

Task1: Run the kmeans multiple times and see if averaging the cluster centers across runs helps to reduce dependency on initialisation. Here the assumption is that good clustering is more likely than bad one. Dataset: The 5 clusters

Task2: Run the algorithm on concentric circles dataset in r-theta space or some other space you like.

Task3: Recommender system (Do the dataset creation by 7)

You need to install pandas and use python3

1. Download and run the script.py file and the tmdb_5000_movies.csv file from the link below and put them in the same folder.
2. Run script.py using python3 in the same folder as the tmdb_5000_movies.csv.
3. Select a movie preference out of the 5 that are shown. Do this till you have 30 movies at least.
4. A final.json file will be generated in the folder from which you ran script.py
 - a. Rename final.json to final_<yourname>.json
 - b. Put this file in the Final Datasets folder in the link below
5. Download the data after 7PM today and try clustering based on various movie preferences.

[Link for script, dataset and Final Dataset folder](#)