

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO

Marcos Cesar Ribeiro de Camargo - 9278045
Rafael Augusto Monteiro - 9293095

Exercício 5 - PCA

1. Introdução

O Conjunto de dados Iris contém informações sobre flores e suas espécies. Neste projeto, utilizamos PCA para facilitar o agrupamento de flores pelos atributos presentes no dataset.

2. Métodos e Procedimentos

O notebook “PCA.ipynb” contém o código fonte e todos os procedimentos utilizados na elaboração do projeto.

Inicialmente, o dataset Iris é carregado. Em seguida, todos os valores são centralizados pela média e normalizados pelo desvio padrão. A seguir, foi calculada a variância e a soma das variâncias de cada atributo, como descrito na tabela abaixo:

	sepal_length	sepal_width	petal_length	petal_width
Variância Acumulada	0.73	0.96	0.99	1

Tabela 1: variância acumulada

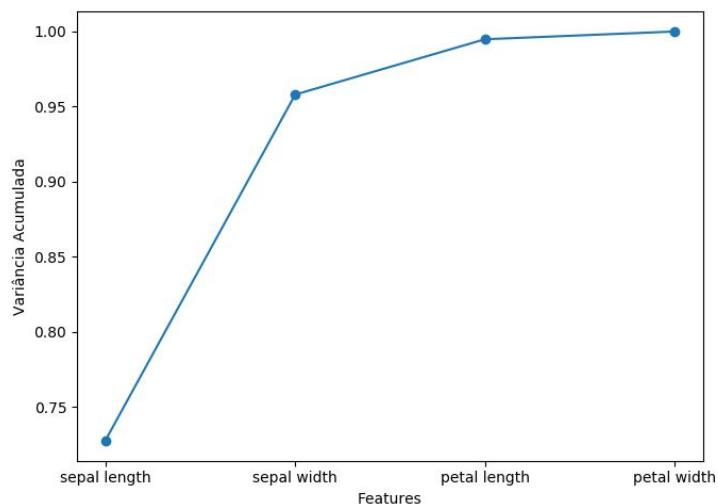


Imagem 1: variância acumulada. Note que o eixo Y começa em 0.73

Em seguida, foi realizado o PCA. Os resultados são apresentados a seguir:

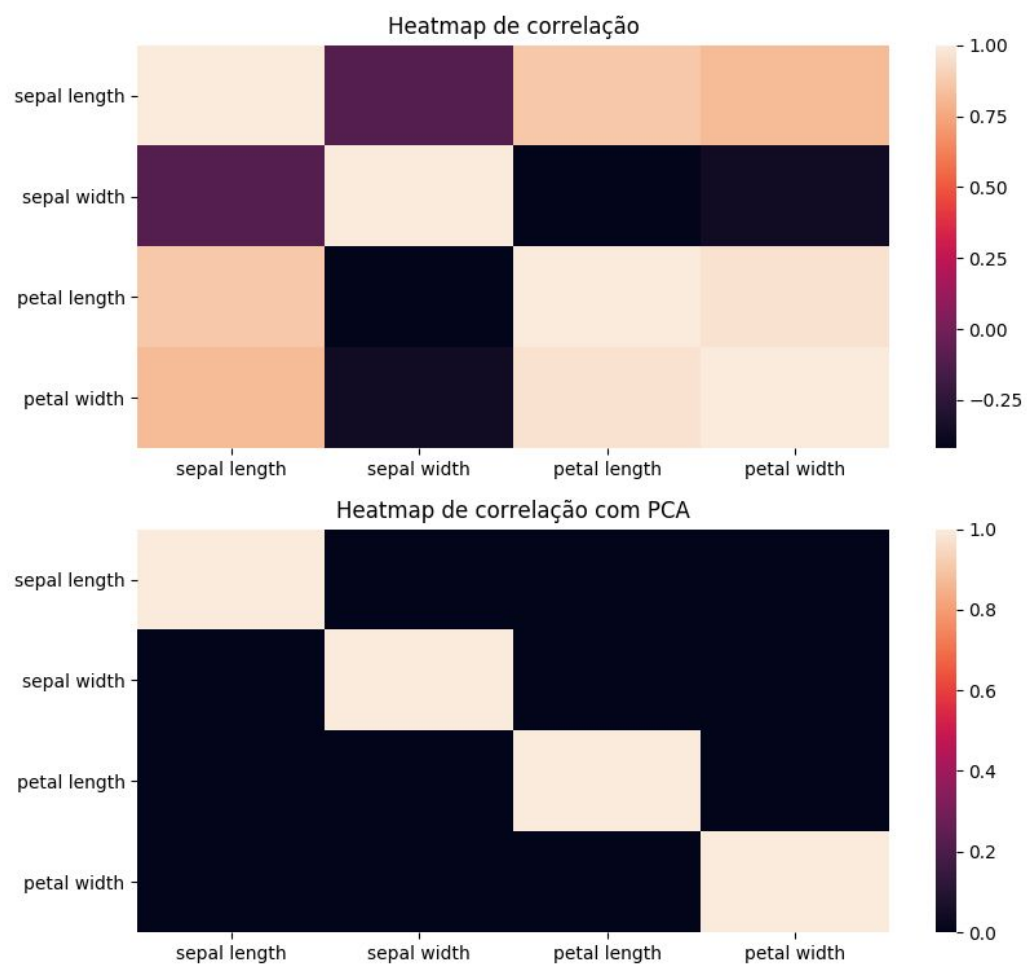
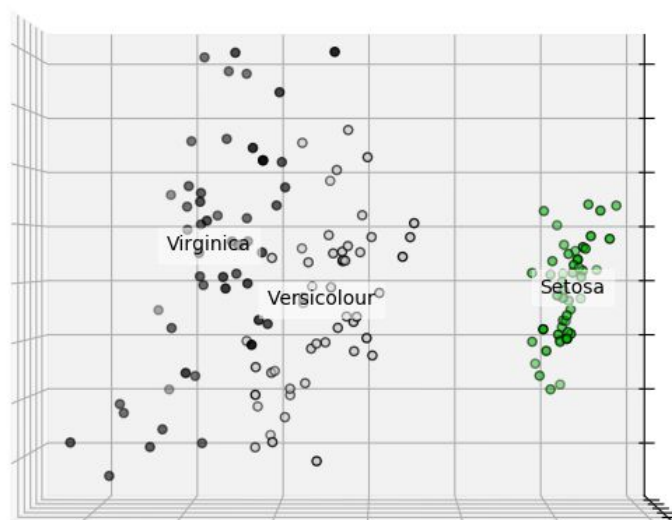


Figura 2: Matriz de correlação. É possível observar que o PCA reduziu a correlação entre os atributos.

Visualização 3D dos dados com PCA



Visualização 3D dos dados sem PCA

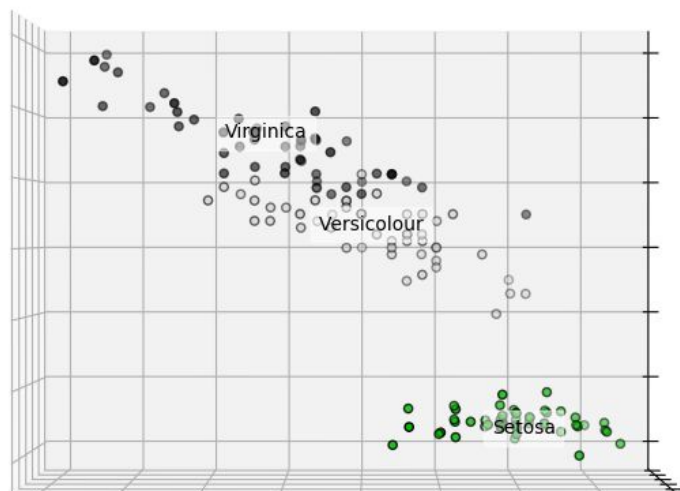


Figura 3: Projeção 3D dos dados. É possível observar que o PCA aumentou a diferenciação entre as classes