Summer 2021: CSEE5590 – Special Topics

Python_Lesson_3_Part_2: Machine Learning: Clustering and PCA

Lesson Overview:
In this lesson, we will focus on clustering techniques
Clustering using K-Means and Dimension reduction using PCA.

Use Case Description:
1. K-Means Clustering and PCA

Programming elements:
K-Means Clustering and Data Analysis

Source Code:
Provided in the assignment repo & Canvas use-case file.

Assignment:

1. Apply K means clustering to credit card dataset: CC.csv

    - Remove any null values by the mean.
    - Use the elbow method to find a good number of clusters with the K-Means algorithm
    - Calculate the silhouette score for the above clustering.

2. Try feature scaling and then apply K-Means on the scaled features. Did that improve the Silhouette score? If yes, can you justify why.

3. Apply PCA on the same dataset. Apply K-Means algorithm on the PCA result and report your observation if the silhouette score improved or not?

***** Bonus points**

4. **Visualize the clustering of first question.**

Online Submission Guidelines (for Online students):
1. Submit your source code and documentation to GitHub and represent the work in a ReadMe file properly (submit your screenshots as well. The screenshot should have both the code and the output)
2. Comment your code appropriately
3. Video Submission (1 – 3 min video showing the demo of the assignment, with brief voice over on the code explanation)

Note: *Cheating, plagiarism, disruptive behavior and other forms of unacceptable conduct are subject to strong sanctions in accordance with university policy. See detailed description of university policy at the following URL: https://catalog.umkc.edu/special-notices/academic-honesty/*