

TA04 - Classificação

Alunos: Erick Eckermann Cardoso GRR20186075
Anderson Carmo Frasão GRR20204069

Professor: Eduardo Todt

Link do repositório com o código: <https://github.com/Carmofrasao/trab-4-visao>

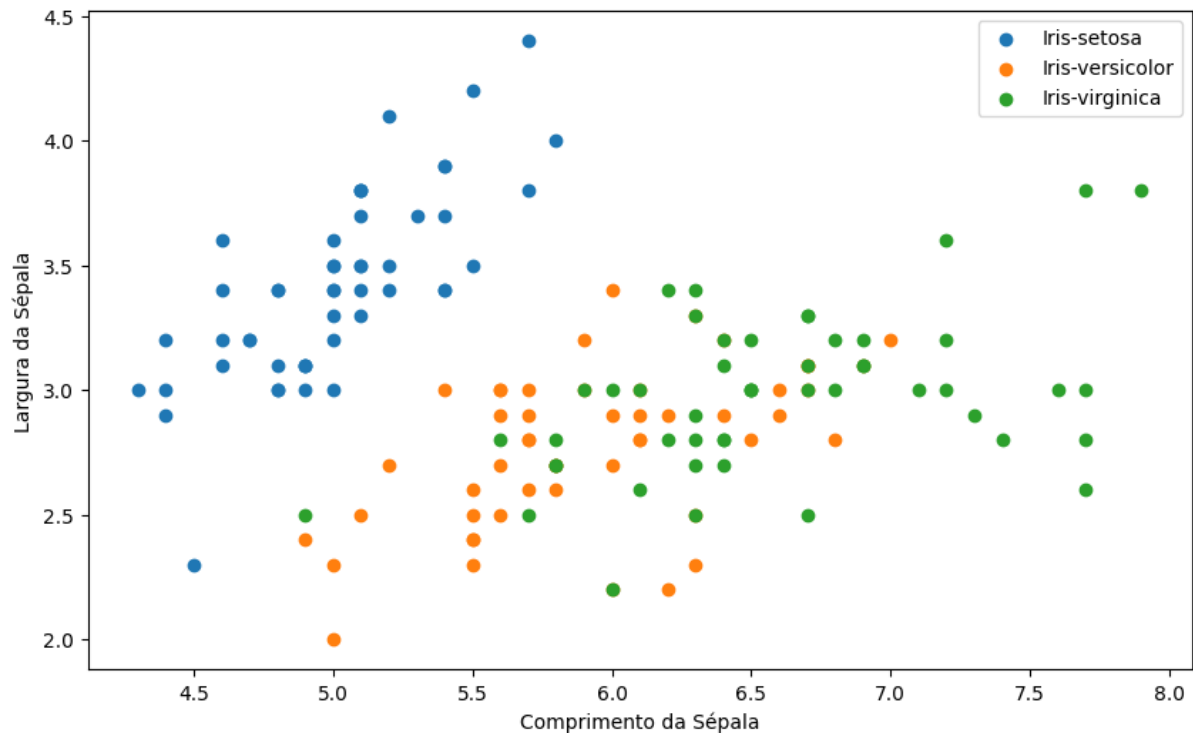
Explicação

Experimentos de classificação com o dataset iris, encontrado em <https://archive.ics.uci.edu/ml/datasets/iris>. Para realizar esses experimentos, foi criado o código `analise.py`, onde seus principais resultados são::

1. **Análise estatística básica:** O código calcula a média, desvio padrão e moda para cada medida (comprimento e largura da sépala e da pétala) em todo o conjunto de dados e por categoria (espécie de flor). Isso fornece uma visão geral das tendências e características das medidas.
2. **Frequência das categorias:** O código calcula a contagem de ocorrências de cada categoria (espécie de flor) no conjunto de dados. Isso permite visualizar a distribuição das espécies de flores e identificar possíveis desequilíbrios nos dados.
3. **Gráficos de dispersão:** O código gera gráficos de dispersão para visualizar a relação entre o comprimento e a largura da sépala e da pétala, separados por categoria (espécie de flor). Isso ajuda a identificar padrões e diferenças entre as espécies de flores com base nas medidas.
4. **Box-plot:** O código gera Box-plots para visualizar a distribuição das medidas (comprimento e largura da sépala e da pétala) por categoria (espécie de flor). Esses gráficos mostram a variação e a presença de outliers nas medidas, permitindo uma comparação mais detalhada entre as espécies.
5. **PCA:** O código aplica a técnica de PCA para reduzir a dimensionalidade dos dados para duas componentes principais. Isso ajuda a visualizar os dados em um espaço bidimensional e identificar padrões e agrupamentos das espécies de flores.
6. **Classificação e acurácia:** O código divide o conjunto de dados em conjuntos de treinamento e teste, treina um classificador K-NN e avalia a acurácia do modelo. A acurácia mede a precisão do modelo em prever corretamente a espécie de flor com base nas medidas da sépala e da pétala.

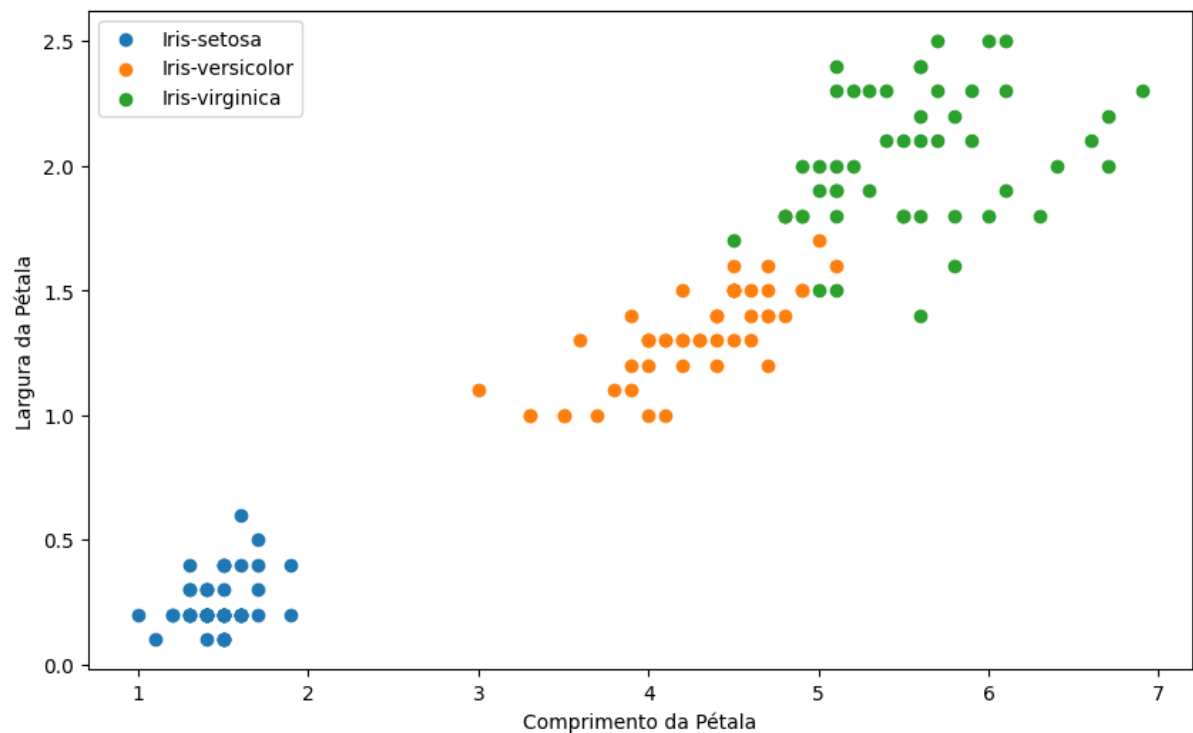
Resultados

Gráfico da relação entre comprimento e largura da sépala em cada categoria

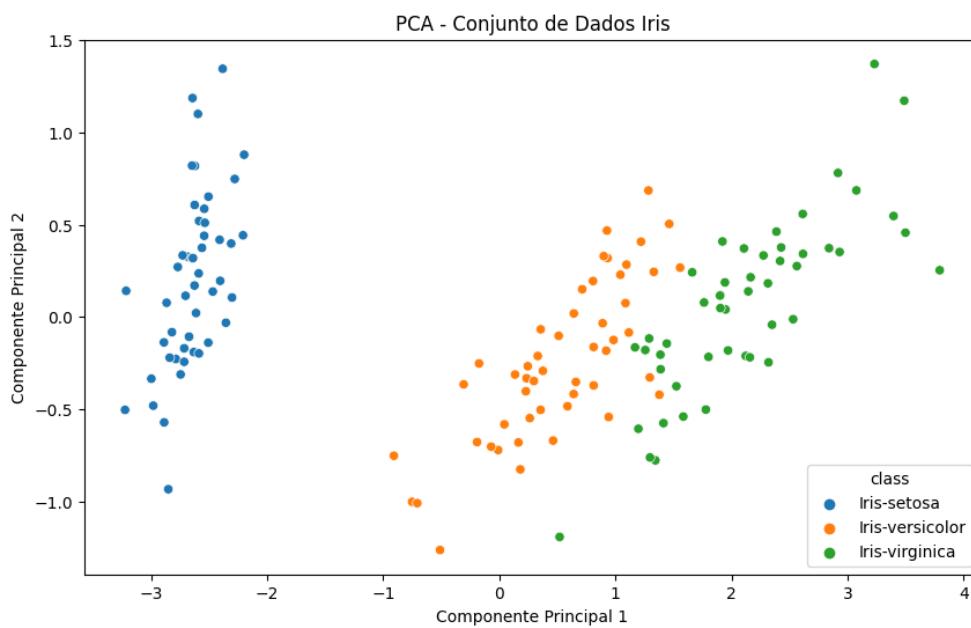
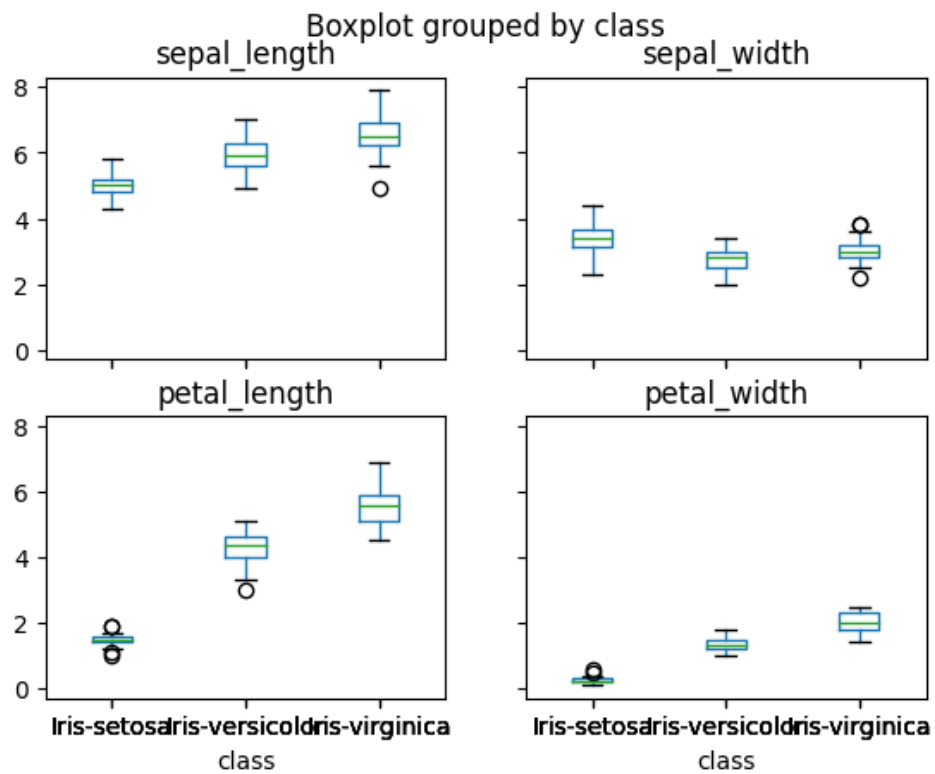


Pode-se perceber que a as relações para as sépalas da iris-setosa se mantêm unidas e bem distintas da iris-versicolor e virginiana, que se misturam por sua vez. Mesmo assim continuam separáveis.

Gráfico da relação entre comprimento e largura da Pétala em cada categoria



Já analisando as relações da pétala, as três categorias ficam bem mais separáveis.



Resultados da classificação

Dimensões do conjunto de treinamento: (120, 4) (120,)

Dimensões do conjunto de teste: (30, 4) (30,)

Acurácia do modelo K-NN: 1.0

Percebe-se que as relações entre as classes são bem separáveis, portanto a classificação obteve um resultado extremamente satisfatório.

Outros resultados

Média geral:

```
sepal_length  5.843333
sepal_width   3.054000
petal_length   3.758667
petal_width    1.198667
```

Média por categoria:

```
      sepal_length sepal_width petal_length petal_width
class
Iris-setosa      5.006      3.418      1.464      0.244
Iris-versicolor  5.936      2.770      4.260      1.326
Iris-virginica   6.588      2.974      5.552      2.026
```

Desvio padrão geral:

```
sepal_length  0.828066
sepal_width   0.433594
petal_length   1.764420
petal_width    0.763161
```

Desvio padrão por categoria:

```
      sepal_length sepal_width petal_length petal_width
class
Iris-setosa      0.352490  0.381024  0.173511  0.107210
Iris-versicolor  0.516171  0.313798  0.469911  0.197753
Iris-virginica   0.635880  0.322497  0.551895  0.274650
```

Moda geral:

```
sepal_length    5.0
sepal_width     3.0
petal_length     1.5
petal_width     0.2
class      Iris-setosa
```

Moda por categoria:

```
0      sepal_length sepal_width petal_length petal_width \
class
Iris-setosa      5.0      3.4      1.5      0.2
Iris-versicolor  5.5      3.0      4.5      1.3
Iris-virginica   6.3      3.0      5.1      1.8
```

Frequência das categorias:

```
Iris-setosa    50
Iris-versicolor 50
Iris-virginica 50
```