

INSTITUTO DE CIÊNCIA E TECNOLOGIA

Redes Neurais

Prof. Marcos Quiles

Análise de comportamento do procedimento:

PCA – Principal Component Analysis

Tamires Beatriz da Silva Lucena - 111866

Dezembro/2018

A Rede

Deseja-se com a implementação da rede PCA (Principal Component Analysis) reduzir a dimensionalidade de um conjunto de dados com vários atributos, a fim capacitar possíveis análises visuais de tais conjuntos.

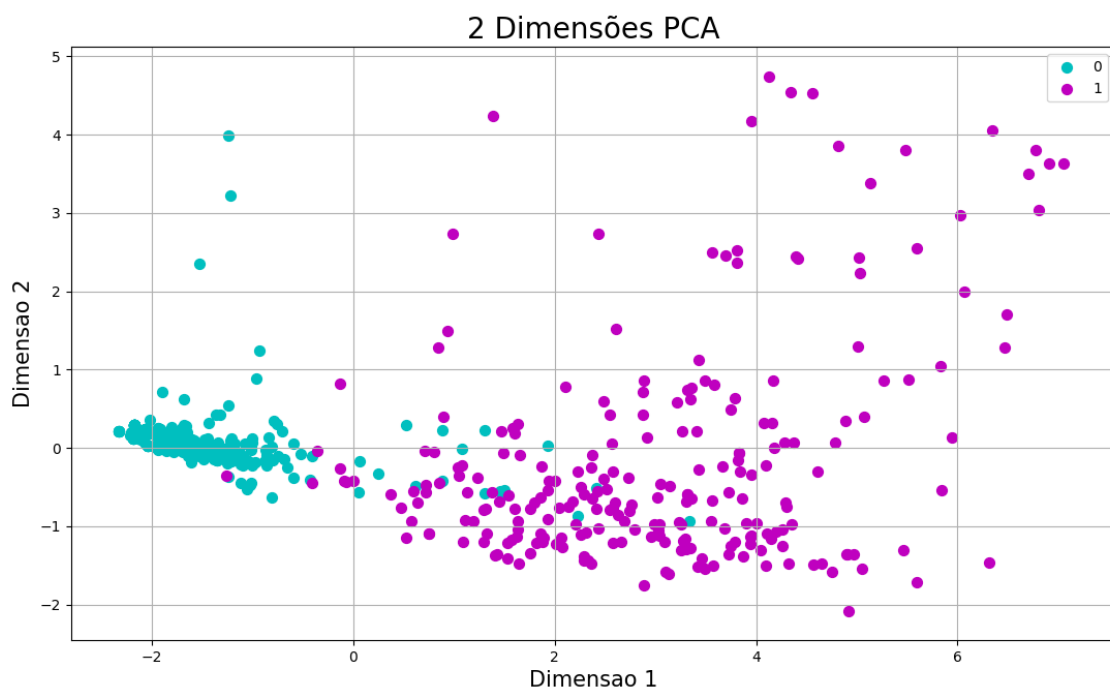
Conjuntos Analisados

Breast Cancer

O primeiro dataset escolhido pra ser analisado é o *breast cancer* que trata de informações sobre pessoas com câncer de mama, classificando o tipo de câncer em M- Maligno e B- Benigno. Os atributos são características celulares retiradas de imagens da massa mamária, como por exemplo, tamanho da célula e uniformidade da forma da célula. O *dataset* conta com 9 atributos.

Tendo realizada a redução de dimensionalidade para vários numeros (inclusive, reduzindo para 2 dimensões como demonstrado na figura abaixo), obteve-se melhor resultado com a redução para 7 dimensões, onde foi preservada 96% da variancia original.

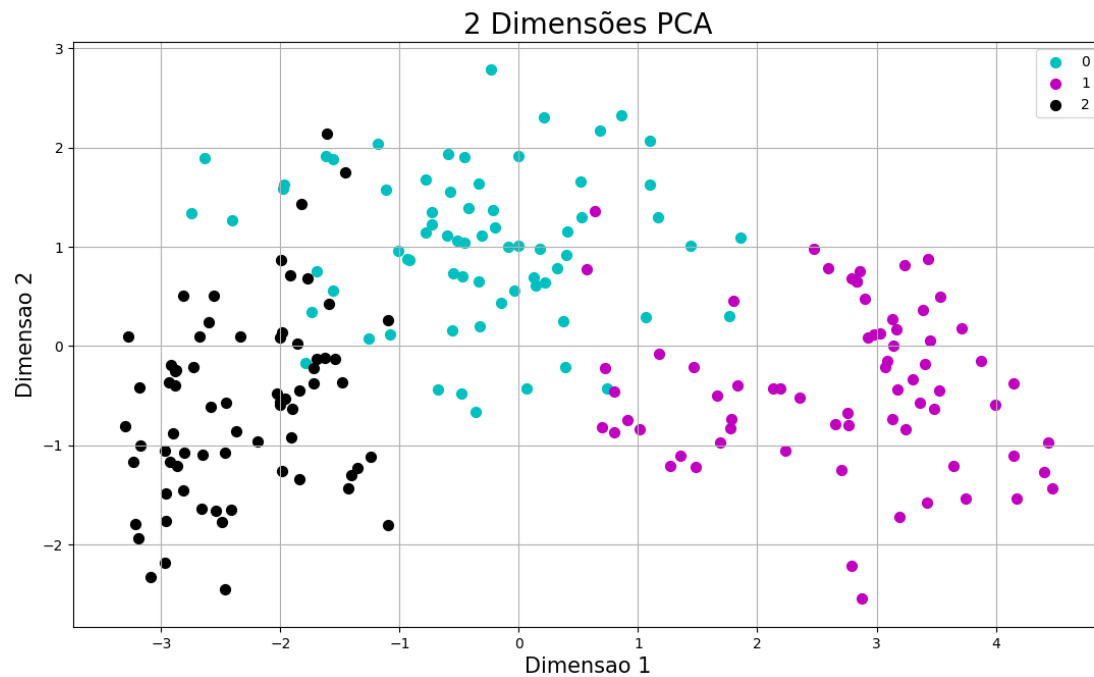
É importante citar que mesmo reduzindo a 2 dimensões, a rede apresentou um resultado satisfatório onde manteve 74% da variancia original.



Seeds

O conjunto diz respeito a grãos de trigo, sendo três, os tipos de classificação. São eles: Kama, Rosa e Canadense. Para construir o *dataset* foram medidos 7 parâmetros geométricos, entre eles: área, perímetro e coeficiente de simetria, por exemplo.

Feitos os testes de redução, logo obteve-se bom resultado, onde reduzindo para 3 dimensões, a variancia foi mantida em 99% da original. A figura abaixo ilustra a redução para 2 dimensões, que também traduz bem o conjunto, mantendo 89% da variancia original.



Glass

A análise do conjunto de dados *glass* objetiva identificar o tipo de vidro, tendo como motivação a investigação criminal, visto que este material, quando encontrado em locais em que houveram situações suspeitas, pode ser considerado como algum indício. O *dataset* conta com 9 atributos iniciais.

Este conjunto não pode ser bem representado por apenas 2 dimensões – representadas na figura abaixo - visto que, foi mantido apenas 51% da variância original. Entretanto, reduzindo para 6 dimensões, pode-se obter boa representação tendo 95% da variância mantida.

