

Workshop Intensivo de Aplicações Modernas de Ciência de Dados com Machine Learning

Quarto dia

Paulo Cysne Rios Jr.

Modelagem

Satisfação de Vida e Riqueza (PNB)

This repository

Search

Pull requests Issues Marketplace Explore

paulocr2 / Workshop-App-ML-UNI7

<> Code Issues 0 Pull requests 0 Projects 0 Wiki Insights

Documentação e programas do Workshop em Aplicações Modernas de Machine Learning dad

Add topics

68 commits 1 branch 0 releases

Branch: master New pull request Create new file

paulocr2 Atualização do Dia 3

data	data added
day1	update2
day2	updated
img	img added
<u>Life Satisfaction and GDP - Linear Model Ply.ipynb</u>	Day 3
README.md	Atualização do Dia 3
<u>Workshop - Day 3.pdf</u>	Day 3

README.md

Bias Variance Tradeoff



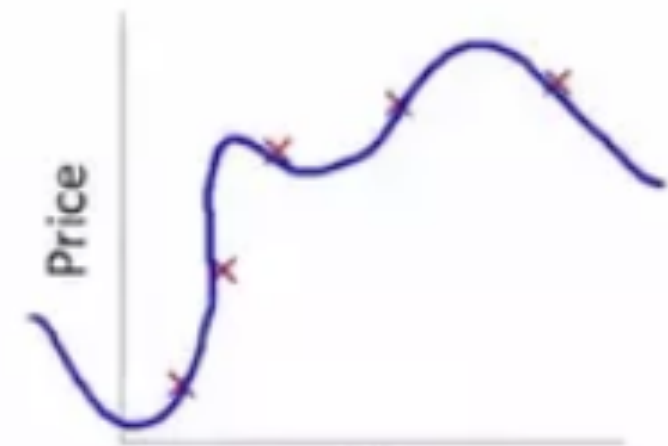
Size
 $\theta_0 + \theta_1 x$

High bias
(underfit)



Size
 $\theta_0 + \theta_1 x + \theta_2 x^2$

"Just right"



Size
 $\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$

High variance
(overfit)

Comparação das Modelagens



Linear Simples

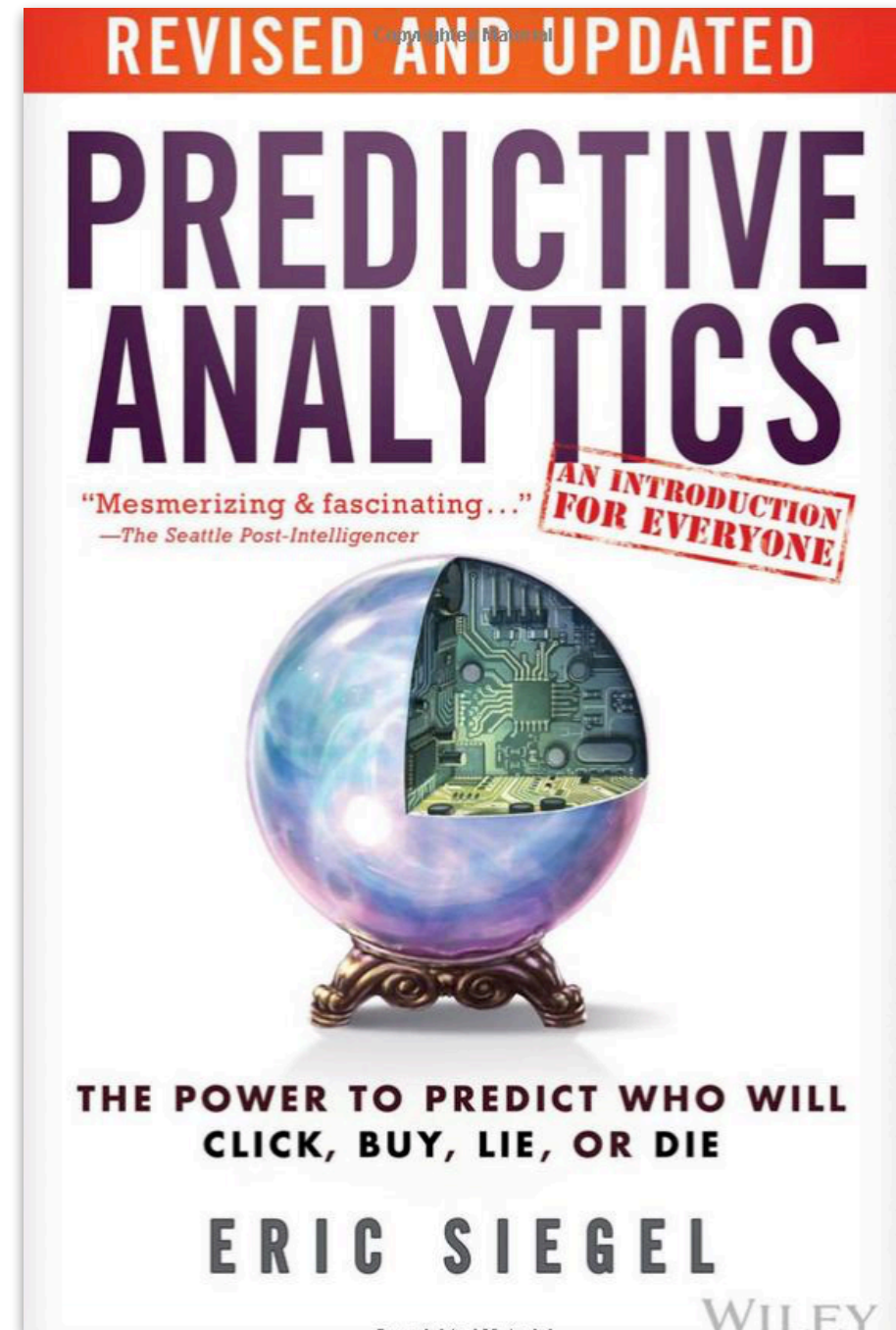


Linear with Polinômio de Alto Grau

Programa Python com nova Modelagem

Ciência de Dados e Analítica Preditiva

- Qual a diferença entre **Ciência de Dados** e **Analítica Preditiva**?
- Resposta....



Resposta

- **Analítica Preditiva (Predictive Analytics)**
- **Aprendizagem de Máquina (Machine Learning)**
- *é a mesma coisa!* (segundo o autor deste livro muito vendido mundialmente que já está na segunda edição)
- Mas é mesmo??

Prática

Contéudo

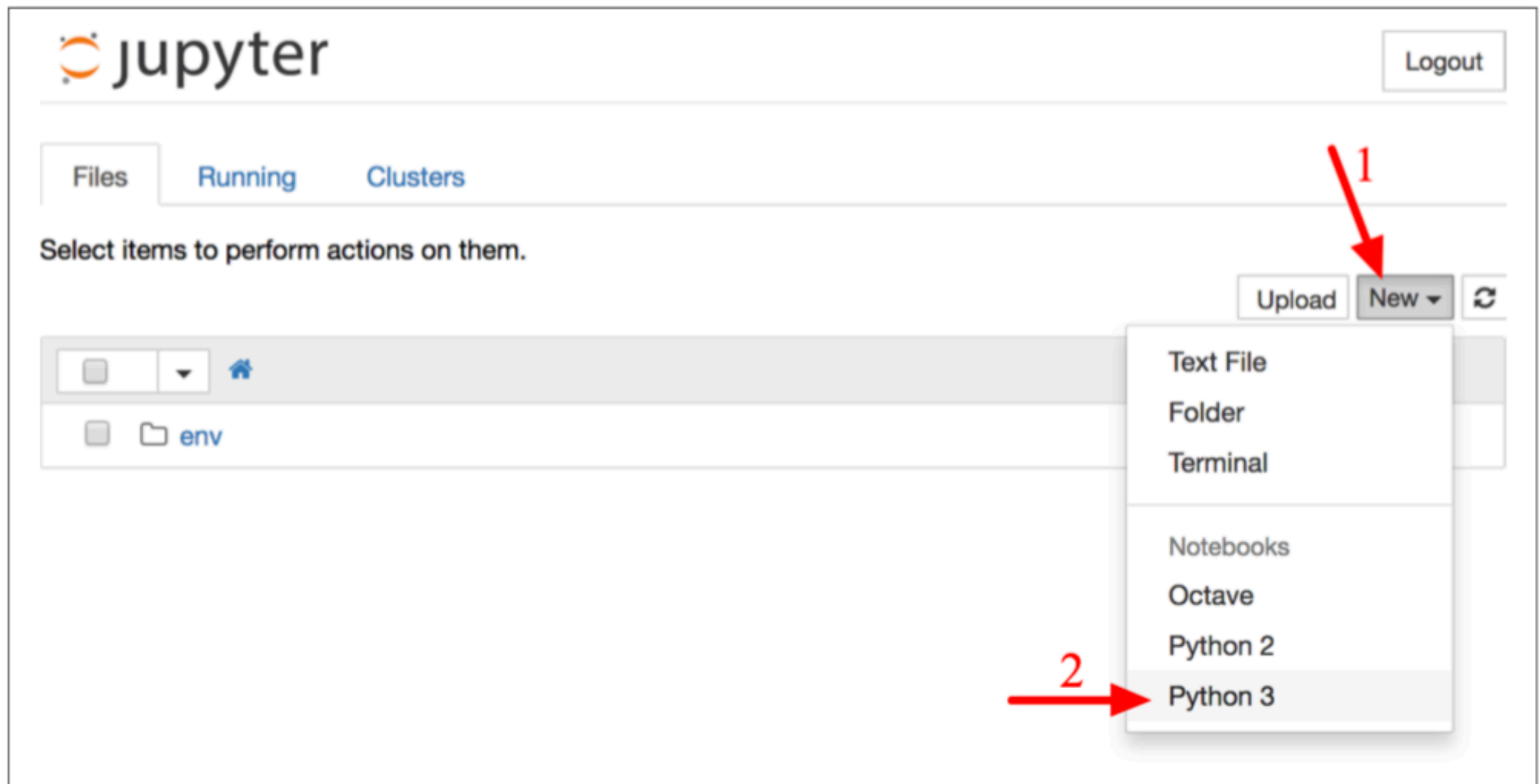
Módulo III – Curso Rápido de Python para Machine Learning

- Ambiente de Programação
 - Anaconda
 - Jupyter Notebooks
- Os conceitos básicos
 - Estruturas básicas da linguagem: if/else, loops
 - Listas, dicionários, tuplas e conjuntos
- Programação funcional e orientada a objetos
 - Funções, Lambdas, Classes
 - Modules e Packages
 - Funções zip, enumerate, all, any, map/reduce/filter
- Numpy
 - Criação a partir do zero, criação a partir de listas
 - Indexing, slicing, reshaping
 - Concatenação e divisão (splitting)
 - Agregação, Min, Max e mais
 - Broadcasting
 - Comparação, masks, lógica booleana, sorting

Conteúdo

- Pandas
 - Series, DataFrames
 - Indexação e Seleção em Series e em DataFrames
 - Operando com dados com valores nulos
 - Encontrando valores distintos
 - Explorando Series e DataFrames
 - Combinando dados
 - Agregação e agrupamento, pivot tables
 - eval() e query()
- Visualização de Dados
 - Matplotlib
 - Seaborn
- Preprocessamento de Dados com SciKit Learn
 - Limpeza de dados
 - Feature Engineering
 - Hot encoders
 - Scaling
 - Principal Component Analysis (PCA)
 - Pipelines
 - Estimator API

Criando um novo notebook



Curso Rápido de Python

Estruturas de Dados