

A series of overlapping geometric shapes, primarily triangles and quadrilaterals, in shades of teal and purple, located in the top-left corner of the slide.

AceLeraDev Data Science

Continuação Pensamento estatístico em Python

A series of overlapping geometric shapes, primarily triangles and quadrilaterals, in shades of teal and purple, located in the bottom-right corner of the slide.

0 problema da dimensionalidade

/

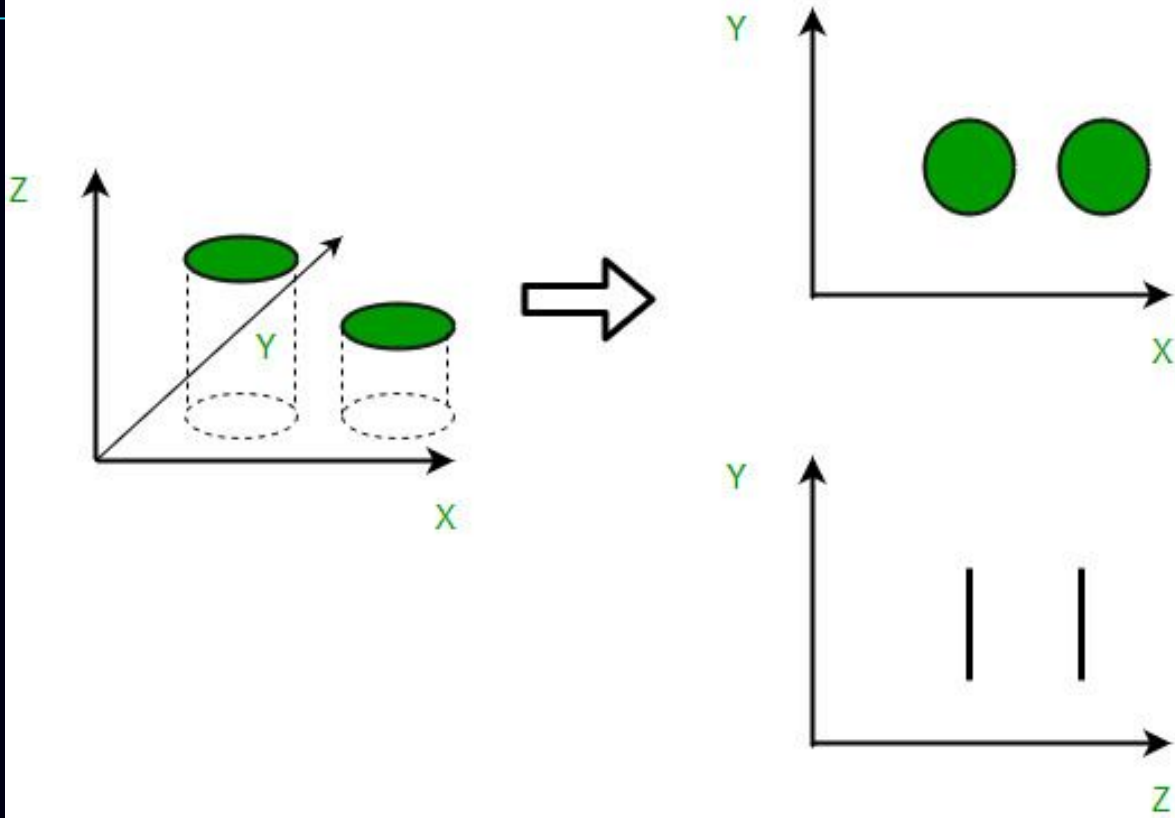
- 0 problema da dimensionalidade, também conhecido como curse of dimensionality e como comportamento de curva em U, é um fator muito relevante para decidir-se a dimensionalidade ideal a ser adotada em um problema de reconhecimento de padrões.
- [ref](#)

Redução de dimensionalidade

/

- Refere-se ao processo de conversão de um conjunto de dados que têm vastas dimensões em dados com dimensões menores;
- Assegurando que ele transmite informações semelhantes de forma concisa;
- [ref](#)
- [ref2](#)

Dimensionality Reduction



Seleção de Features

/

- O processo de seleção de um subconjunto de recursos relevantes (variáveis, preditores) para uso na construção do modelo.
- Simplificação de modelos para facilitar sua interpretação;
- Tempos de treinamento mais curtos,
- Evitar problemas de dimensionalidade,
- Generalização aprimorada reduzindo o overfitting [sobreajuste] (funciona só para o treino/teste inicial)

Feature Selection

Full Feature Set



Identify Useful Features



Selected Feature Set



The slide features decorative geometric lines in teal and purple. In the top-left corner, a teal line forms a partial hexagon, and a purple line extends diagonally downwards. In the bottom-right corner, a teal line forms another partial hexagon, and a purple line extends diagonally upwards. These lines frame the central text area.

Extração de Features

/Extração de Features

- Diminuir o número e recursos para explicar seus dados;
- Interpretar o comportamento de cada features
- Parte do processo de redução de dimensionalidade



AceLeraDev Data Science

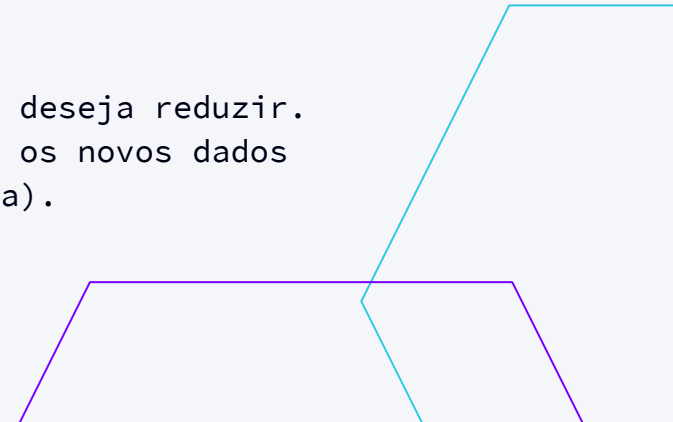
Continuação Pensamento estatístico em Python

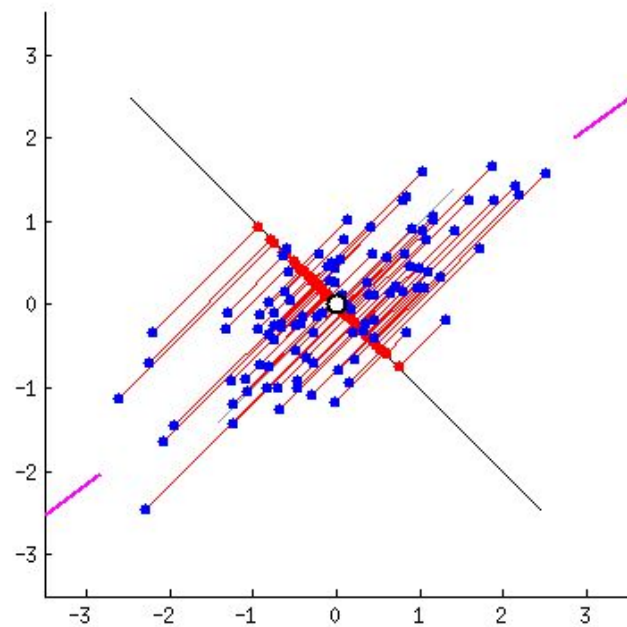


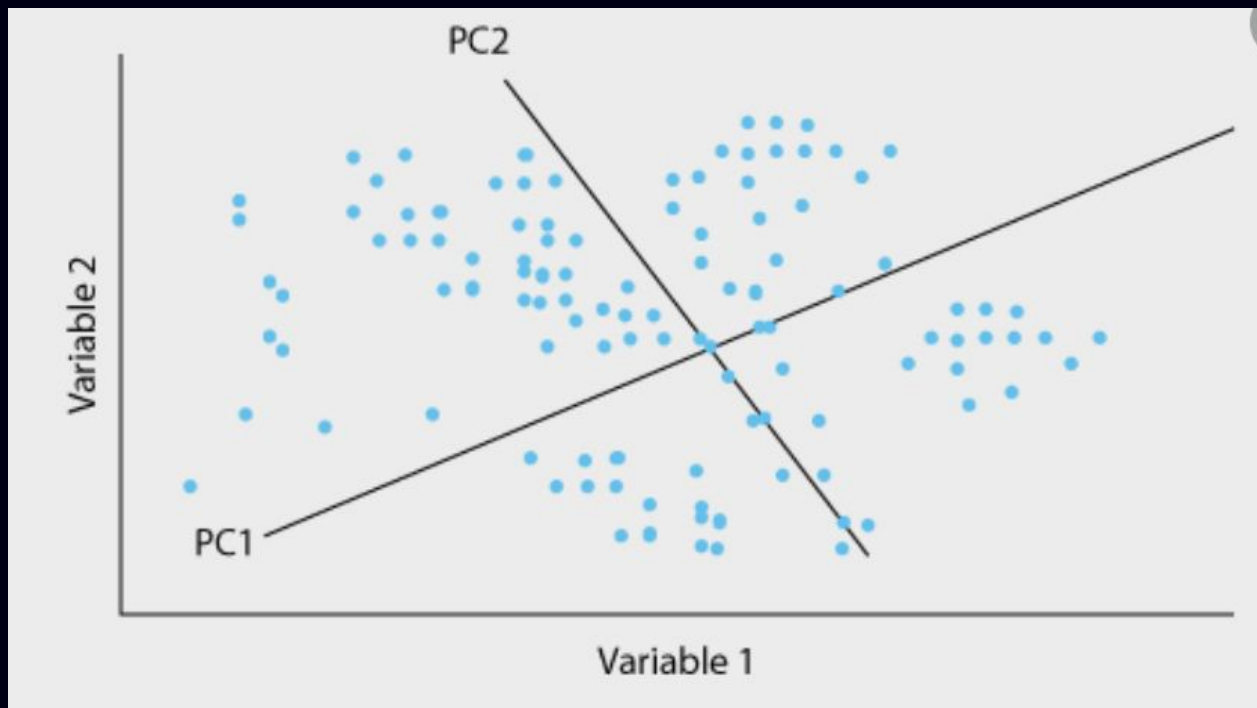


Análise de Componentes Principais – PCA

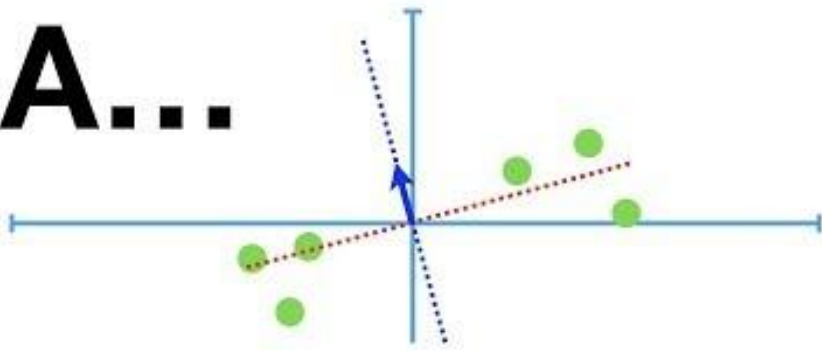
/

- Ferramenta matemática para reduzir a dimensionalidade dos dados.
 - Bastante útil quando tratamos problema com altas dimensões.
 - Identifica a relação entre as características extraídas de dados;
 - Exemplo de uso:
 - Escolhe-se um número de variáveis que se deseja reduzir.
 - Escolhe-se o 'quão fidedigno' deseja que os novos dados sejam aos anteriores (variância explicada).
- 





PCA...



Step-by-Step!!!



AceLeraDev Data Science

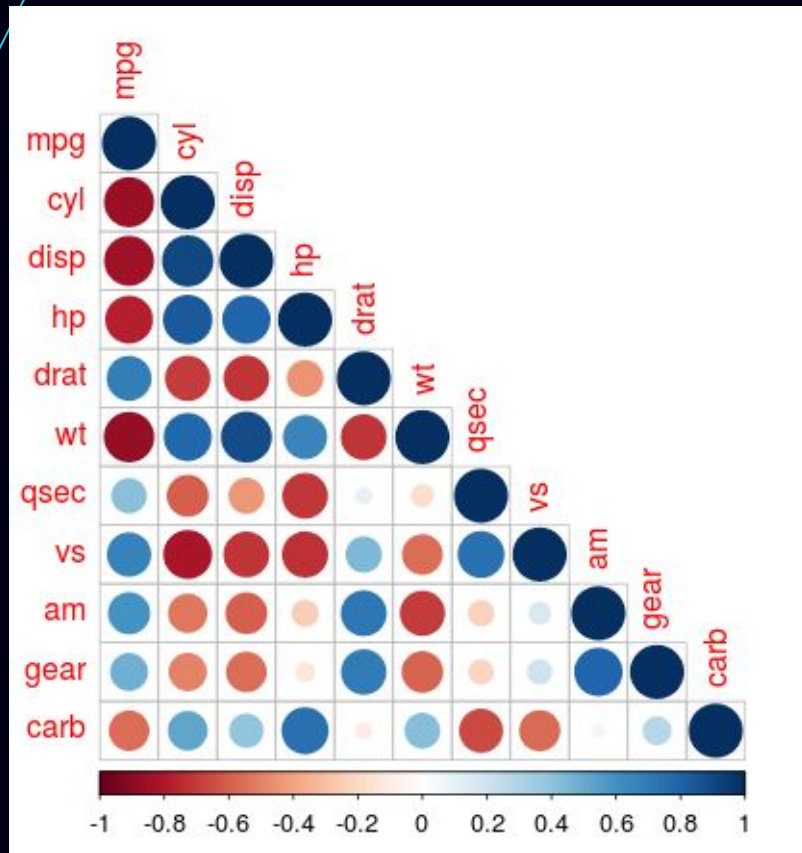
Continuação Pensamento estatístico em Python



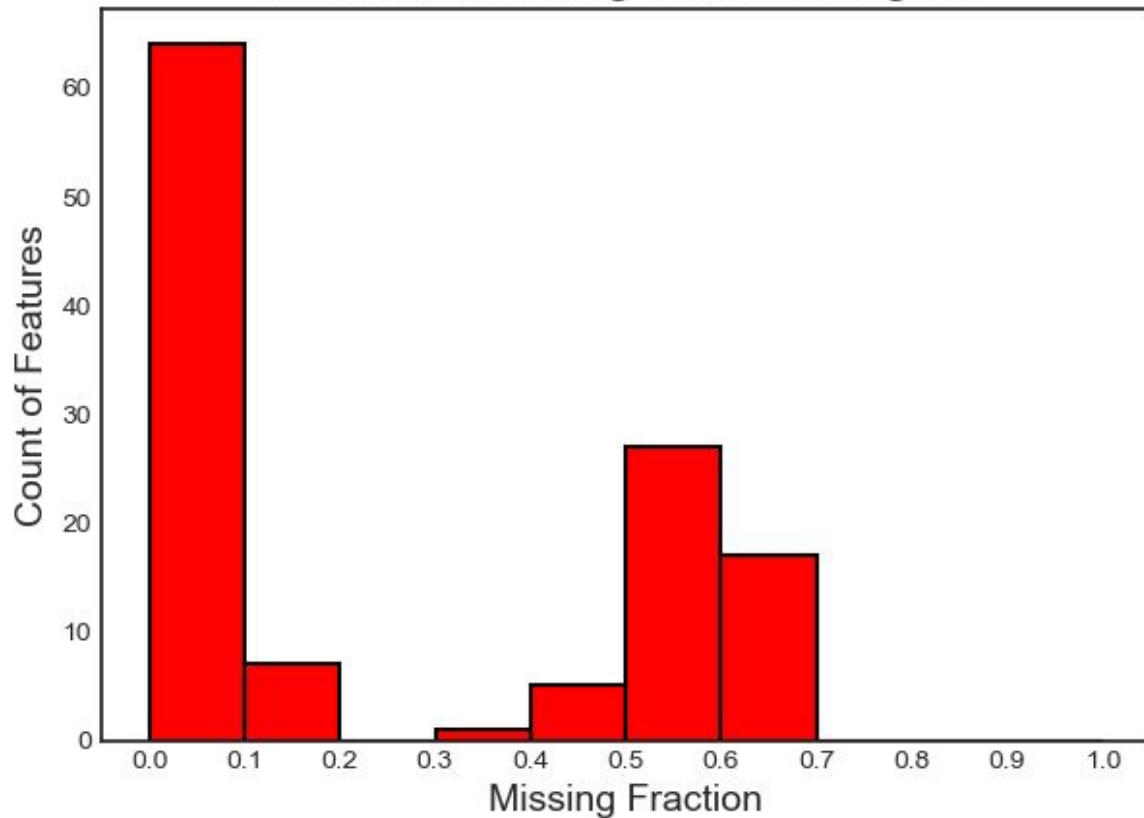
Seleção de Variáveis

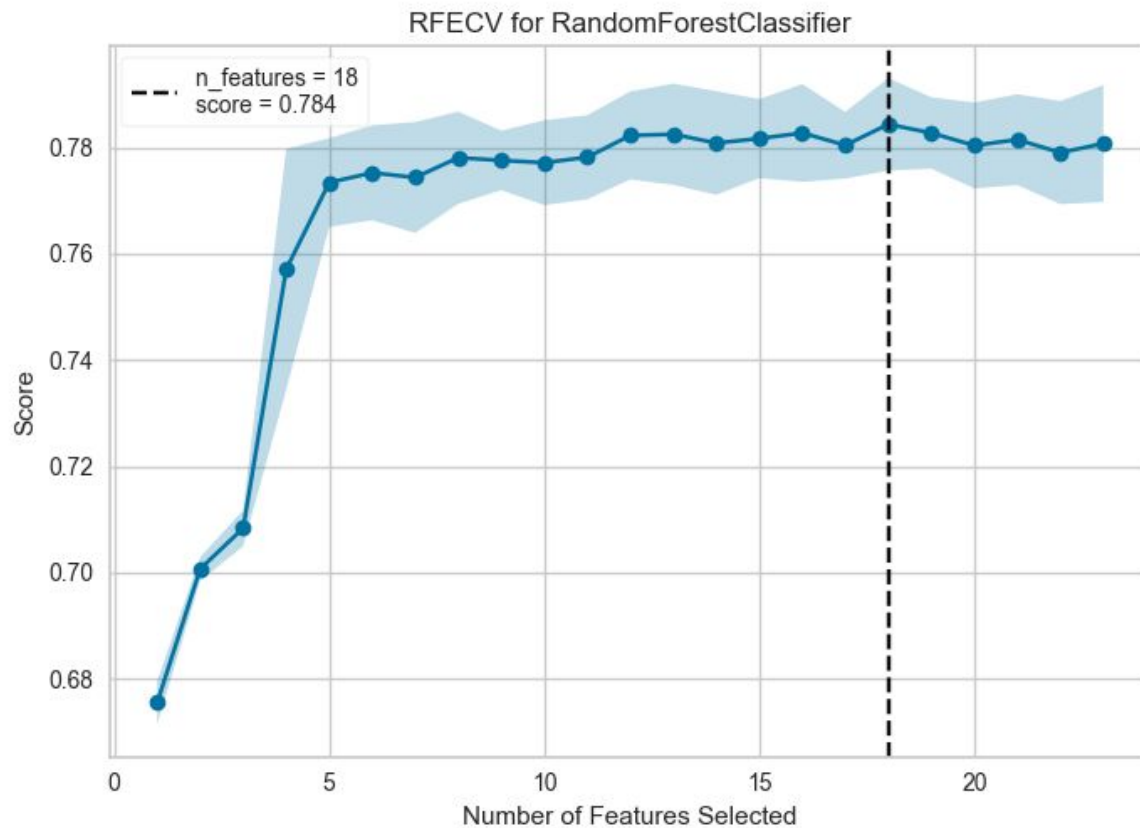
/

- Seleção por correlação, completude e variância;
- Recursive Feature Elimination;
- Algoritmo de Minimum Redundancy Maximum Relevance (mRMR);



Fraction of Missing Values Histogram





Seleção de Variáveis

/

<https://skymind.ai/wiki/eigenvector>

<https://www.vooo.pro/insights/guia-de-aprendizado-para-iniciantes-em-tecnicas-de-educacao-de-dimensoes-dimension-reduction/>

<http://www2.ic.uff.br/~aconci/PCA-ACP.pdf>

https://sebastianraschka.com/Articles/2014_pca_step_by_step.html

<https://towardsdatascience.com/a-feature-selection-tool-for-machine-learning-in-python-b64dd23710f0>

<https://towardsdatascience.com/why-how-and-when-to-apply-feature-selection-e9c69adf2fabf2>

<https://machinelearningmastery.com/feature-selection-in-python-with-scikit-learn/>

<https://medium.com/@aptrishu/understanding-principle-component-analysis-e32be0253ef0>