

ICMC - USP

SCC0270 - Redes Neurais

Exercício 5

PCA

Docente: Roseli Aparecida Francelin Romero

Aluno: Lucas Antognoni de Castro, 8936951

São Carlos

Novembro 2017

Enunciado

Implementar e testar a técnica ***PCA*** (*Principal Components Analysis*) com o conjunto de dados ***Iris Plants Database***.

Plataforma utilizada

O referido trabalho foi confeccionado no sistema operacional ***Linux*** (*Arch*), utilizando um ambiente de desenvolvimento com a ***IDE Visual Studio Code*** e terminal. O código foi desenvolvido na linguagem de programação ***Python*** e suas bibliotecas.

Dados

A base de dados utilizada é a ***Iris Plants Database***. Esta base contém 150 instâncias, com 50 para cada uma das três classes possíveis, a saber:

- Iris Setosa
- Iris Versicolour
- Iris Virginica

Para cada instância são apresentados 4 atributos:

- Comprimento da sépala
- Largura da sépala

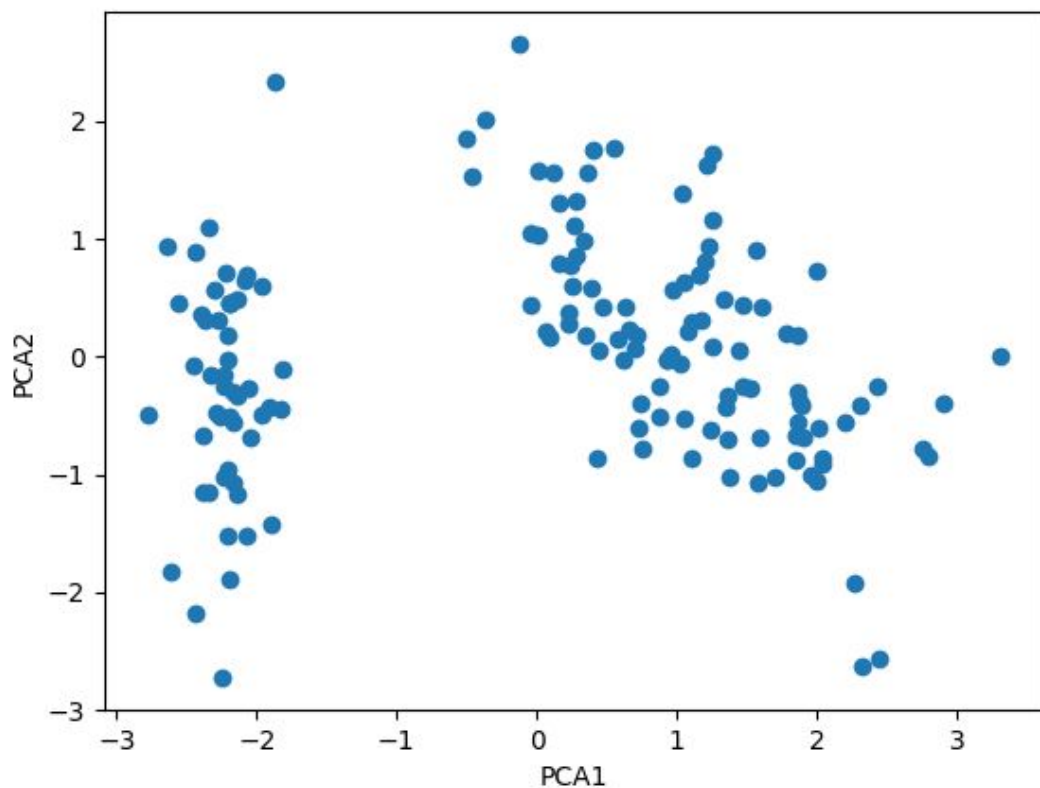
- Comprimento da pétala
- Largura da pétala

Esta base de dados está disponível em:

- <https://archive.ics.uci.edu/ml/datasets/iris>

Testes e Resultados

Realizou-se a projeção da base de dados em um subespaço de tamanho 2, o que gerou o seguinte gráfico:



Analisando os dados podemos definir intervalos aproximados para cada classe:

- Iris Setosa

- PCA1: [-2.77, -1.83]
 - PCA2: [-2.75, 2.4]
- Iris Versicolour
 - PCA1: [-0.5, 1.35]
 - PCA2: [-0.52, 2.65]
- Iris Virginica
 - PCA1: [0.87, 3.3]
 - PCA2: [-2.62, 1.32]

Referências

[1] <https://plot.ly/ipython-notebooks/principal-component-analysis/>

[2] *Dimensionality Reduction: Principal Components Analysis*,

<https://www.youtube.com/watch?v=ZqXnPcylAL8>

<https://www.youtube.com/watch?v=NUn6WeFM5cM>

<https://www.youtube.com/watch?v=7WqWoEKUdQY>

[3] Shlens, Jon. *A Tutorial on Principal Component Analysis: Derivation, Discussion and Singular Value Decomposition*. March 2003