CS 510 Data Clustering Spring 2023

Informal Documentation for Coding project

Text Clustering

Amazon food reviews clustering using K-Means, K-Means++ and Agglomerative Clustering

Collaborators - Rajashree Pethkar, Upasana Chaudhari

Goals –

- Cluster data using k-means, k-means++
- Visualize the clusters.
- Compare the results.

How do we achieve the above goals?

- Preprocessed the data.
- Extracted features using TF-IDF
- Implemented PCA
- Performed K-means, k-means++ and Agglomerative clustering.
- Generated the output by plotting the clusters.
- Compare the results.

How to execute the source code to reproduce your results?

- We have shared an ipynb file and a sample data set of approximately 5000 rows which we have selected randomly.
- To execute our code you can use any python editor example- jupyter notebook or google colab.
- Upload the given data file and the ipynb file in your project directory.
- Update the file path for the data file in your code.
- Execute the source code as we usually do to execute the cells one by one (shift+enter).