Universidade Federal Rural de Pernambuco

Unidade Acadêmica de Garanhuns

Bacharelado em Ciência da Computação

Reconhecimento de Padrões II 2016.2

Atividade 1

Prof. Tiago de Carvalho

Aluno

João Antônio Chagas Nunes

1. Introdução

O objetivo dessa atividade foi conhecer a base *Iris,* aprender o que é o *PCA*, para que serve e como usa-lo. Para isso, foram gerados alguns gráficos da base, com e sem aplicar o *PCA*, para observar como seus dados sem comportam. Também foi calculado a taxa de acerto para a base com e sem o uso do *PCA*. Todas as classificações foram feitas com o classificador *K*-NN.

2. Metodologia

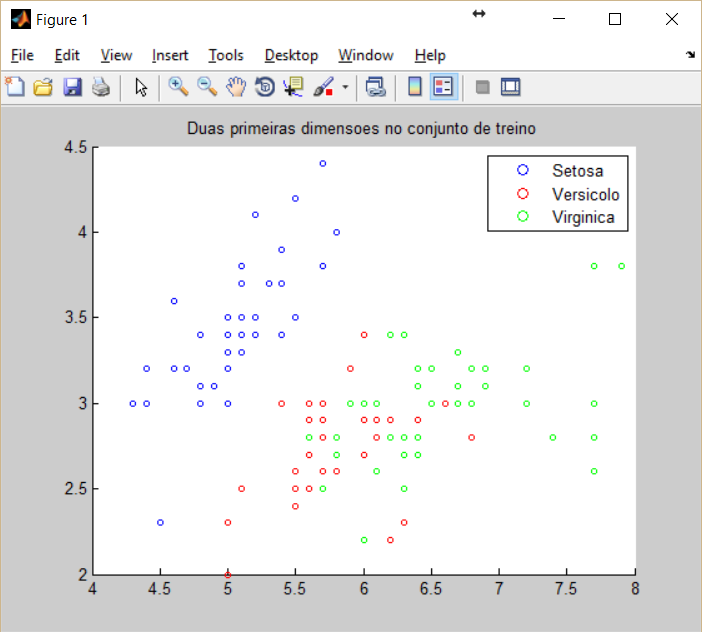
Primeiramente a base *Iris* foi separada em conjuto de treino e de teste. Os 10 primeiros exemplos de cada classe foram para o conjunto de teste, o restante foi para o conjunto de treino.

O classificador 1-NN foi usado para calcular a taxa de acerto da base utilizando todas suas 4 dimensões. Também foi calculado uma taxa de acerto para a base utilizando apenas a primeira e segunda dimensão. Junto com a taxa de acerto, foi plotado um gráfico do conjunto de treino e do conjunto de teste dessas duas dimensões.

Depois calculamos um *PCA* para a base *Iris* a partir de seu conjunto de treino. Projetamos o conjunto de treino e teste da base com o *PCA* calculado. Plotamos um gráfico para o conjunto de treino e outro para o conjunto de teste utilizando os dois componentes principais do *PCA* calculado e calculamos uma nova taxa de acerto para base utilizando apenas esses dois componentes principais.

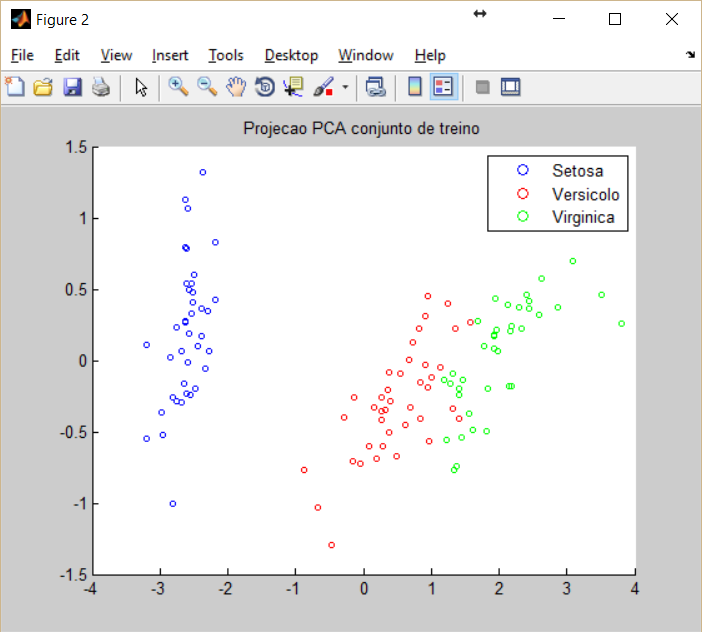
3. Resultados

O resultado do uso do *PCA* foi satisfatório, utilizando apenas duas dimensões calculadas como *PCA* conseguimos uma taxa de acerto de 96,67% que foi igual a taxa de acerto original da base, a taxa de acerto sem o uso de *PCA* e utilizando todas as 4 dimensões da base.



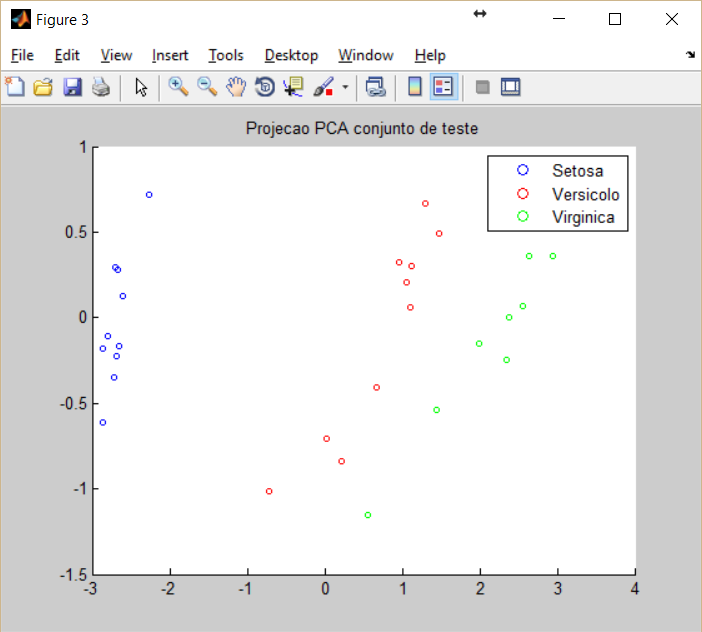
*Imagem 1* – *Gráfico do conjunto de treino sem PCA.*

Na *Imagem 1* podemos ver o gráfico das duas primeiras dimensões do conjunto de treino da base *Iris* sem o uso do PCA. A taxa de acerto obtida utilizando apenas essas duas dimensões e sem o uso do *PCA*  foi de 70%. O que nos da um resultado 26,67% pior do que o obtido usando todas as 4 dimensões.



*Imagem 2 – Gráfico conjunto de treino com PCA*

Na *Imagem 2* temos o gráfico da projeção dos dois componentes principais do conjunto de treino da base *Iris*. Notamos que os dados já estão bem mais separados comparado a o gráfico da imagem 1.



*Imagem 3 - Gráfico conjunto de teste com PCA*

Na *Imagem 3* temos o gráfico da projeção dos dois componentes principais do conjunto de teste da base *Iris*. Notamos que os dados do conjunto de teste também estão estão bem mais separados comparado a o gráfico da imagem 1. Calculamos também a taxa de acerto nesse conjunto que foi de 96,67%. Assim percebemos que podemos conseguir a mesma taxa de acerto na base *Iris,* utilizando apenas metade de suas dimensões e usando *PCA*.

4. Conclusão

Conclui-se que o PCA é um método muito bom para extração de características. Ele consegue separar muito bem os dados e diminuir muito a quantidade de dimensões necessárias para conseguir-se um bom resultado.