



Faculdades Integradas de Bauru – FIB
Bacharelado em Ciências da Computação

Ciência de Dados I

Atividade Avaliada 2 – 1º Bimestre

Lista de Exercícios

FIB
2021

Prof. Ms. André Farina

Metodologia

- Desenvolva seu trabalho **individualmente**.
- Use exclusivamente os cálculos de estatística e álgebra linear que desenvolvemos em aula, escreva 2 módulos Python para isso
- Postar o trabalho no Teams até às 23:59h do dia 11/04.
- Esta atividade terá peso 1,0 na média da disciplina.

Descrição

Para o desenvolvimento desta atividade vamos adotar o **dataset Seeds** que pode ser obtido do repositório da UCI (<https://archive.ics.uci.edu/ml/datasets/seeds>) ou no Teams, no formato textual (.txt). Esse dataset é composto por **210 instâncias** pertencentes a três classes, **70 instâncias** cada, correspondendo respectivamente a cada uma das variedades de trigo: **Kama**, **Rosa** e **Canadian**. A tabela abaixo mostra os atributos de cada classe que correspondem às características do grão de trigo, totalizando sete atributos do tipo “real” e mais o campo enumerativo correspondente à classe (variedade do trigo).

Atributos de cada instância:

1. Área
2. Perímetro
3. Compacidade (ou compactação),
4. Comprimento do grão
5. Largura do grão
6. Coeficiente de assimetria
7. Comprimento do sulco do grão

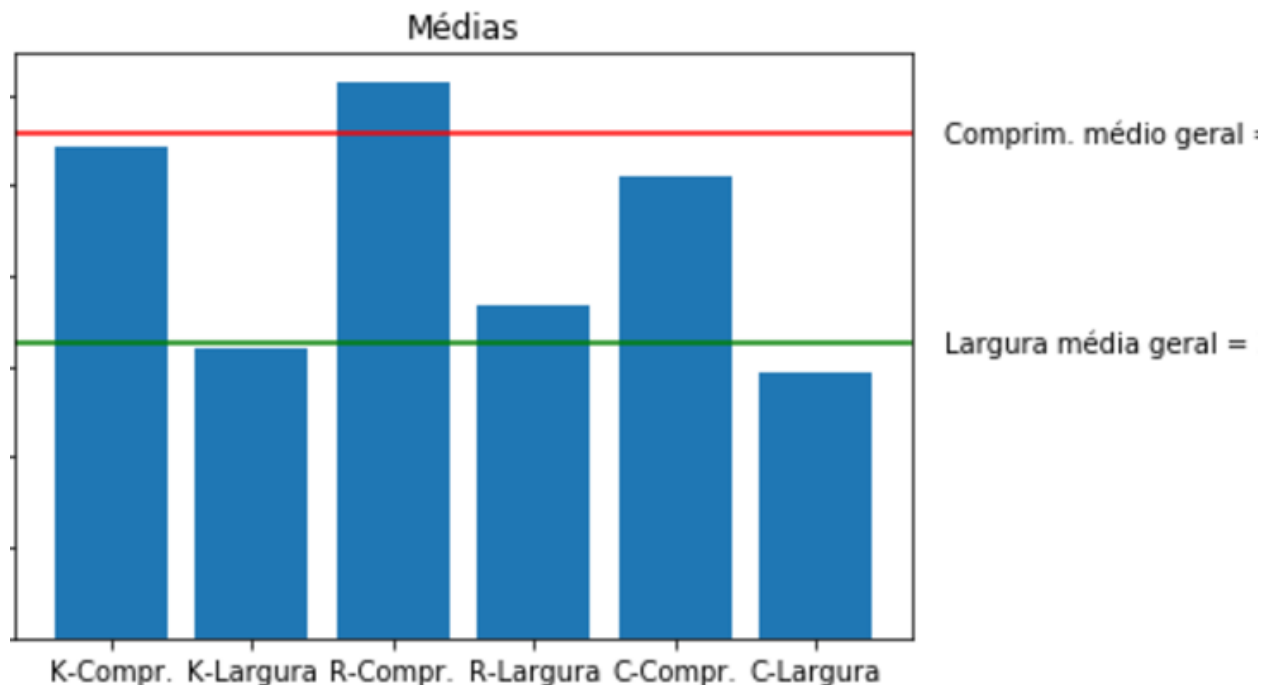
Considere as instruções a seguir para desenvolver o seu trabalho.

Instruções

- Baixe o dataset Seeds do site da UCI ou do Teams
- Abra o Jupyter Notebook e crie um novo ambiente Python 3 (new -> Python 3)
- Importe a API Numpy
 - `import numpy as np`
- Carregue o dataset desconsiderando as colunas de 1 a 7
 - `data = np.genfromtxt('seed-dataset.txt', delimiter=',', usecols=(0, 1, 2, 3, 4, 5, 6))`

Questões:

1. Escreva dois módulos Python, um para conter os cálculos que aprendemos na aula de estatística e outro para os cálculos da aula de álgebra linear.
2. Imprima a moda e a mediana do comprimento e da largura de todas as instâncias.
3. Imprima o desvio padrão dos comprimentos de todos os grãos e explique o resultado.
4. Imprima o desvio padrão das larguras de todos os grãos e explique o resultado.
5. Imprima as correlações entre o comprimento e a largura de cada espécie e explique cada resultado.
6. Gere um gráfico de barras que mostre o comprimento médio e a largura média de cada espécie, com as médias gerais expressas em linhas horizontais, como no exemplo abaixo:



Obs.: Use os comandos abaixo para gerar as linhas horizontais:

```
plt.axhline(valor, color="cor") #Gera a linha  
plt.text(pos_x, pos_y, texto, va='center', ha="left") #texto
```

7. Gere um gráfico de dispersão que mostre a relação entre a área e o perímetro de cada grão de trigo.

8. Usando o cálculo da distância entre dois vetores que aprendemos em aula, informe qual é a semente que mais se parece com a primeira semente da base de dados (índice 0). Sua resposta deverá ser composta pelo índice no vetor e pelo valor da distância, por exemplo: 97, 0.11234.