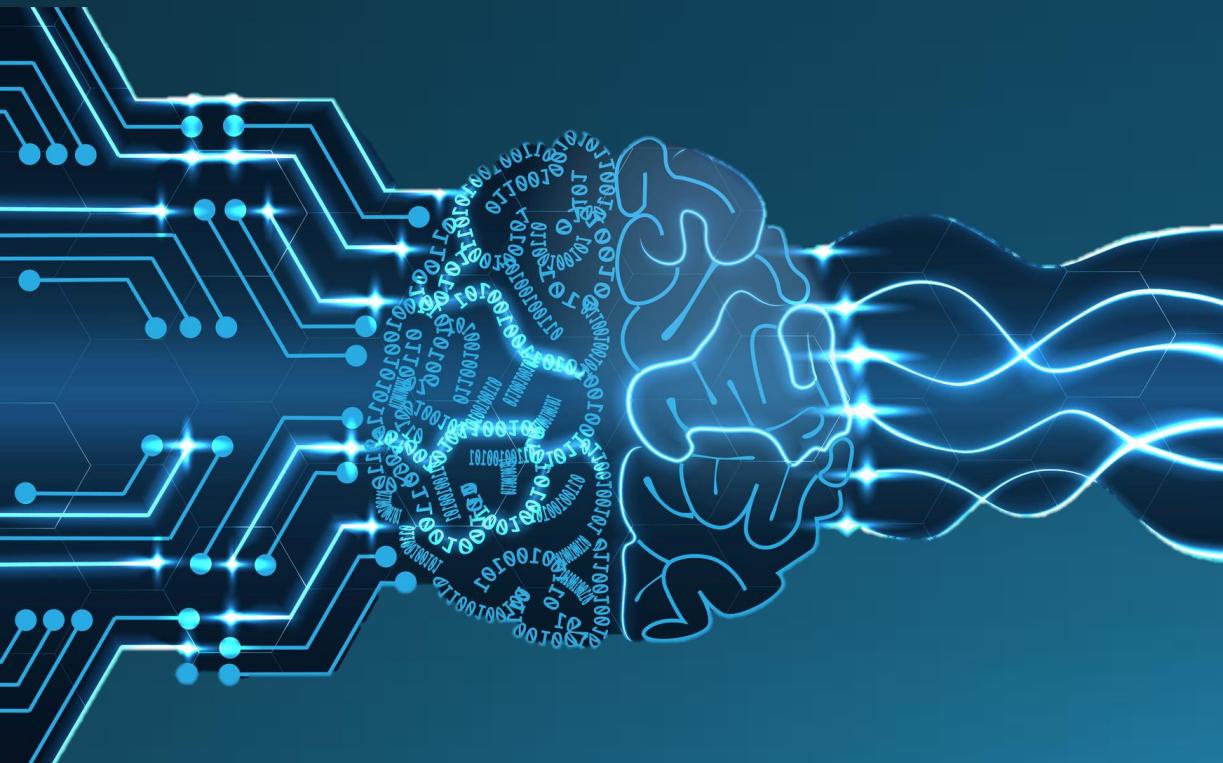


# Machine Learning

## com ferramentas Open Source



[linkedin.com/in/andrade-junior](https://linkedin.com/in/andrade-junior)



[kaggle.com/andradejunior](https://kaggle.com/andradejunior)



[twitter.com/\\_andrade\\_junior](https://twitter.com/_andrade_junior)



[github.com/andradejunior](https://github.com/andradejunior)



[t.me/andradejunior](https://t.me/andradejunior)

# O que é Machine Learning?

- Ciênci(a)e arte) de programar computadores para que eles possam aprender a partir de dados.
- *"É o campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados."*  
- Arthur Samuel, 1959



# Exemplos e Aplicações

- Reconhecimento de caracteres (OCR)
- Reconhecimento de Imagens
- Algoritmos de recomendação
- Segmentação de clientes
- Reconhecimento de voz
- Detecção de fraudes
- Diagnóstico Médico
- Filtros de spam



# Bibliotecas Open Source



Tensorflow - Google Brain (Python, C++, Haskell, Java, Go, Rust)



Scikit-Learn (Python)



Shogun (C++)



Mllib – Apache Spark (Python, Java, Scala)



Weka (Java)

# Pré-requisitos

- Probabilidade e Estatística
- Cálculo
- Álgebra Linear
- Programação



# Tipos de sistemas de Machine Learning

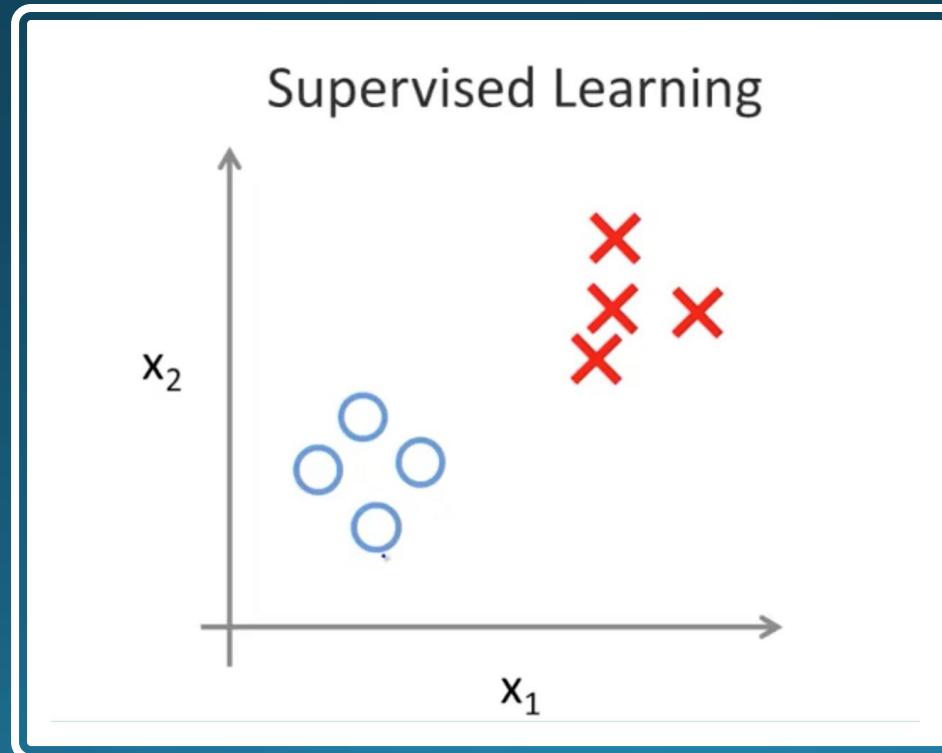
- Aprendizagem supervisionada
- Aprendizagem não-supervisionada
- Aprendizagem semi-supervisionada
- Aprendizagem por reforço



# Aprendizagem supervisionada

- Entradas e saídas desejadas (rótulos)
- O objetivo é achar a relação entre as entradas e saídas.

- Classificação.
- Regressão.

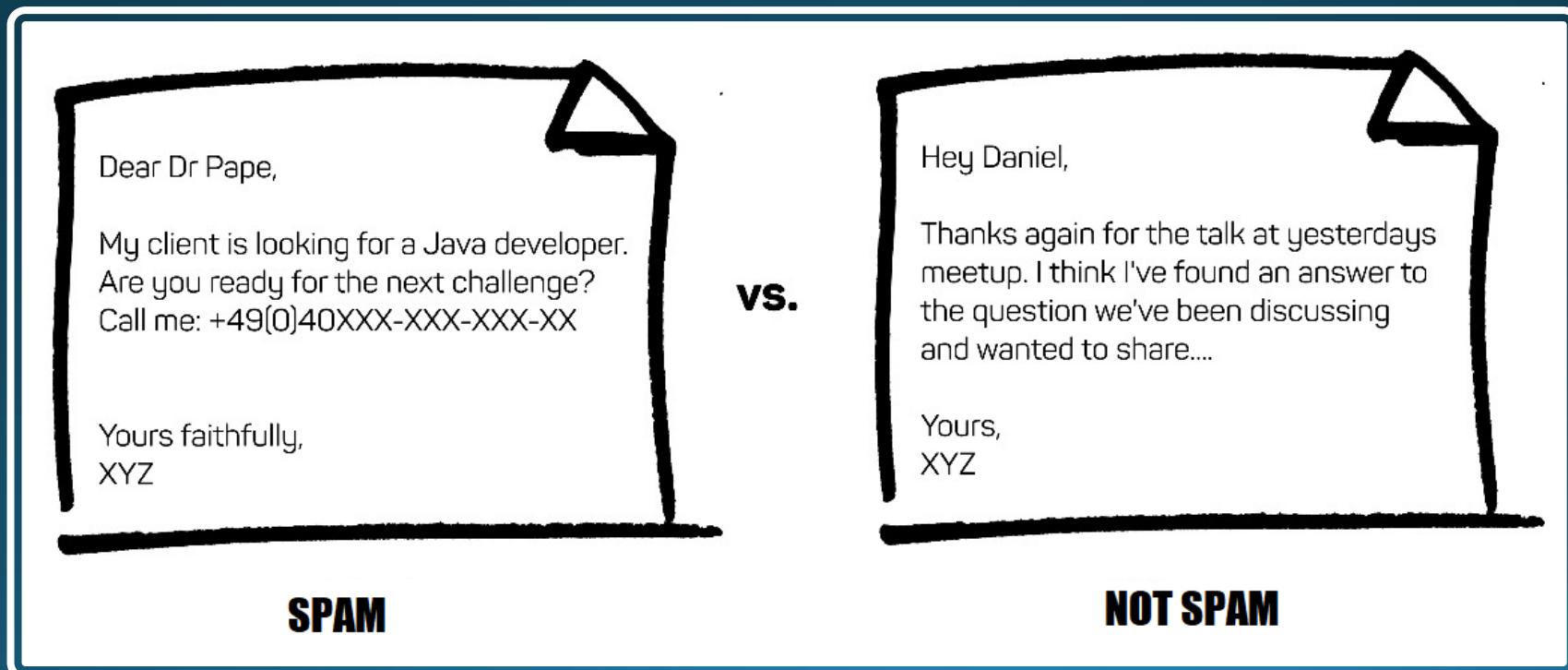


# Aprendizagem supervisionada

- Classificação

Aprender a categorizar objetos em categorias fixas.

Ex.: Filtro de spam (As classes são “spam” ou “não spam”)

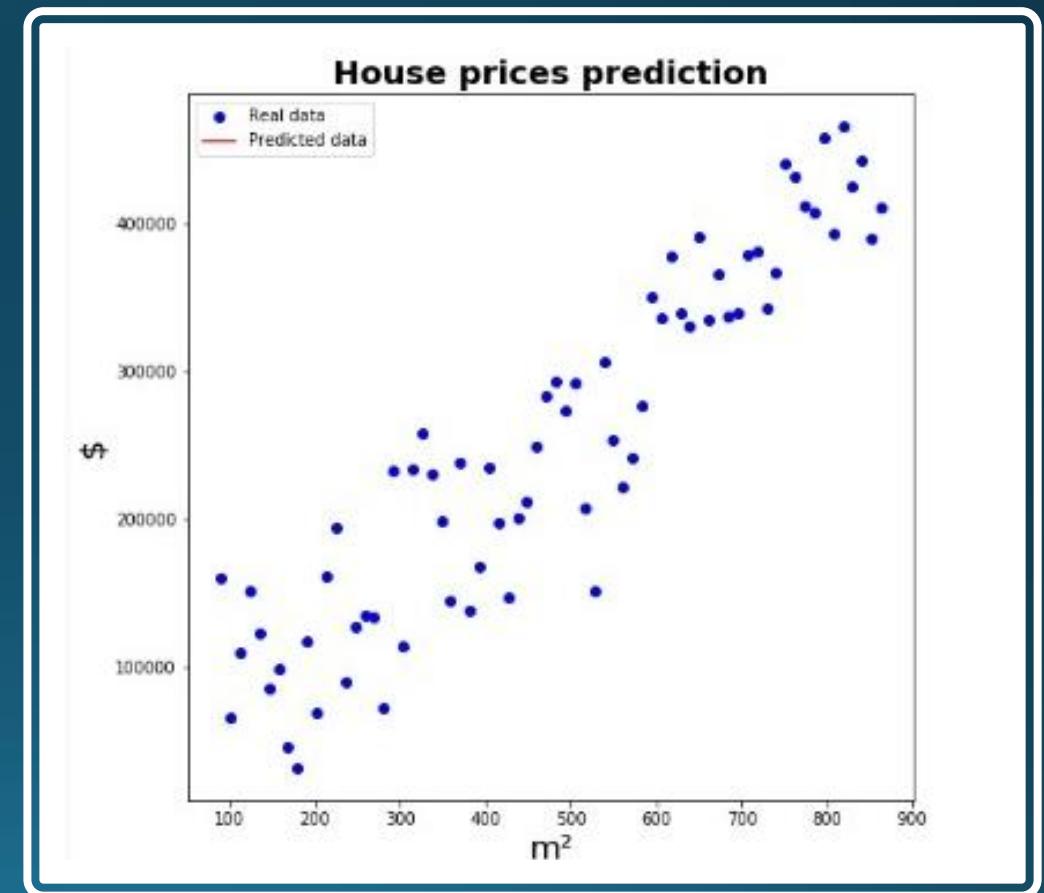


# Aprendizagem supervisionada

- Regressão

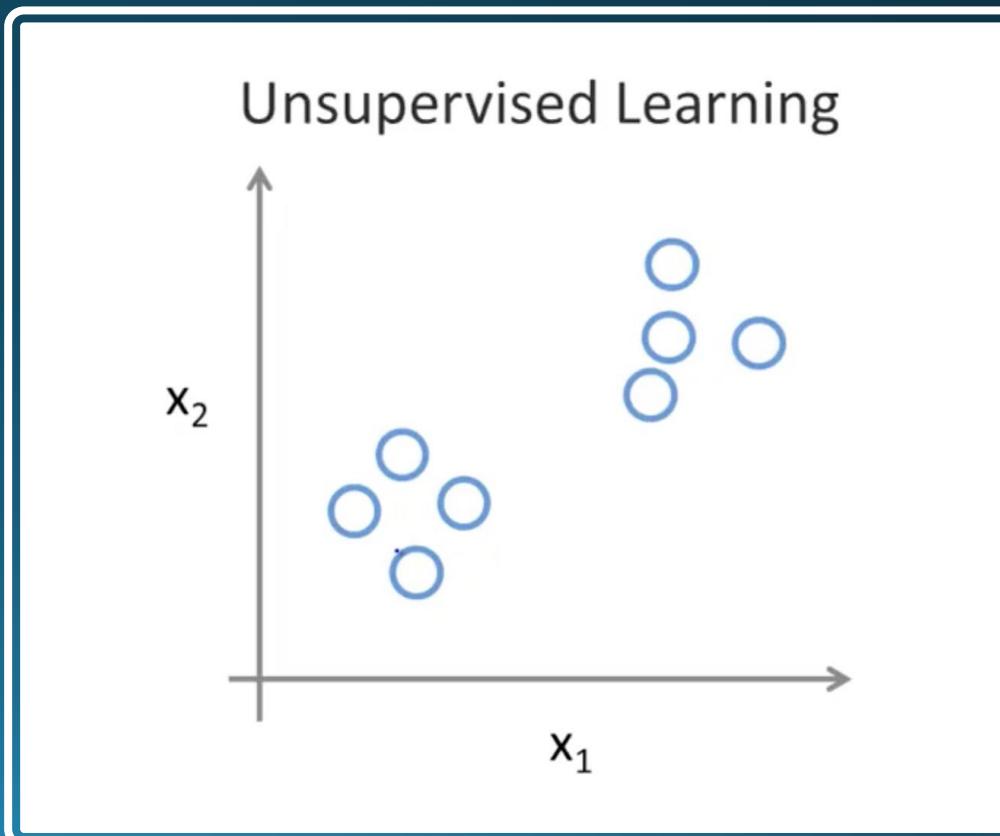
Aprender um valor real.

Ex.: Prever o valor de uma casa com base nas suas características.



# Aprendizagem não supervisionada

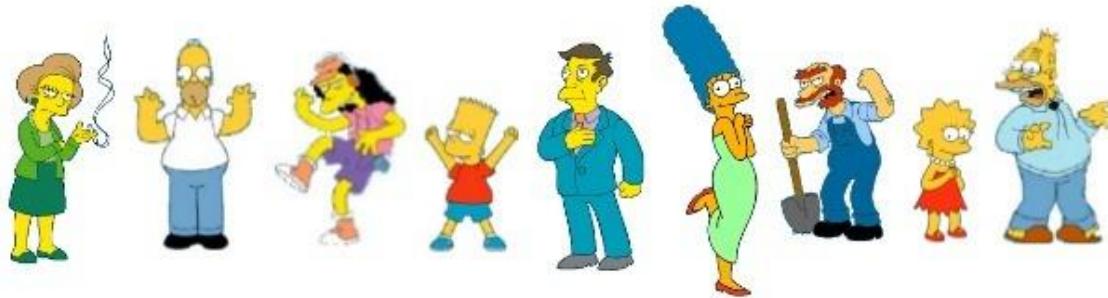
- Apenas entradas, sem saídas desejadas (sem rótulos).
  - O objetivo é encontrar padrões nas entradas.
- 
- Clusterização.
  - Redução de dimensionalidade.



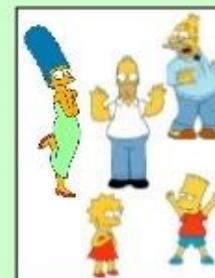
# Aprendizagem não supervisionada

- Clusterização  
Agrupar as entradas em conjuntos diferentes de acordo com suas semelhanças.

What is a natural grouping among these objects?



Clustering is subjective



Simpson's Family



School Employees



Females

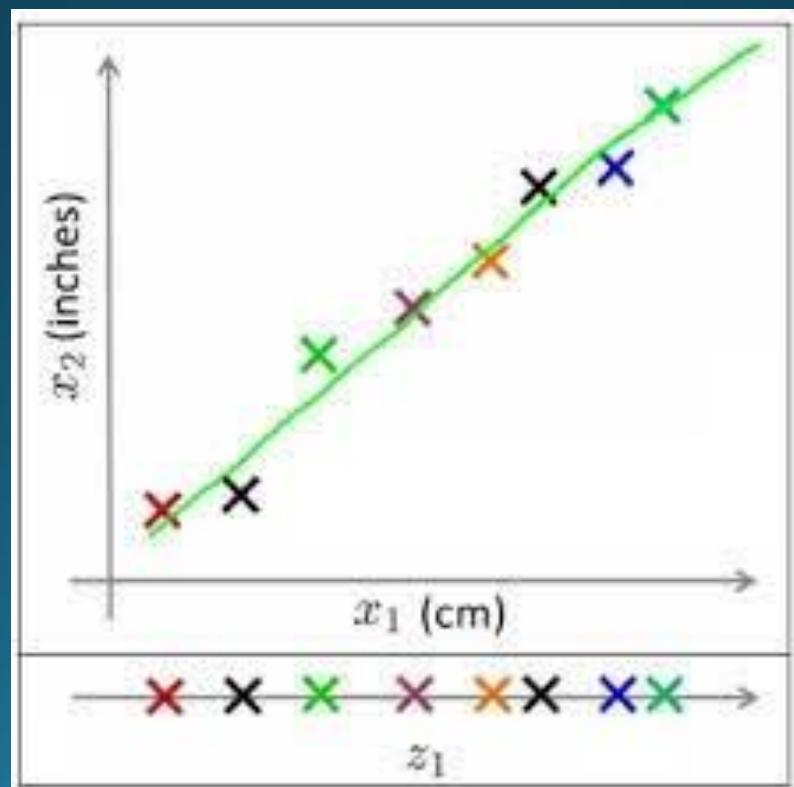


Males

# Aprendizagem não supervisionada

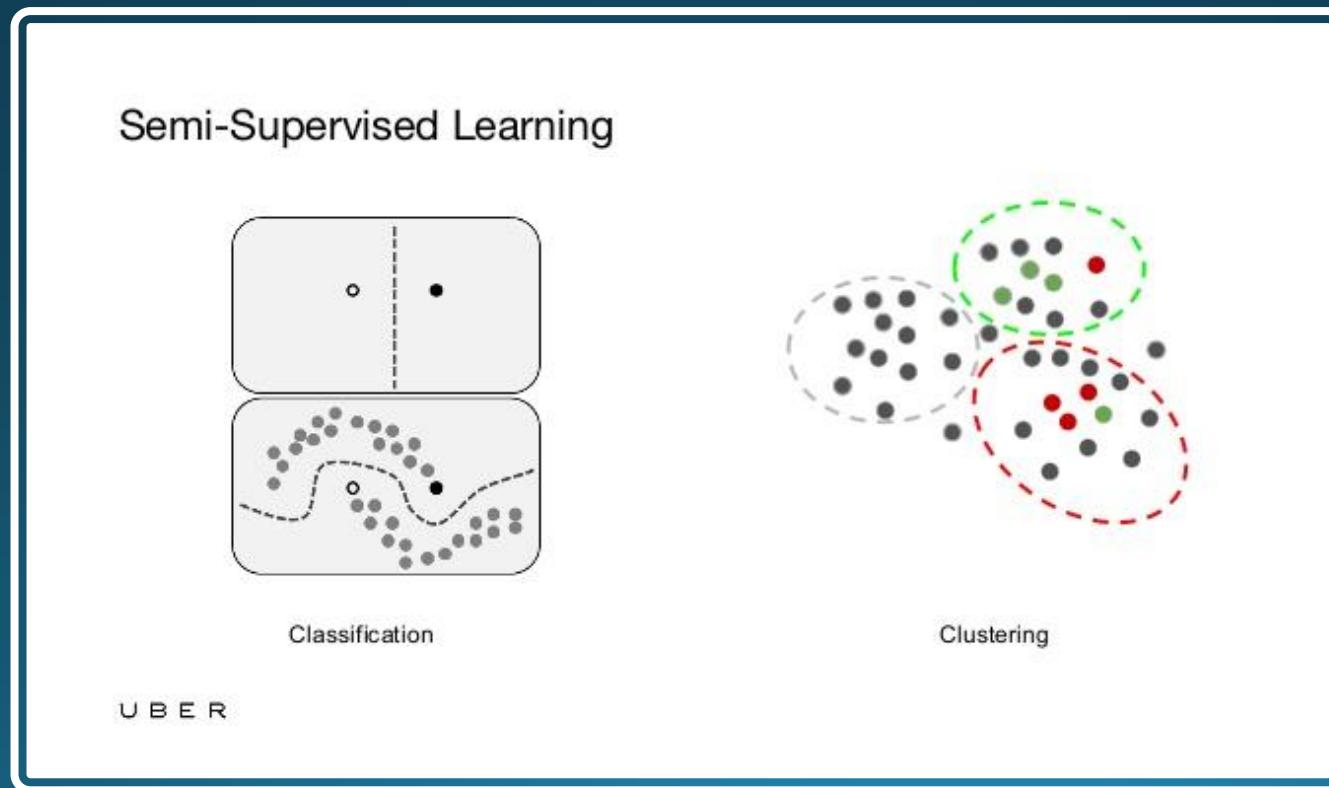
- Redução de dimensionalidade.

Reducir a quantidade de entradas, mapeando-as para um espaço de dimensão menor.



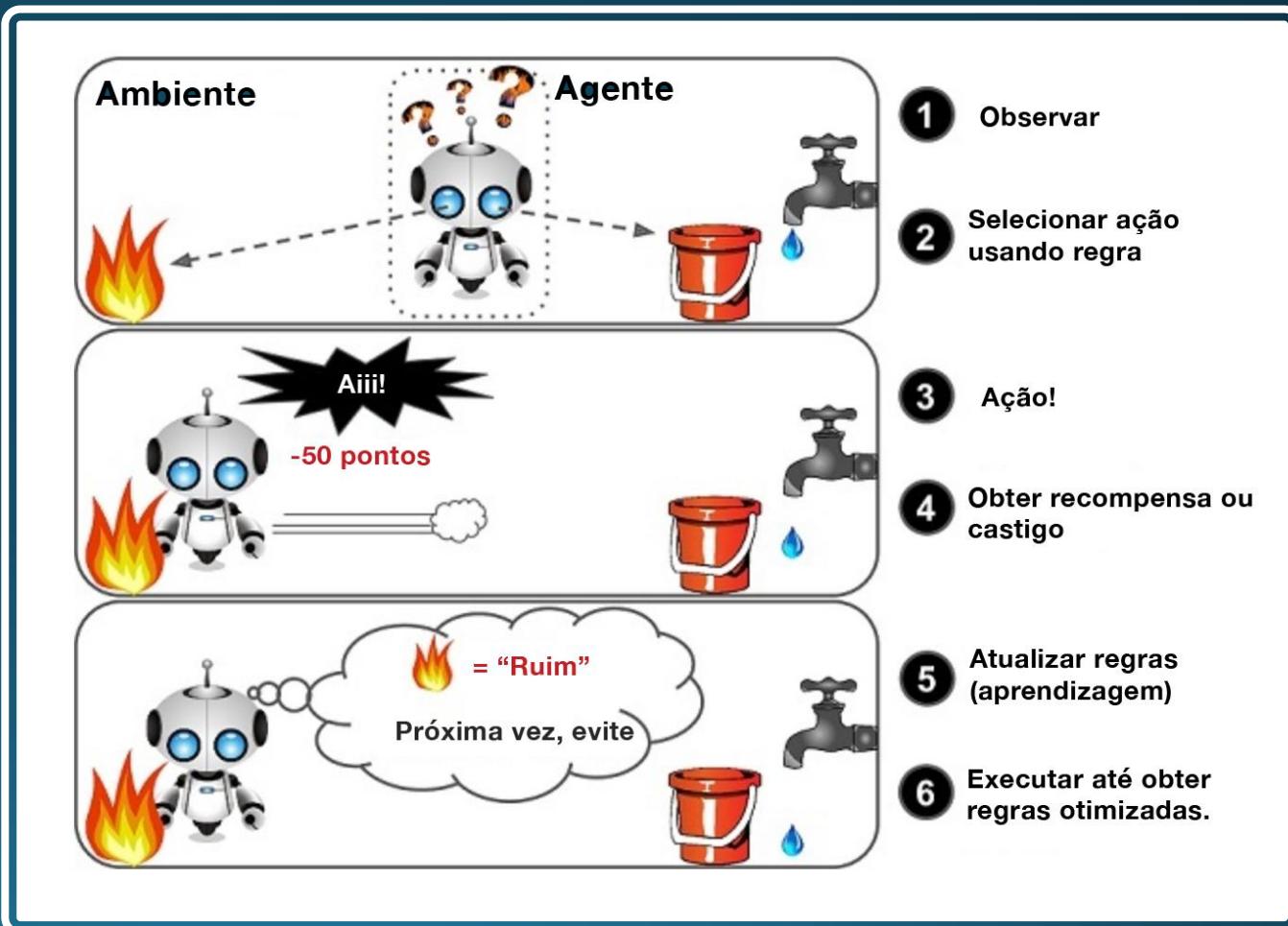
# Aprendizagem semi-supervisionada

- Combina uma pequena quantidade de amostras rotuladas com um grande número de amostras não rotuladas para produzir melhores classificadores.



# Aprendizagem por reforço

- Um agente pode observar o ambiente, selecionar e realizar ações e receber recompensas em retorno (ou penalidades).





ALPHAGO

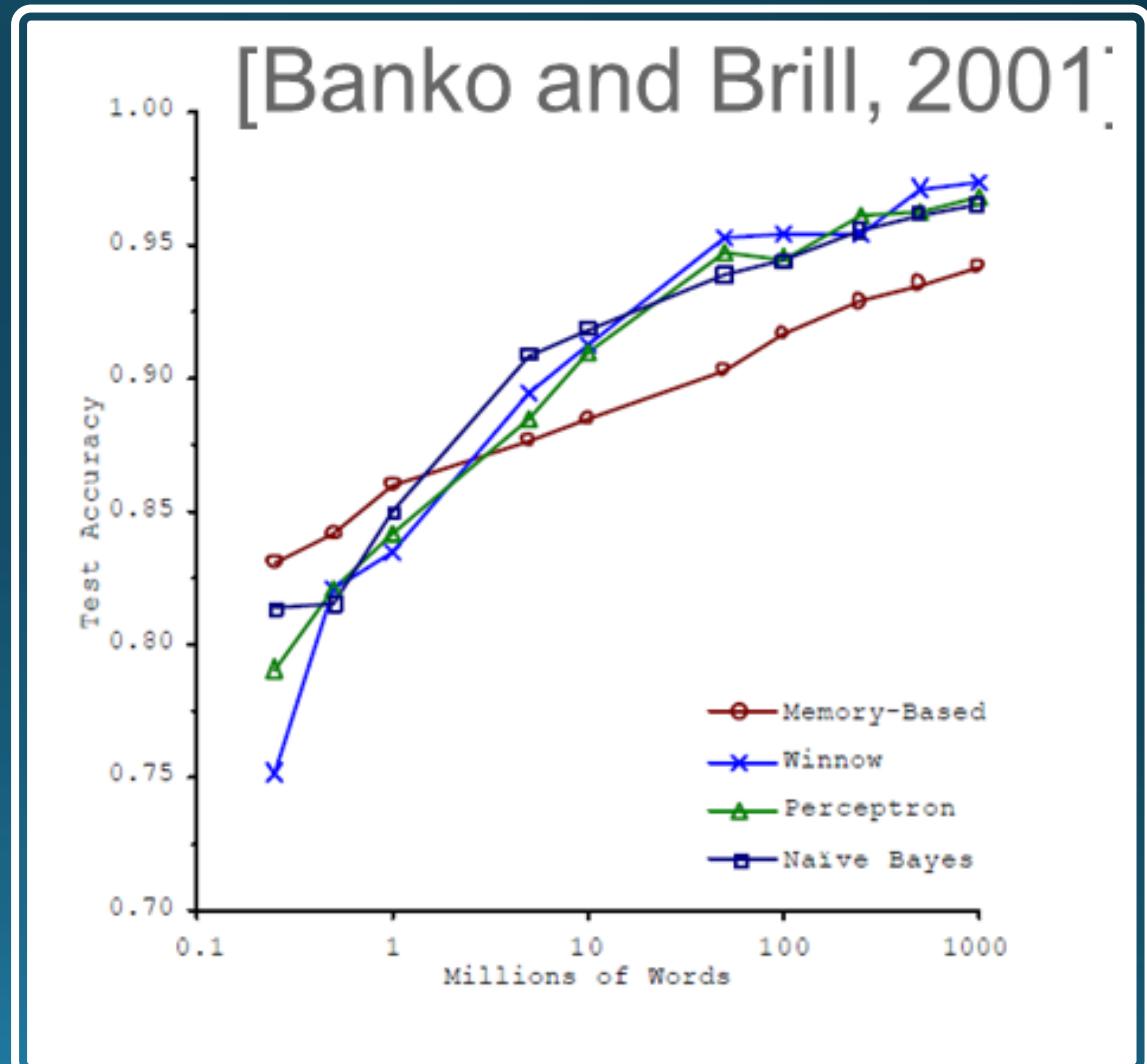
# Principais desafios em Machine Learning

- Baixa qualidade dos dados
- Features irrelevantes
- Dados de treinamentos não representativos
- Quantidade insuficiente de dados de treinamento
- Overfitting (Sobreajuste)
- Underfitting (Subajuste)



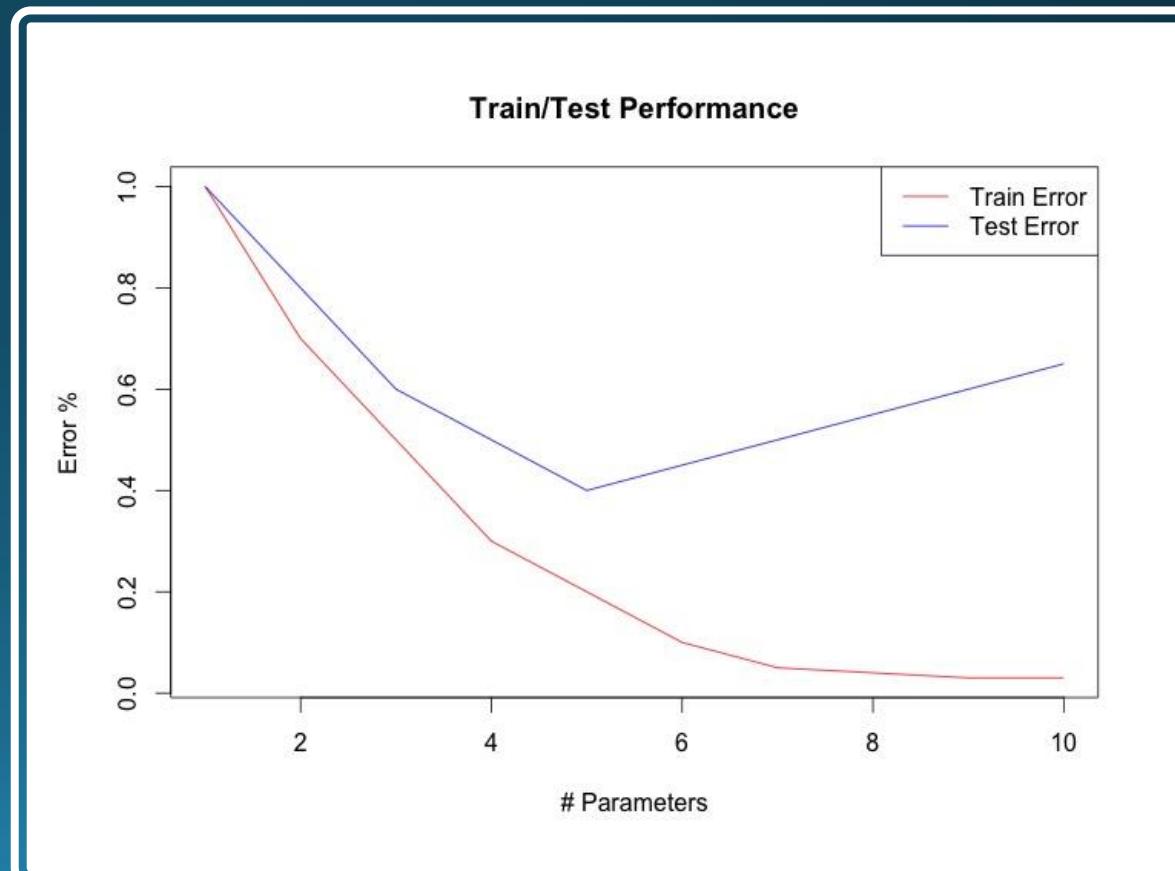
# Dados > Algoritmos

- The unreasonable effectiveness of data (<http://goo.gl/R5enlE>), publicado em 2001.



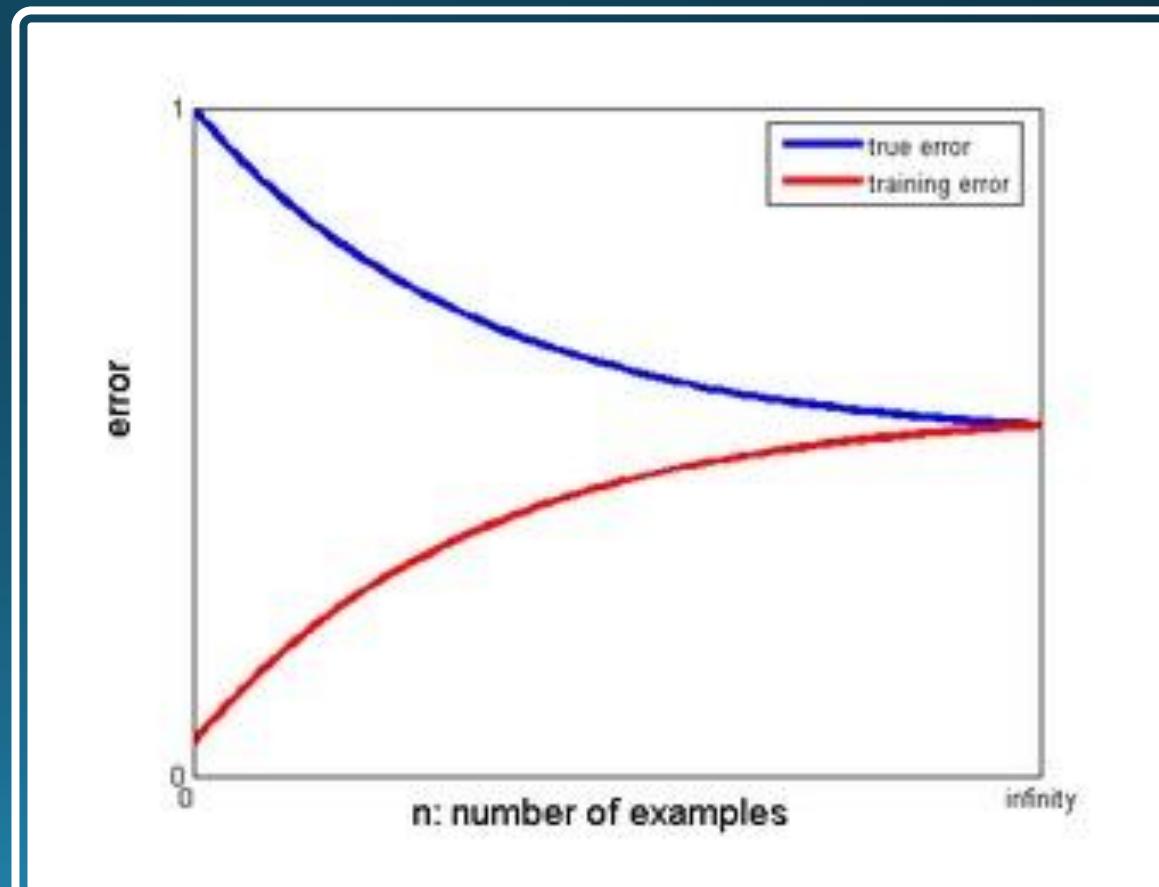
# Overfitting

- Quando o modelo se ajusta muito bem ao conjunto de dados de treinamento mas se mostra ineficaz para prever novos resultados.
- Possíveis soluções:
  - ✓ Reduzir o número de atributos
  - ✓ Coletar mais dados de treinamento
  - ✓ Limpar os dados

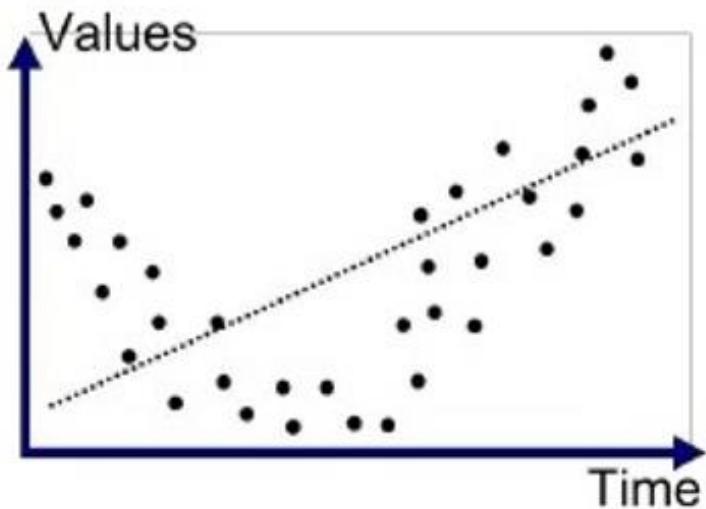


# Underfitting

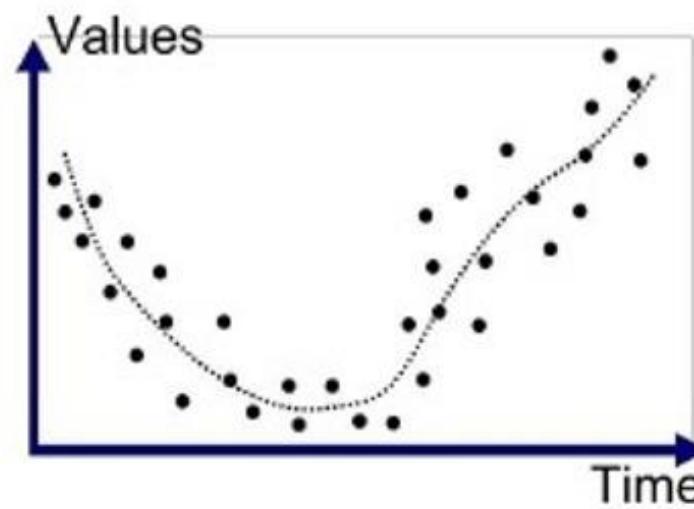
- Oposto ao Overfitting, quando um modelo é muito simples para aprender o padrão dos dados.
- Soluções:
  - ✓ Selecionar um modelo mais complexo
  - ✓ Adicionar melhores Features



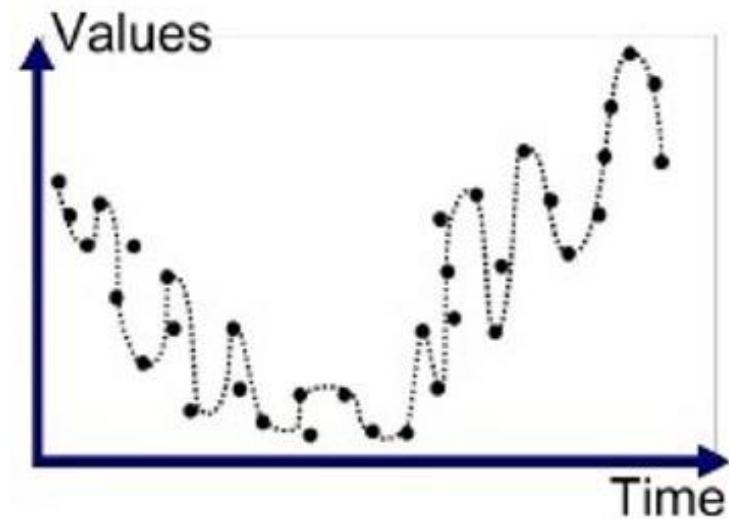
# Underfitting VS. Overfitting



Underfitted



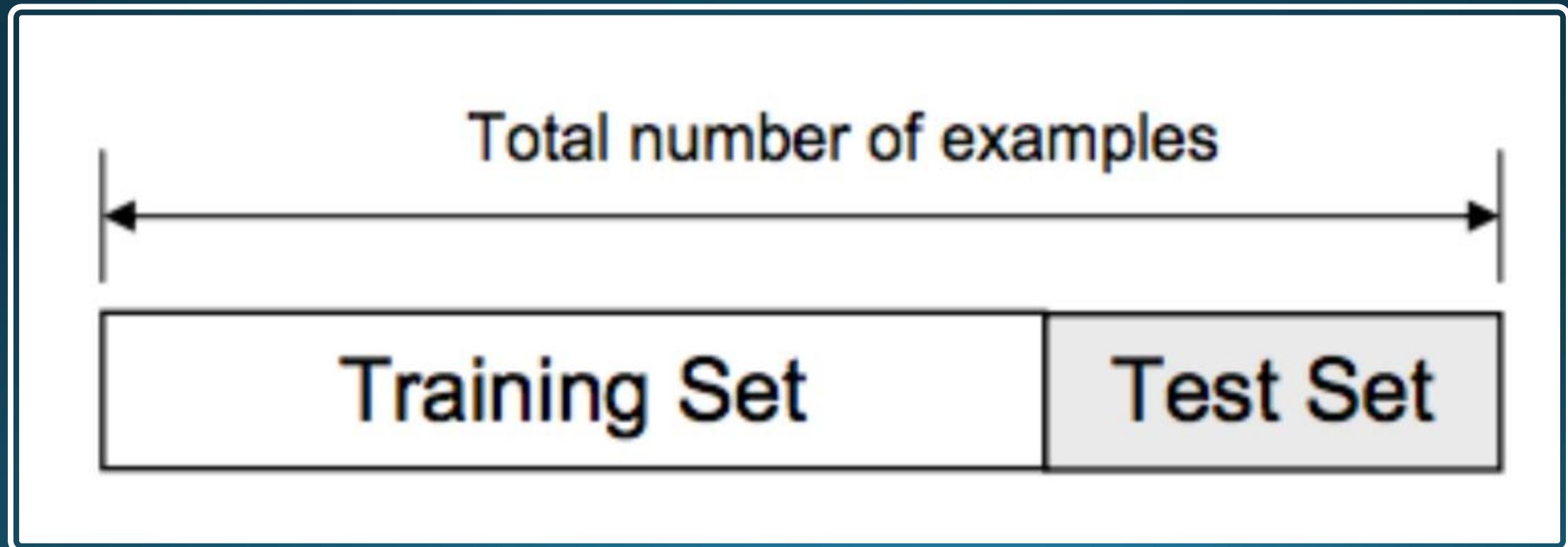
Good Fit/Robust



Overfitted

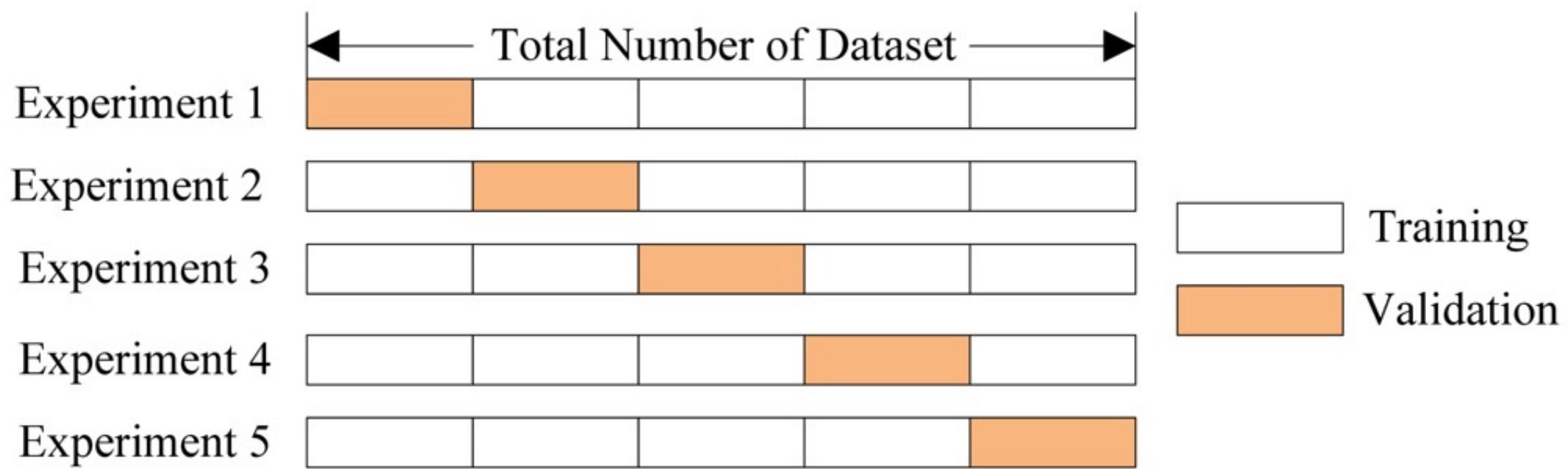
# Teste e Validação

- Para validar corretamente, é necessário dividir nossos dados em dois conjuntos: Conjunto de Treinamento e Conjunto de Teste.



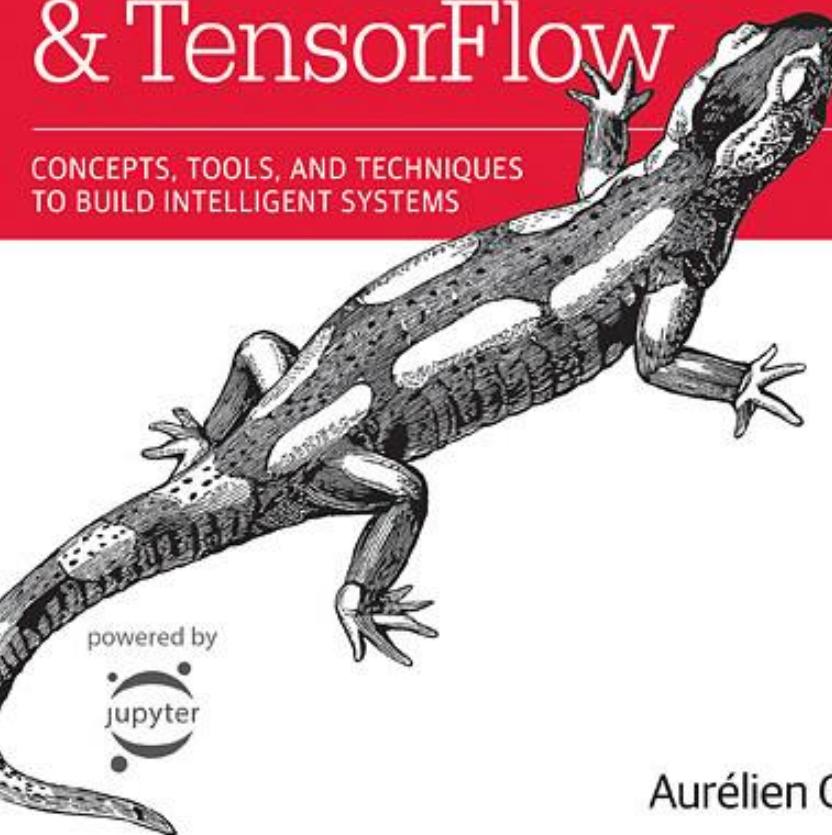
# Cross-Validation

## (Validação Cruzada)



# Hands-On Machine Learning with Scikit-Learn & TensorFlow

CONCEPTS, TOOLS, AND TECHNIQUES  
TO BUILD INTELLIGENT SYSTEMS



Aurélien Géron

# Referências

- Hands-On Machine Learning with Scikit-Learn and TensorFlow by Aurélien Géron (O'Reilly) © 2017.



# Dúvidas?



# ENOUGH TALKING

