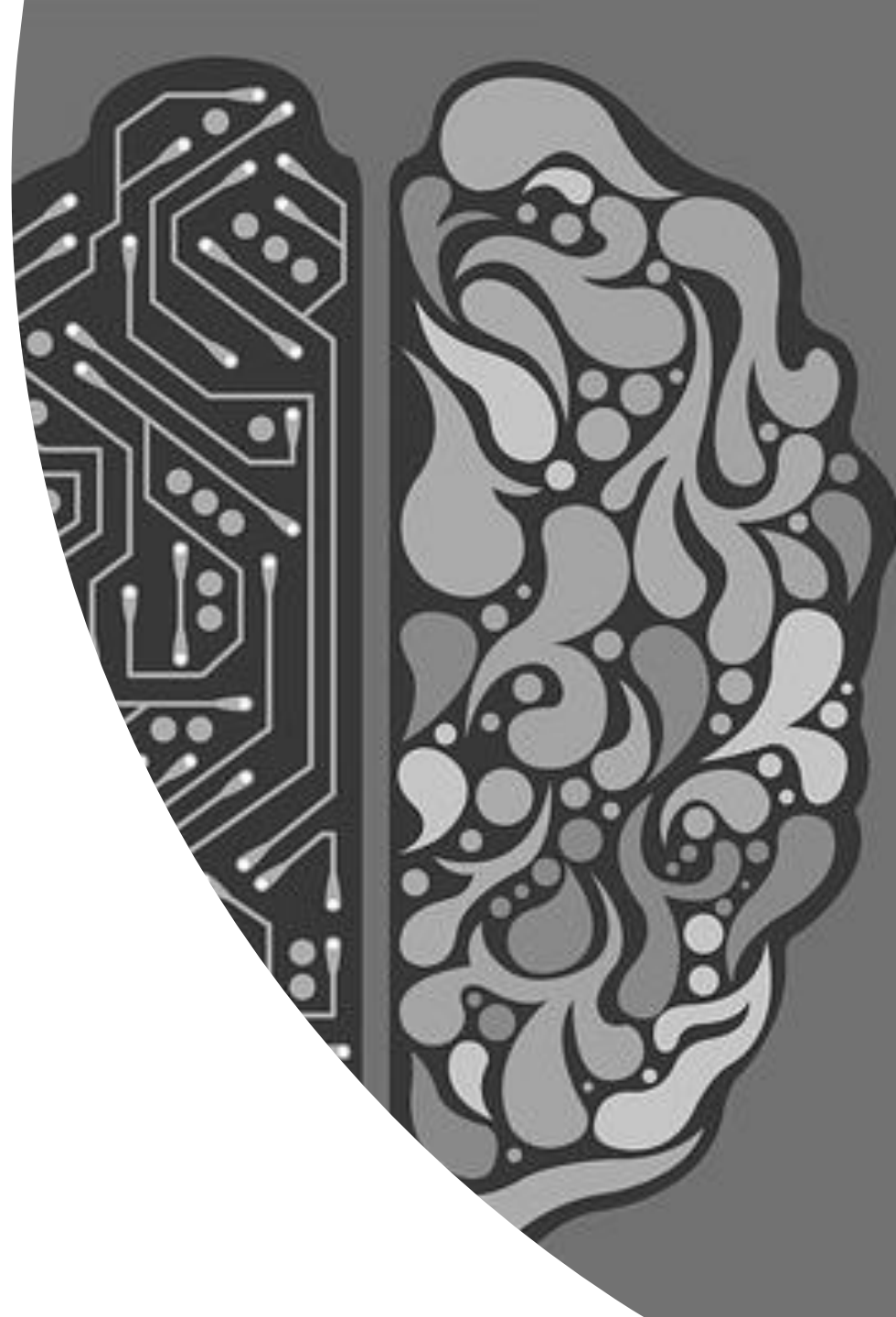
Problema do XOR:

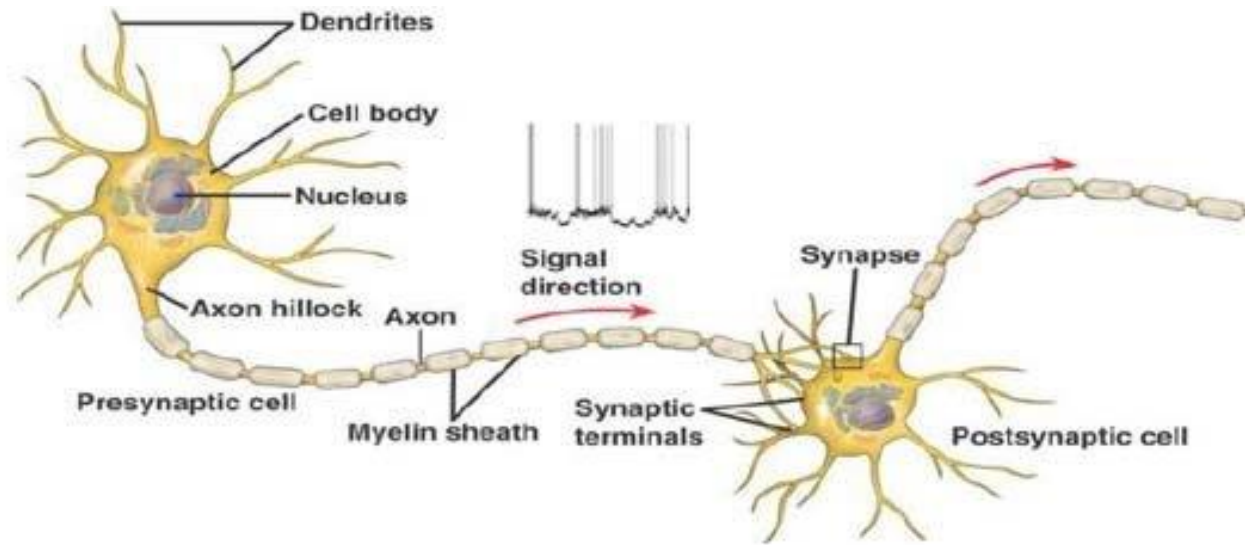
- Simples, 2 variáveis
- Não-linear



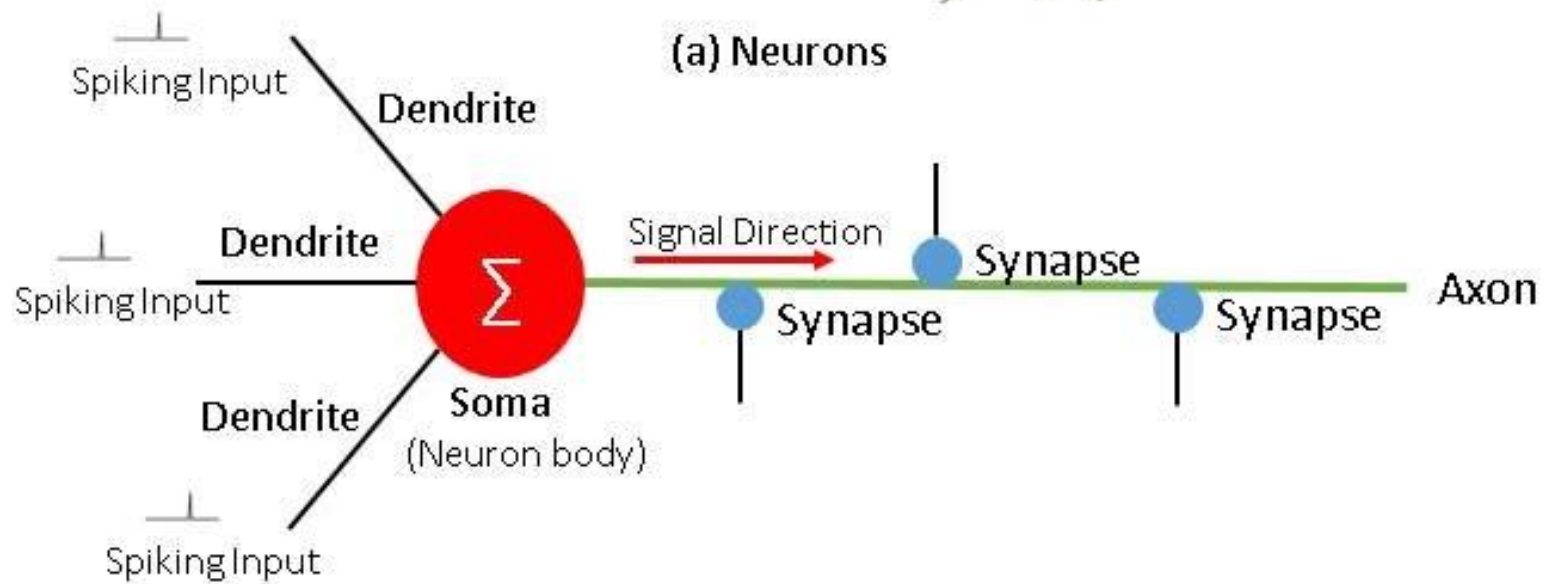
Redes Neurais Artificiais (RNN)

- Aproximadores universais funções
- Inspiração Biológica
- Caixa preta





(a) Neurons



(b) Neuron Structure

História das RNN



Conceito em 1943 – McCulloch e Pitts

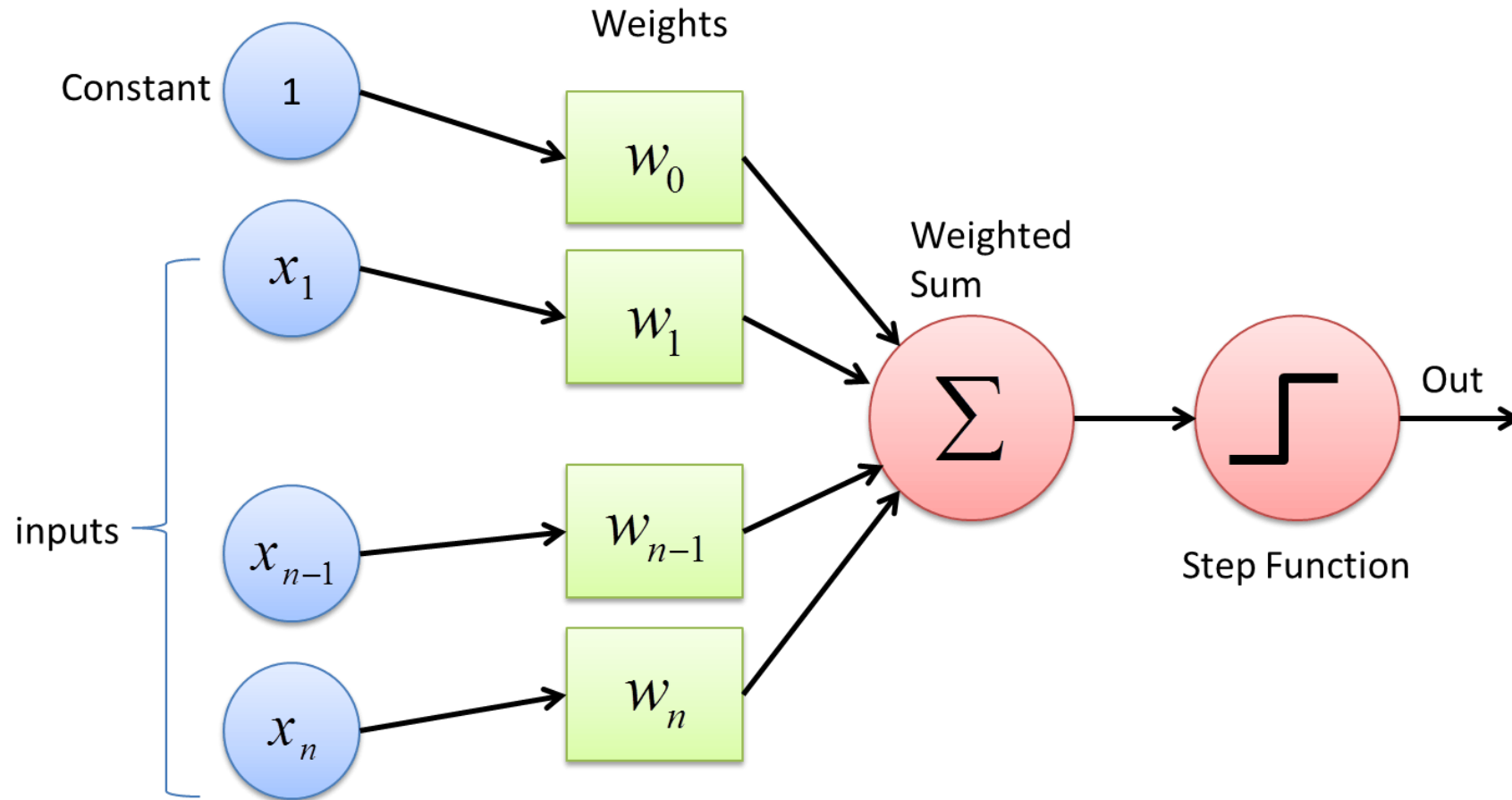


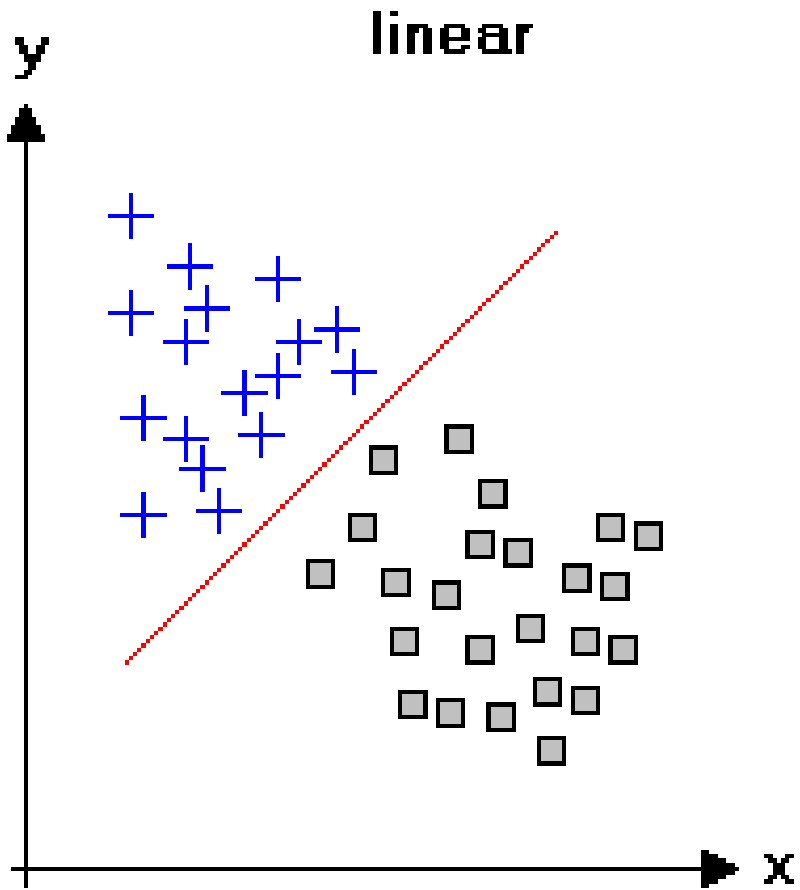
Primeira RNN em 1951 – Marvin Minsky (Princeton)



Livro Perceptron (Minsky & Papert, 1961)

Perceptron





Perceptron:
limitações

História das RNN:

Renascença



Linnainmaa, 1970 –
retropropagação eficiente



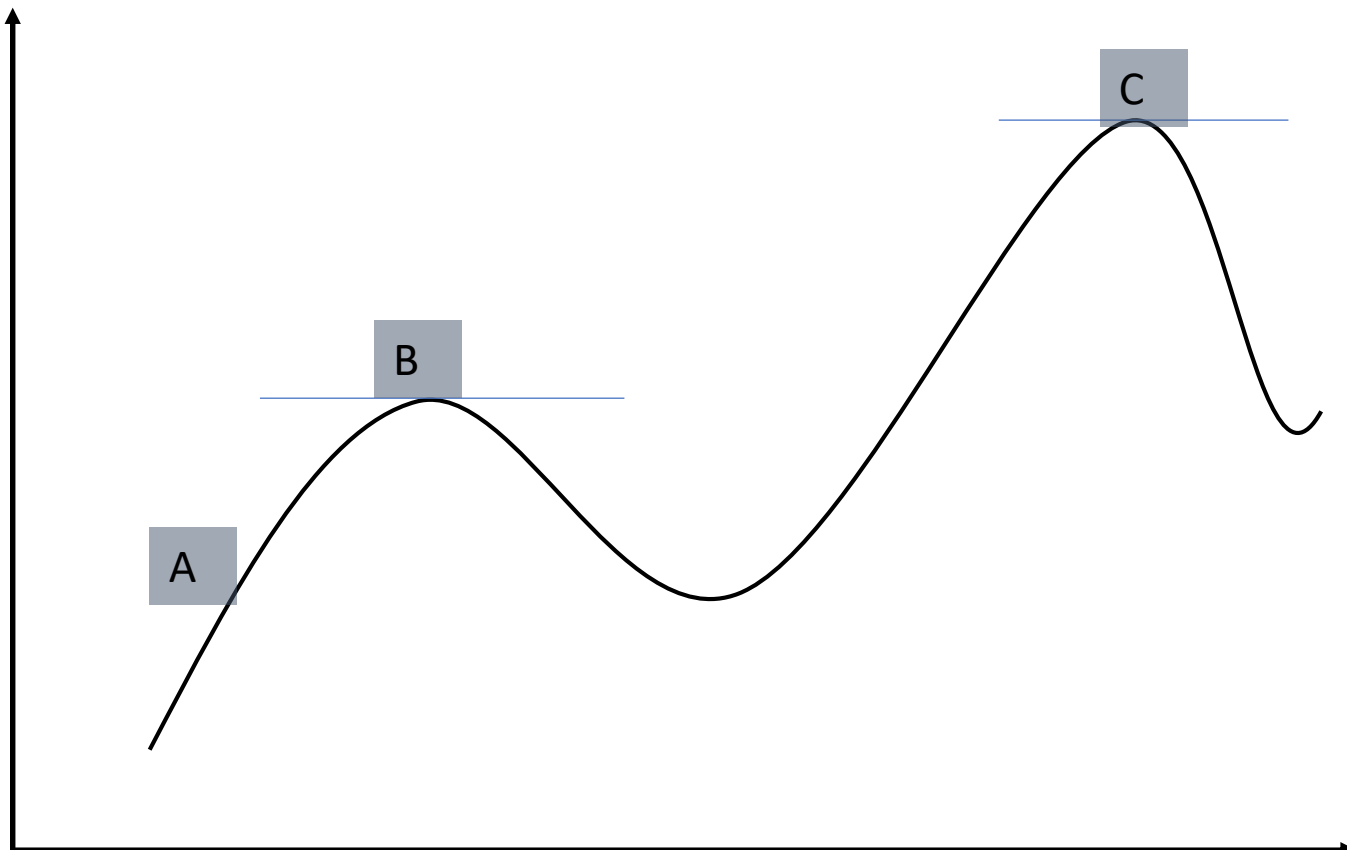
Parker, 1985 & LeCun, 1985
redescoberta da
retropropagação

Retropropagação – Como funciona

- Otimização (escolha dos parâmetros)
- Baseado em Gradiente
- Serve de base para os principais algos usados no treinamento
- Não convexo (pré-processamento ajuda)

Gradiente

Salário



Anos de estudo

História das RNN – AlexNet

- Desafio ImageNet
- 1.3 mi imagens em alta resolução
- 1000 classes diferentes

RNN Hoje!

- Popular no Kaggle
- ONGs : DeepMind / Open.Ai
- Google (TensorFlow)

Aplicações Interessantes

- Redes Recorrentes
- Reinforcement Learning
- GANs

Reinforcement Learning – casos:

- DeepMind, StarCraft II
- Open.ai, Dota 2

GANs

- Generative Adversarial Networks
- Redes concorrentes
- Podem ser usadas para criar mais dados
- Também usadas em artes





SHIP

CAR(99.7%)



HORSE

FROG(99.9%)



DEER

AIRPLANE(85.3%)



DEER

DOG(86.4%)



HORSE

DOG(70.7%)



DOG

CAT(75.5%)



BIRD

FROG(86.5%)



BIRD

FROG(88.8%)

Redes Recorrentes

- Interessante propriedade temporal
- LSTM: “Portões” controlam o que a rede lembra ou esquece
- Muito usadas em NLP



Prós:

Elevada Capacidade
Possibilidade de
modelar relações
não lineares
Explora relações



Contras:

Difícil interpretar
(black box)
Overfitting
Ganancioso

RNN – O Modelo

Unidade Básica:

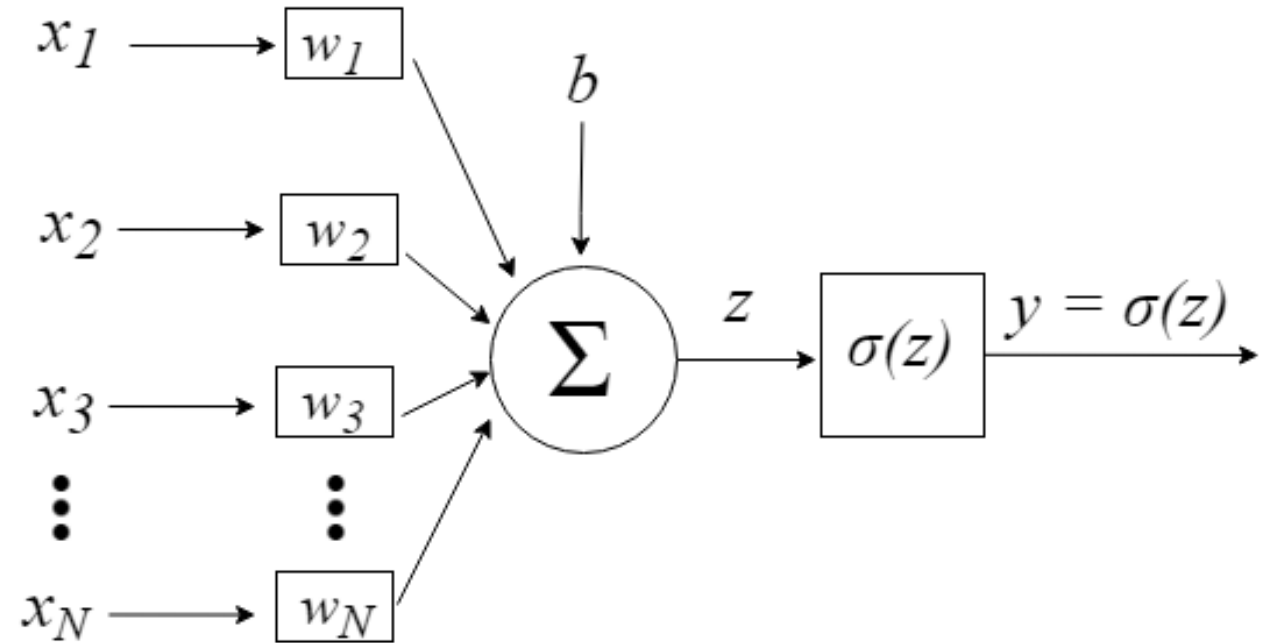
- Neurônio / nódulo
- Aceita diferentes funções de ativação

Nóduo

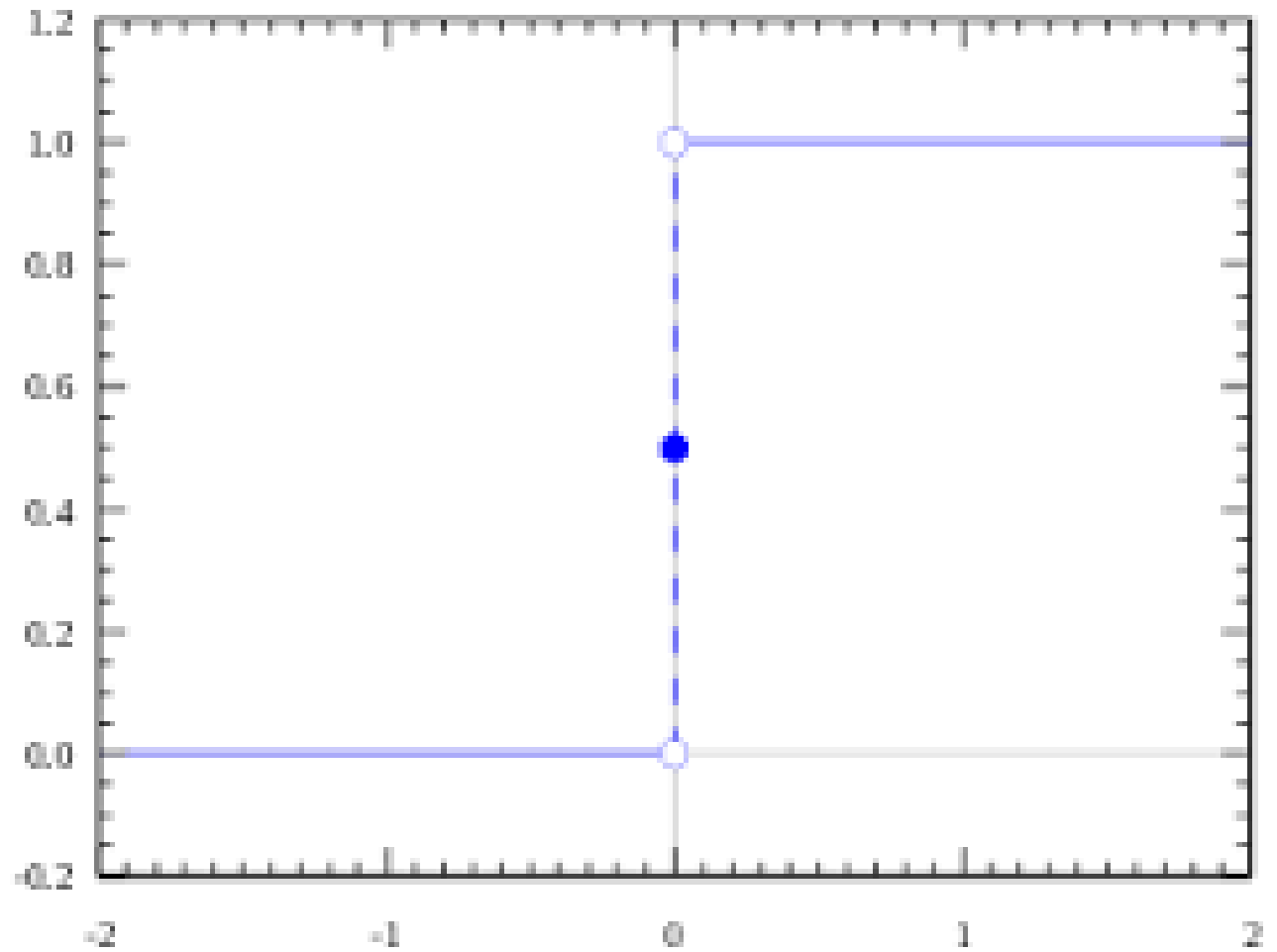
- Entrada:

$$Ai = \sum_{i=0}^n W_i X_i$$

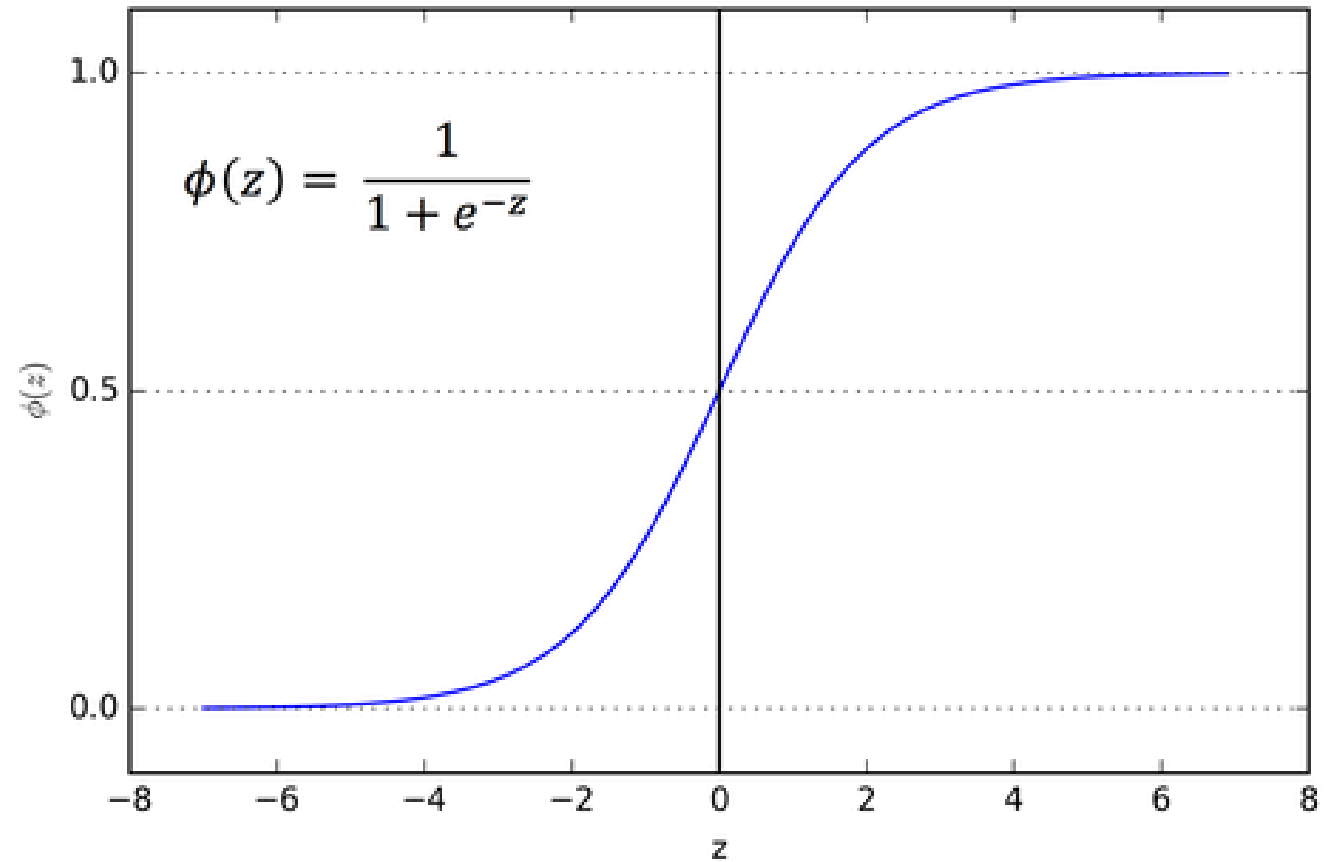
- Saída: $G(Ai)$



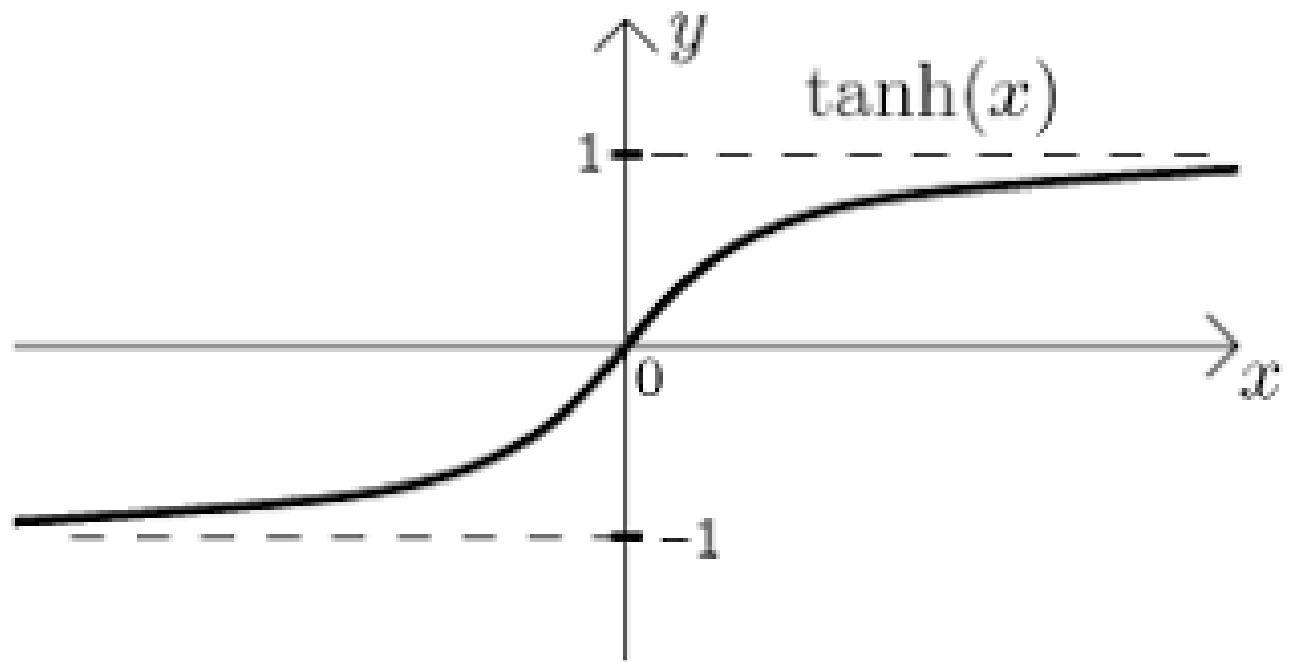
Função de Ativação – Degrau



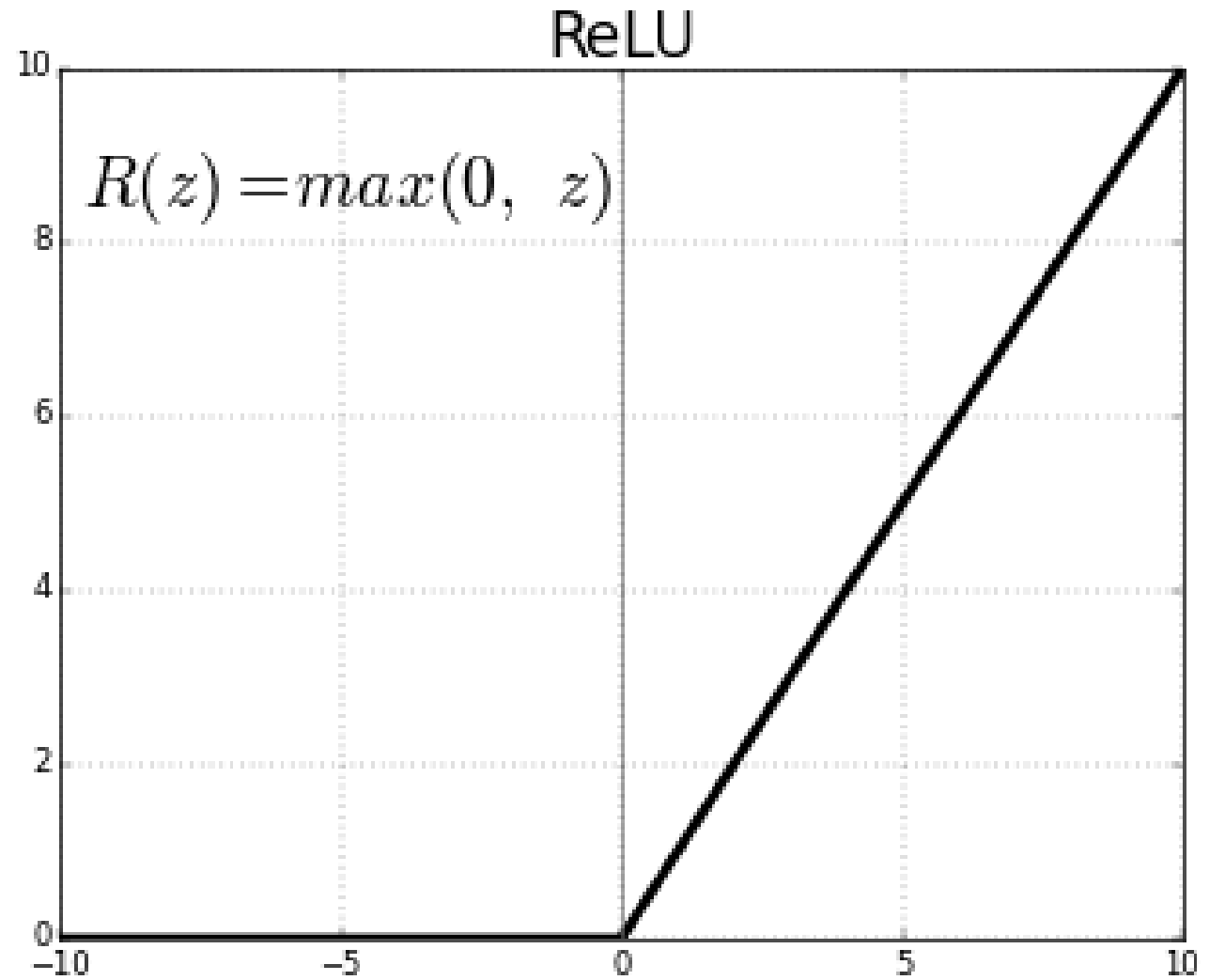
Função de Ativação – Sigmoid



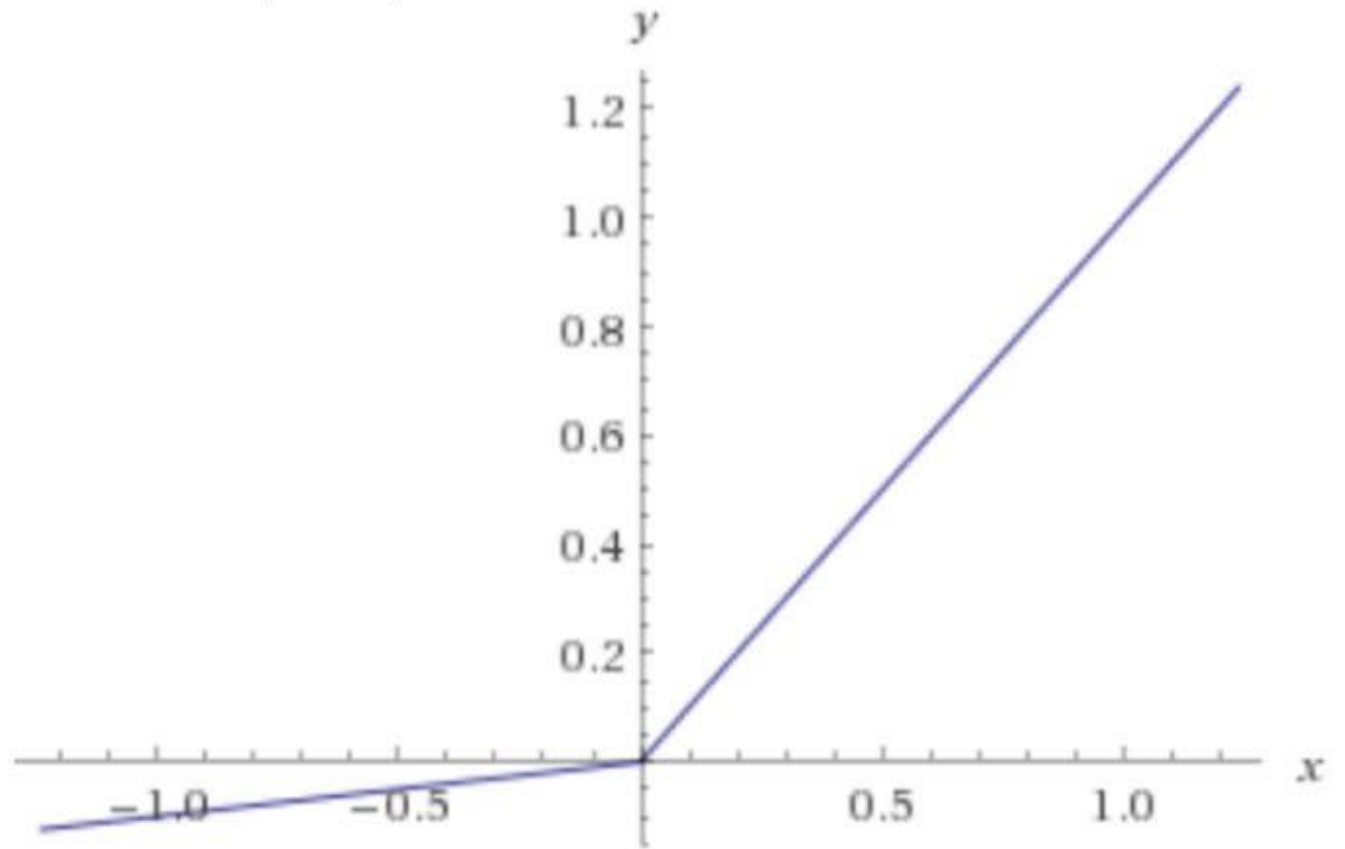
Função de Ativação – Tahn



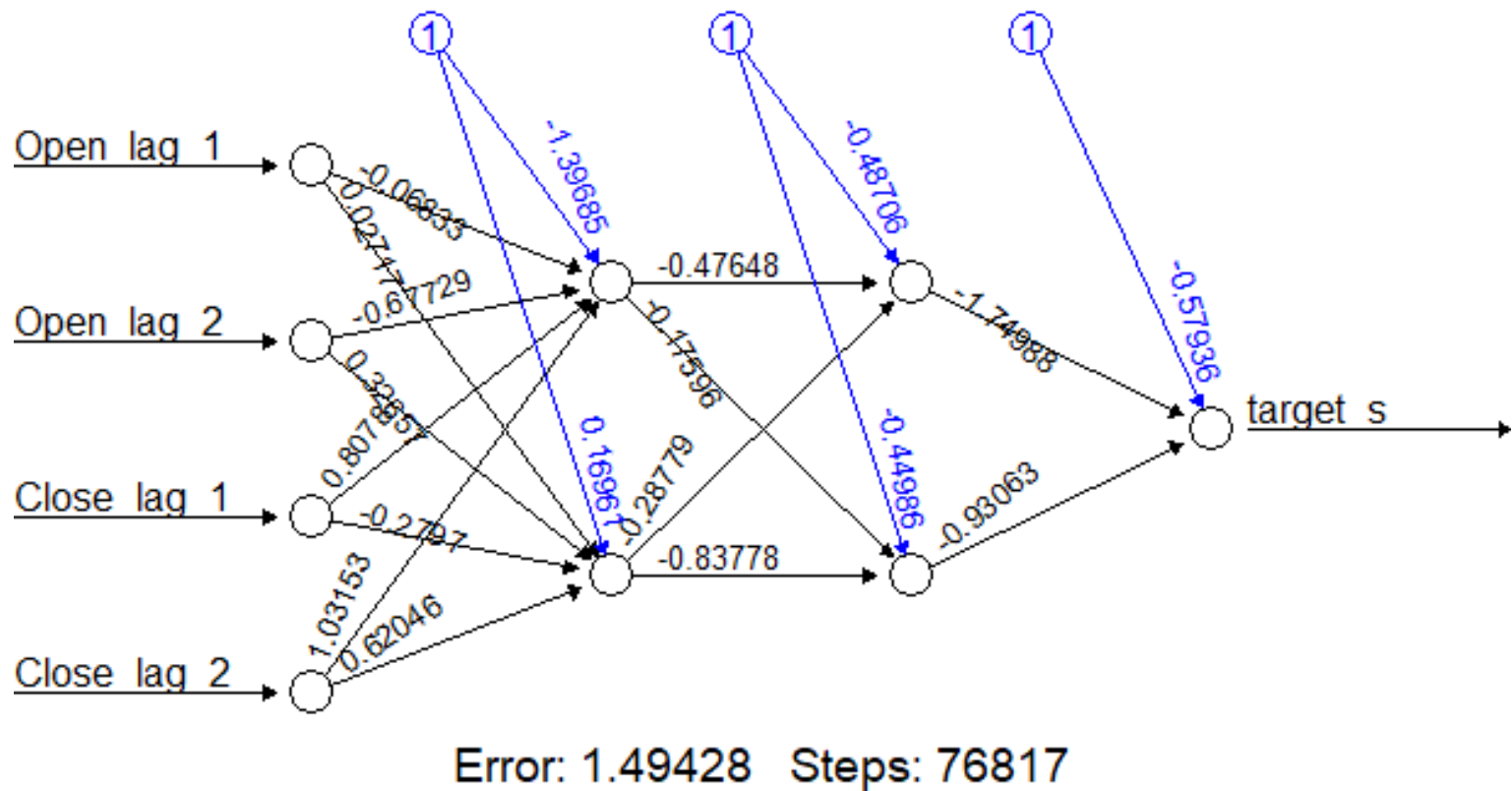
Função de Ativação - ReLU



Função de Ativação – Leaky ReLU



RNN – Nódulos em Conjunto, REDES!



RNN – Camadas

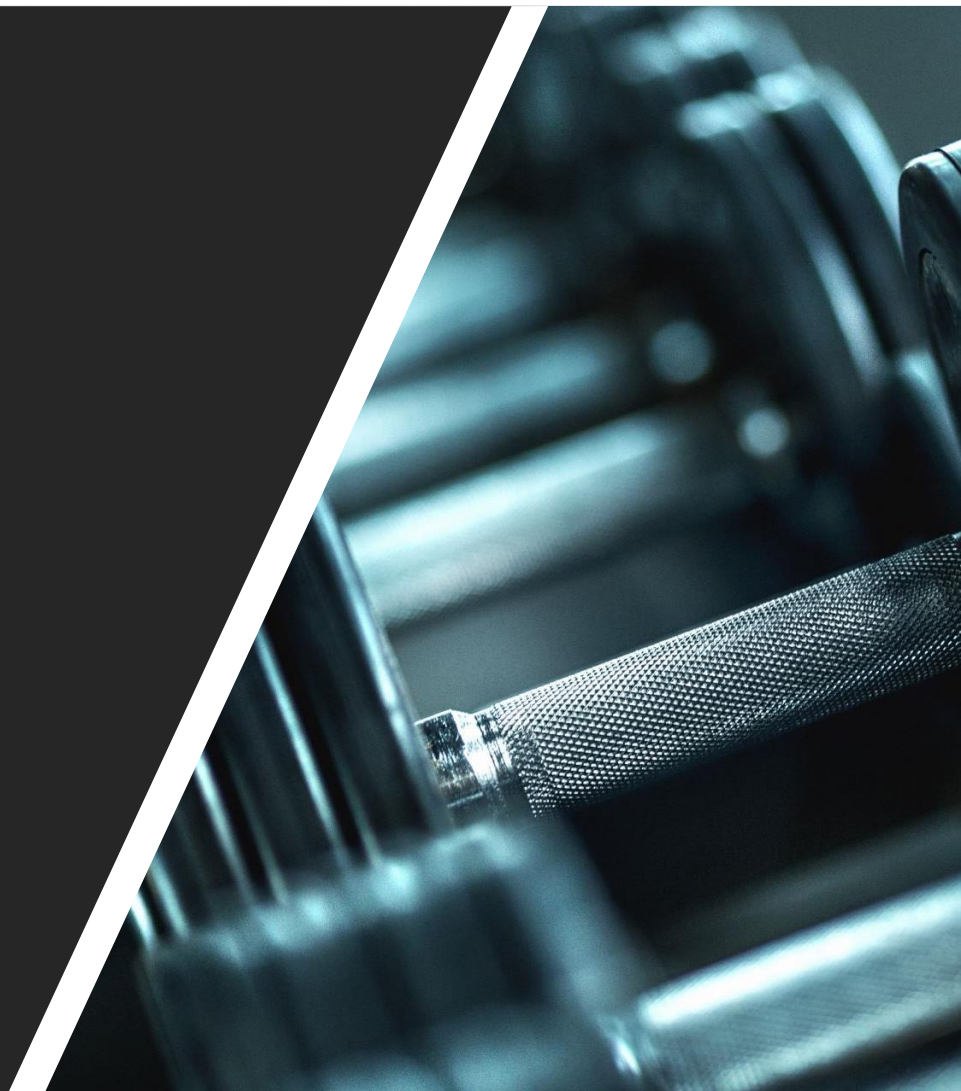
- Input / Entrada
- Hidden / Saída
- Output / Saída

RNN – Pré-treino

- Glutamina + BCAA
- Limpeza dos dados
- Normalização

Treinamento

- Algoritmo de treino
- Fornada
- Épocas



Validação

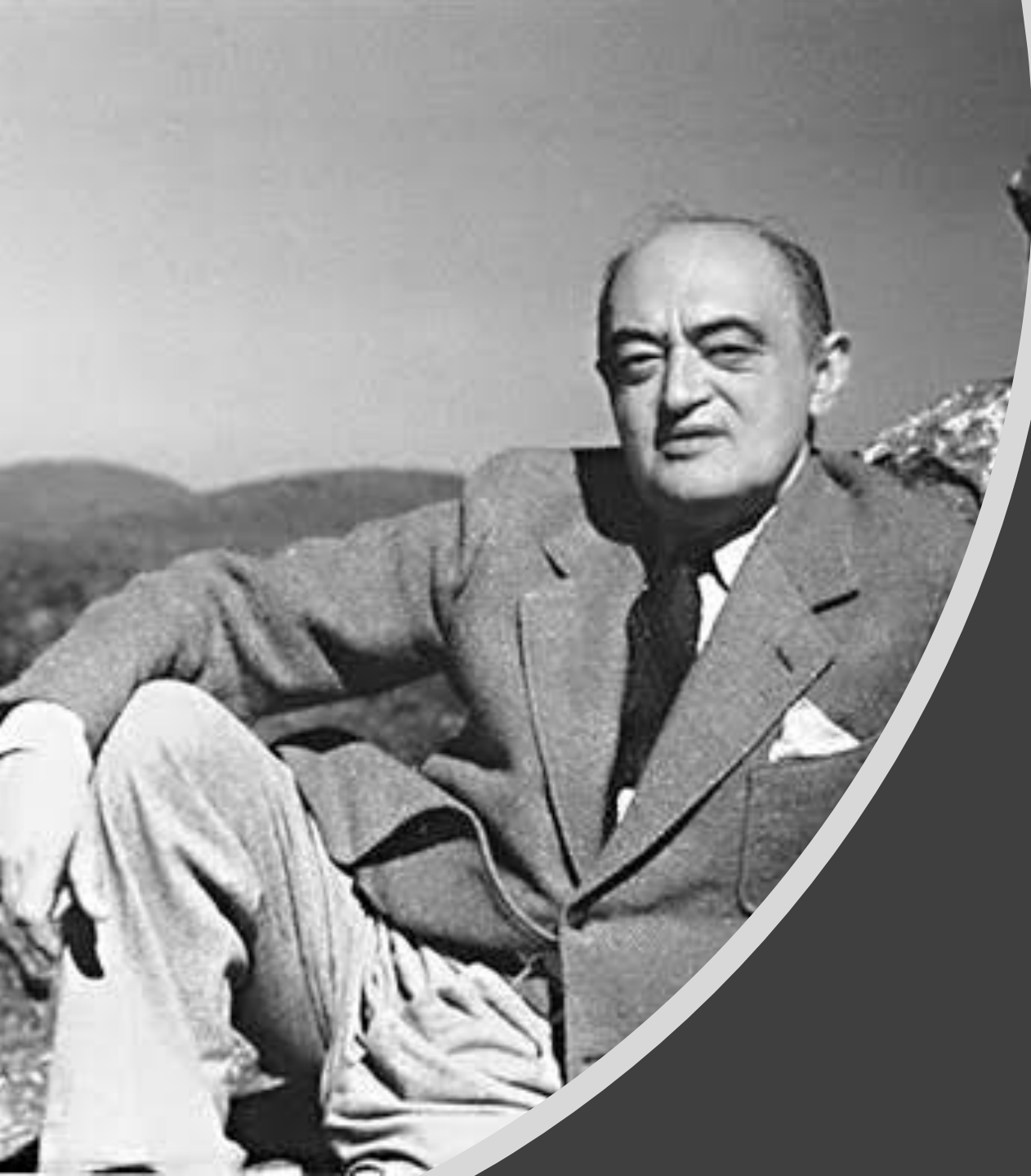
- Cross Validation (CV)
- Holdout
- Leave One Out (LOO)





- Software livre
- Amplo conjunto de bibliotecas
- Comunidade amistosa

RNN no R



Machine Learning vs Econometria

Destruição Criativa ... ? Com Certeza não.

Obrigado UEM

vbianchilanzetta@gmail.com

@vitorlanzetta

