

# PANDA UFSCar

## Introdução ao Aprendizado de Máquina com scikit-learn

Por Joãozinho



Apresentação inspirada nos capítulos do livro:  
**Hands-On Machine Learning**, Aurélien Géron

# What is machine learning?

“ Machine learning (ML) is the study of computer algorithms that improve automatically through experience. It is seen as a subset of artificial intelligence. Machine learning algorithms build a model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so. ”

Fonte: [Machine learning\\_\(Wikipedia\)](#).

Parte dessa noção de ML já existia em **1959!!!** [Arthur Samuel, 1959](#)  
Outro autor "antigo": [Tom Mitchell, 1997](#)

# O que é aprendizado de máquina?

“ O aprendizado de máquina (ML) é o estudo de algoritmos de computador que melhoram automaticamente com a experiência. É visto como um subconjunto da inteligência artificial. Os algoritmos de aprendizado de máquina constroem um modelo baseado em dados de amostra, conhecidos como "dados de treinamento", a fim de fazer previsões ou decisões sem serem explicitamente programados para isso. ”

Fonte: [Machine learning \(Wikipedia\)](#) + [Google Tradutor](#)

# Supervisionado e Não-supervisionado

- **Aprendizado supervisionado:** o computador "conhece" a resposta desejada para cada situação usada para treino  
*Ex: um classificador de spam que é treinado com vários e-mails e junto deles a informação se eles são spam ou inofensivos*
- **Aprendizado não-supervisionado:** o computador não conhece essa resposta desejada para os dados  
*Ex: um algoritmo que agrupa pessoas semelhantes dentro da universidade, com base no gosto musical e de séries*

# Overfitting e Underfitting

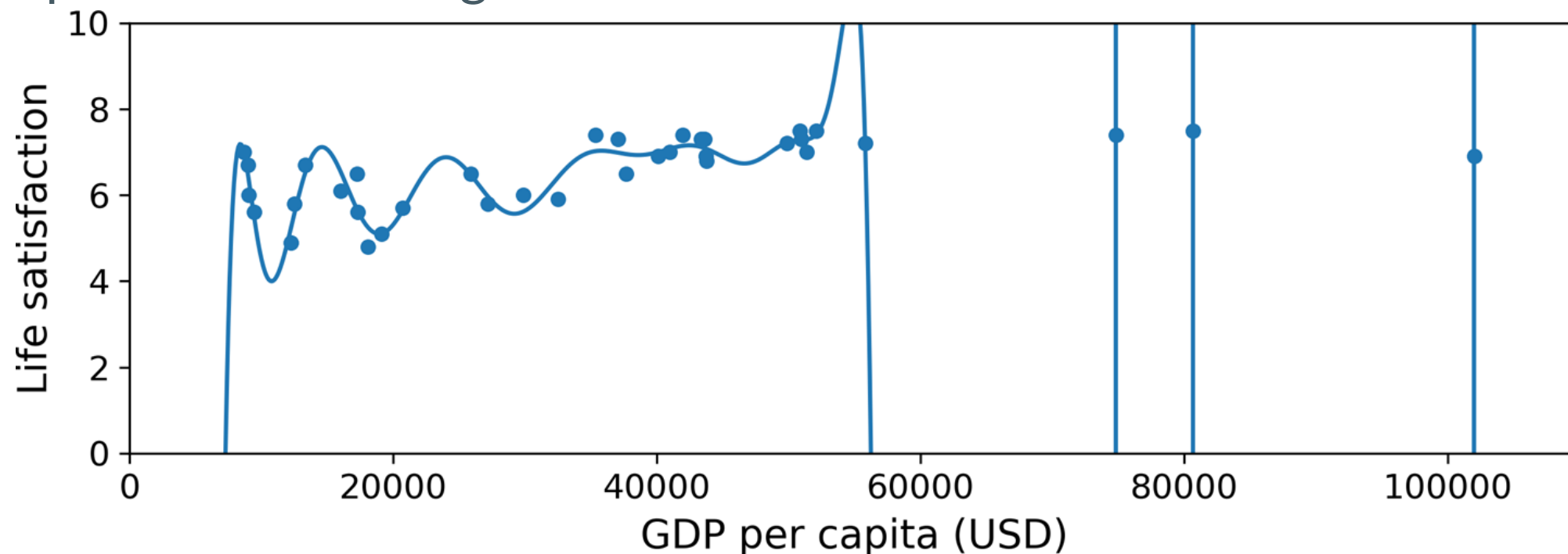
- Um modelo pode se sair muuuuito bem na hora de ser treinado, mas quando chega na hora de generalizar para dados novos...

**OVERFITOU!**



# Overfitting e Underfitting

Exemplo de overfitting

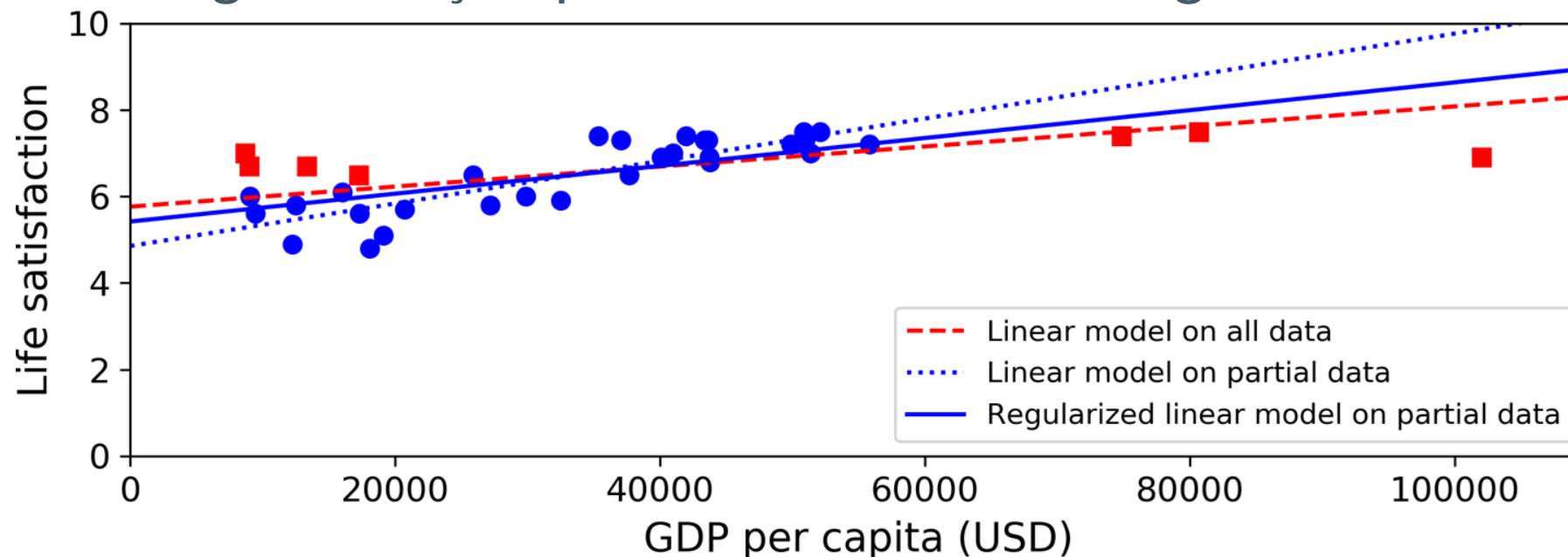


Fonte: Aurélien Géron, 2019

# Overfitting e Underfitting

SIMPLIFIQUE!!!

Exemplo de regularização para conter overfitting



Fonte: Aurélien Géron, 2019

# Overfitting e Underfitting

- Eis que você decide simplificar. Um modelo simples não vai trazer problemas, certo?

**ERRADO. UNDERFITOU!**





# Treino e Teste

- Um modelo de aprendizado de máquina é primeiramente treinado. Depois disso, como saber se o modelo é bom?
- E se usarmos o mesmo conjunto que treinamos para testar?  
***Ruim**, há chance de termos um modelo que não generaliza bem para dados fora do conjunto de treino.*
- **Saída:** dividirmos o nosso conjunto em um conjunto de **treino** e um de **teste** (geralmente 80/20 ou 75/25)! Assim, podemos avaliar o nosso erro out-of-sample (ou de generalização).