

# PCA Clássico

thauan.goncalves

November 2018

## 1 Desenvolvimento

Neste exercício foi implementado a técnica *principal components analysis*(PCA) e testado sobre o dataset *iris*. A função que realiza o PCA recebe um parâmetro que indica o quanto de informação a respeito do dataset original deseja-se manter após realizarmos o PCA; isso se dá utilizando as propriedades do PCA que leva em consideração os autovalores da matriz de covariância dos atributos exploratórios do dataset.

A Figura 1 mostra a distribuição dos atributos antes e depois da aplicação do PCA. Como se pode notar, após a aplicação do PCA com uma taxa de diminuição dos dados com o limite de 25%, o dataset diminui em duas dimensões. Para comparar o desempenho de utilizar esta taxa de PCA para o caso iris foi utilizado o algoritmo *knn*, com k valendo 5, e uma taxa de 75% para treino e 25% para teste. A acurácia para ambos os casos foi de aproximadamente 1, como pode ser verificado executando o código *pca.py* que imprime os gráfico representados pelas figuras e também mostra a acurácia para os dois casos.

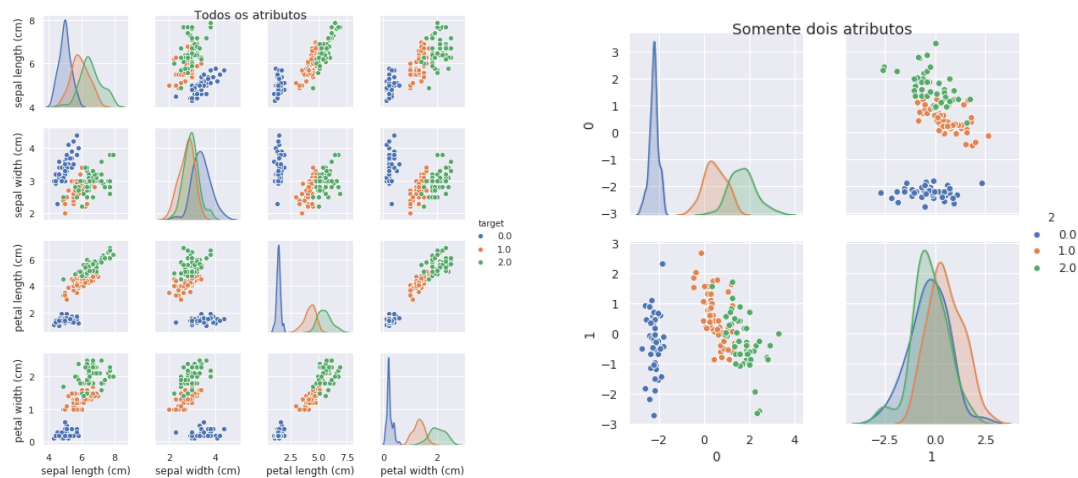


Figura 1: A figura da esquerda mostra a distribuição dos atributos levando em consideração as classes que o iris possui. A figura da direita mostra a mesma distribuição dos dados após a aplicação do PCA. Vale notar que os atributos na figura da direita possuem os seus dados normalizados em torno da média 0. Além disso, os seus atributos não possuem um nome que os indique pois eles são calculados em função dos atributos antigos.