Alexandria University
Faculty of Engineering

Introduction to Machine Learning
Assigned: Sunday, December 15, 2024
Due: Monday, December 23,2024

Assignment 4

The objectives of this assignment are as follows:

- Introducing dimensionality reduction through PCA.
- Introducing Clustering through K-means.
- Correct Implementation of PCA and K-Means
- Evaluate the impact of dimensionality reduction on clustering results.

Problem Statement

Given the Breast Cancer Wisconsin (Diagnostic) dataset, the objective is to perform unsupervised clustering to identify inherent patterns or groupings within the dataset. This dataset contains features computed from digitized images of fine needle aspirates (FNAs) of breast masses, aiming to distinguish between malignant and benign tumors.

Assignment Details

1. Implement K-Means using NumPy.
2. Implement PCA using NumPy.
3. You should have two experiments.
    - First experiment: Cluster the dataset using k-means.
    - Second experiment: Apply PCA then cluster the dataset using k-means.
    - Make Sure you don't use the labels in both experiments.
4. In both experiments.
    - In both experiments apply the elbow method to find the best Number (local minimum) of k clusters.
    - Compare between the sum of square errors/distances in both experiments.
    - Add your notes in both experiments.
5. In the second experiment:
    - Experiment with different numbers of principal components
    - You should visualize the results and compare it with the original labels.

Extra Notes:

- Use dataset from the following link https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data.
- You can't use k-means and PCA from scikit learn, you may validate your answer with it.
- You should work in groups of three. Each team should have one submission Id1_id2_id3.zip.
- Delivery will be ignored if you didn't follow the naming scheme provided in 2, any one of the team ids can be used.

Dr. Marwan Torki

Eng. Abdelrahman ElNainay Eng. Abdelrahman Sakr
Eng. Nour Eddine Hassan Eng. Youssef Nawar
Eng. Mohamed Badran

Grading Scheme:

- PCA and trying different principal components 35%
- K-means and trying different k values 35%
- Elbow Methods 20%
- Comparison and your notes 10%

Dr. Marwan Torki

Eng. Abdelrahman ElNainay Eng. Abdelrahman Sakr
Eng. Nour Eddine Hassan Eng. Youssef Nawar
Eng. Mohamed Badran