Ciência da Computação

Instituto Federal do Ceará

Campus Maracanaú

Professor: Hericson Araújo

Contato: hericson.araujo@ifce.edu.br

Disciplina: Reconhecimento de Padrões

6ª Lista de Exercícios: K-Médias e PCA

Instruções:

- Implementação: Os algoritmos e modelos devem ser implementados do início em qualquer linguagem de programação (Python, Swift, Octave...), exceto quando indicado.

- Pacotes: É permitido o uso de pacotes auxiliares (como sklearn, matplotlib) para manipulação de dados e criação de gráficos.

- Entrega: Soluções podem ser enviadas em PDF ou Jupyter Notebook pelo Classroom. Os códigos também devem ser enviados.

1. Dados de Terremotos (quake.csv):

- Descrição: Conjunto de dados com latitudes e longitudes de locais onde foram registrados terremotos. Detalhes adicionais: https://www.openml.org/d/772.

Tarefas:

(a) Agrupamento com K-Médias:

- Avalie o algoritmo K-médias com distância Euclidiana.

- Escolha o número de grupos entre 4 a 20, com base no índice DB (Davies-Bouldin).

- Plote o melhor resultado de agrupamento obtido.

Observação: Execute o algoritmo várias vezes (ex.: 20 vezes) e escolha a solução com o melhor erro de reconstrução.

2. Dados de Pinguins (penguins.csv):

- Descrição: Conjunto de dados com medições anatômicas de pinguins da Antártica. As colunas contêm 4 atributos e 1 classe (especificando as espécies Adelie, Chinstrap e Gentoo). Detalhes adicionais: https://allisonhorst.github.io/palmerpenguins/.

Tarefas:

(a) Projeção em 2D com PCA:

- Use PCA (Análise dos Componentes Principais) para projetar os dados em 2 dimensões.

(b) Variação Explicada:

- Calcule e exiba a variação explicada ao alterar a dimensão projetada (1, 2, 3 ou 4).

Observação: Normalize os dados antes de realizar os cálculos.