

MBAUSP
ESALQ

INTRODUÇÃO À PROGRAMAÇÃO COM PYTHON

Prof. Dr. Wilson Tarantin Junior

MBAUSP ESALA

A responsabilidade pela idoneidade, originalidade e licitude dos conteúdos didáticos apresentados é do professor.

Proibida a reprodução, total ou parcial, sem autorização.

Lei nº 9610/98

Introdução ao Machine Learning

Machine learning

- Definição
 - Podem ser encontradas muitas definições para o conceito machine learning
 - Porém, de forma geral, pode-se entender como a utilização dos dados e de algoritmos (que contêm também métodos estatísticos) para a produção de informações que serão relevantes para a tomada de decisão
 - Por exemplo: criando modelos preditivos ou analisando a interdependência entre os dados

Banco de dados

- Definição e composição
 - O banco de dados é o objeto onde estão armazenadas as informações de interesse para a análise ou estudo em questão
 - Em muitos casos, o banco de dados contém uma amostra, que nada mais é do que um subconjunto extraído da população
 - O banco de dados é composto por variáveis e por observações
 - Variáveis: características/atributos observados, medidos ou categorizados
 - Observações: as unidades que têm suas características e atributos medidos

Banco de dados

- Estrutura para uso
- Normalmente, o banco de dados é estruturado com as variáveis em colunas e as observações em linhas em uma estrutura tabular

	Idade	Altura	Cidade	Profissão	Renda	...
Pessoa 1						
Pessoa 2						
Pessoa 3						
Pessoa 4						
Pessoa 5						
Pessoa 6						
Pessoa 7						
Pessoa 8						
Pessoa 9						
Pessoa 10						
...						

Tipos de variáveis

- As variáveis podem ser divididas em
 - **Métricas:** são as variáveis quantitativas, isto é, apresentam características que podem ser mensuradas ou contadas
 - **Não métricas:** são as variáveis qualitativas, sendo que indicam características que não podem ser medidas. Tais variáveis contêm categorias, por isto, muitas vezes, são chamadas de variáveis categóricas
 - **A identificação do tipo de variável é fundamental para a escolha do método estatístico que será utilizado na análise dos dados**

Variáveis qualitativas

- Características principais
 - As variáveis qualitativas têm sua representação feita por meio de tabelas de distribuição de frequências ou gráficos
 - Não é possível calcular medidas de resumo como média ou desvio padrão para variáveis qualitativas
 - Exemplos: escalas likert, nacionalidade, cor do veículo, profissão...

Variáveis quantitativas

- Características principais
 - As variáveis quantitativas podem ser representadas por diversas ferramentas, como gráficos, medidas de posição ou localização, dispersão e de forma. A seguir, alguns exemplos:
 - Medidas de posição ou localização: média, mediana, percentis, quartis
 - Medidas de dispersão: variância e desvio padrão
 - Medidas de forma: assimetria e curtose
 - Exemplos: idade, renda em Reais, número de habitantes no município, distância em metros, rentabilidade percentual de uma empresa...

Detalhando as variáveis

- Outras características relevantes
 - Variáveis qualitativas: dicotômica ou policotômica; nominal ou ordinal
 - Dicotômica: duas categorias (binária); Policotômica: mais de duas categorias
 - Nominal: não estabelece relação de grandeza/ordem; Ordinal: estabelece uma ordem
 - Variáveis quantitativas: discretas ou contínuas
 - Discretas: possuem conjunto finito e numerável de valores, em geral, são obtidas a partir de dados de contagem (0, 1, 2, 3, 4, 5...)
 - Contínuas: assumem valores pertencentes ao intervalo de números reais

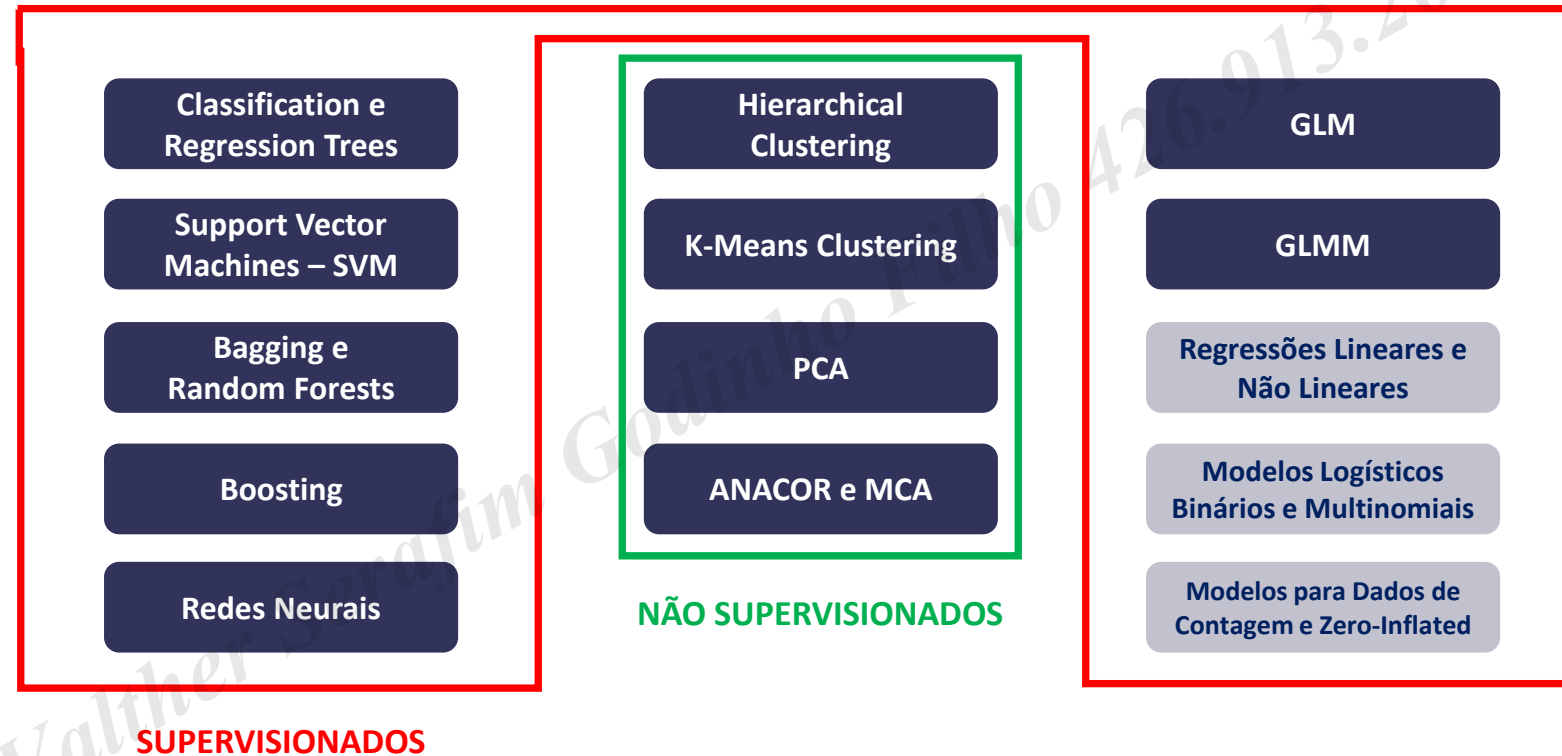
Variáveis e técnicas

- Como identificar a técnica adequada
 - Modelos **não supervisionados** de machine learning
 - Também são conhecidos como modelos **exploratórios** ou técnicas de interdependência
 - Estudo da relação entre variáveis, mas sem a intenção de criar modelos confirmatórios
 - Não há inferência dos resultados encontrados para observações fora da amostra
 - Objetivos: redução dos dados, classificação ou agrupamento de observações e variáveis, correlação ou associação entre variáveis

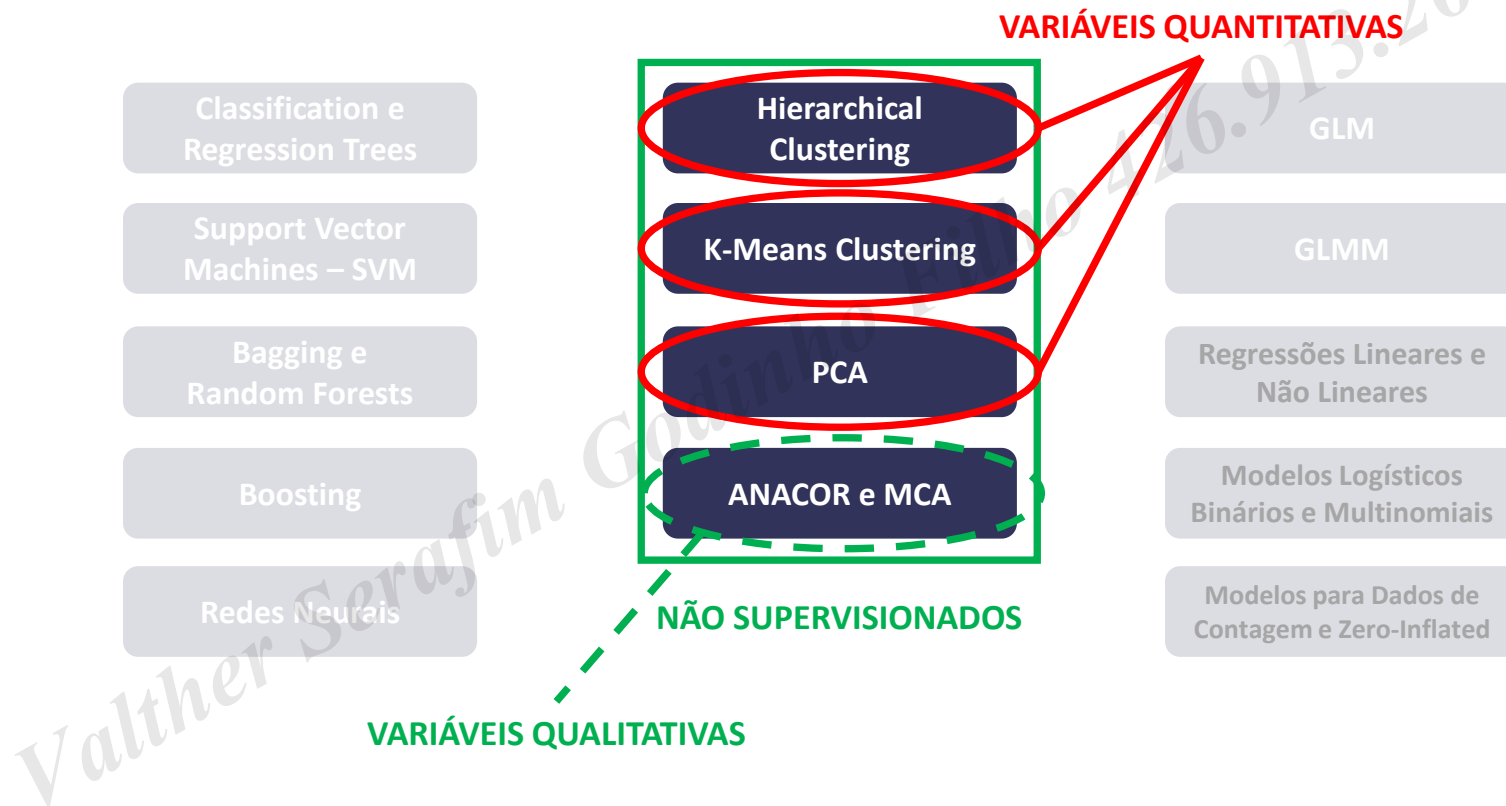
Variáveis e técnicas

- Como identificar a técnica adequada
 - Modelos **supervisionados** de machine learning
 - Conhecidos como modelos **confirmatórios** ou técnicas de dependência
 - O objetivo é estimar modelos, equações, com o intuito de **elaborar previsões**
 - Portanto, há **inferência** dos resultados para outras observações fora da amostra

Classificação de algoritmos



Classificação de algoritmos



Classificação de algoritmos



Referência

- Fávero, Luiz Paulo; Belfiore, Patrícia. (2024). Manual de análise de dados: estatística e machine learning com Excel®, SPSS®, Stata®, R® e Python®. 2 ed. Rio de Janeiro: LTC.

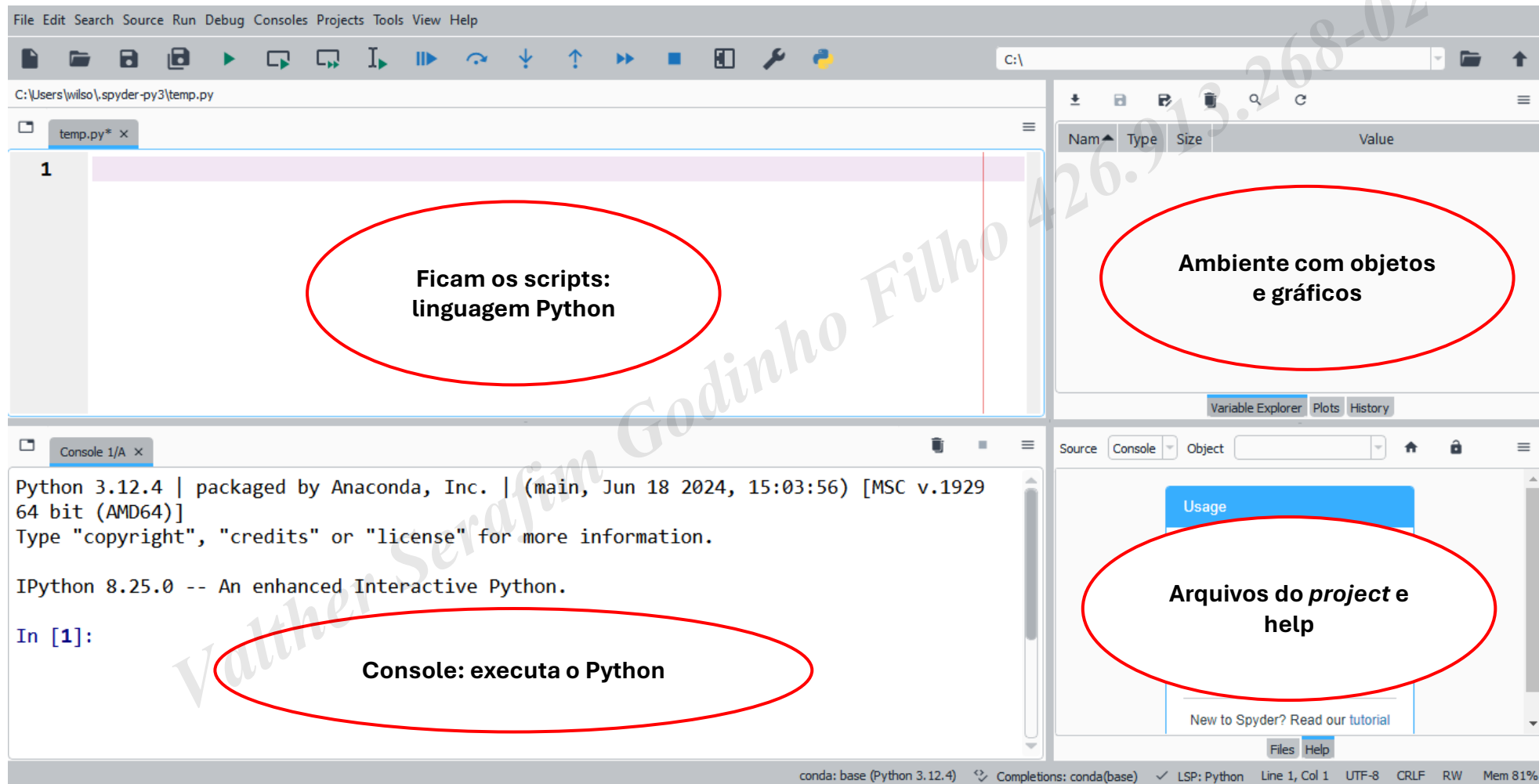
Valther Serafim Godinho Filho 426.913.268-02

Introdução ao Python

Apresentação

- **Python:** é a linguagem de programação que vamos utilizar
- Neste curso, vamos implementá-lo por meio do **Spyder (IDE)**
 - Software que torna o uso do Python mais simples para o usuário
- Para instalar o Spyder, vamos baixá-lo por meio da plataforma Anaconda
 - **Acessar o tutorial de instalação do Anaconda disponibilizado em conjunto com os materiais de aula!**

Ajuste: View > Window layouts > Rstudio



Acessando um *project*

- Sempre que acessar um projeto em execução, acesse o *project*:
 - Retire a pasta do arquivo compactado (caso esteja)
 - No Spyder acesse: Projects > Open Project > Selecione a pasta
 - Os arquivos contidos no *project* aparecerão no ambiente

Pacotes

- Alguns pacotes que utilizaremos:
 - Pandas: manipulação e análise de dados
 - Numpy: funções matemáticas e dados
 - Matplotlib: visualização de dados em gráficos
 - Seaborn: também é um pacote gráfico
 - Plotly: gráficos interativos

Documentação

- Leituras e documentação para consulta
 - <https://pandas.pydata.org/docs/index.html>
 - <https://numpy.org/doc/stable/>
 - <https://matplotlib.org/>
 - <https://seaborn.pydata.org/>
 - <https://plotly.com/python/>

MBAUSP ESALQ

Obrigado!

Wilson Tarantin Junior

[linkedin.com/in/wilson-tarantin-junior-359476190](https://www.linkedin.com/in/wilson-tarantin-junior-359476190)