Assignment-2

Data Mining

Instructions

1. Do the code using python; it can be used in Jupyter notebook.
2. Create a Github private repo where all your assignments and projects would be stored. At an opportune time, you would be asked to share your repo with our evaluation team.
3. Your submission should be sufficiently original to be considered for evaluation.
4. Submissions would include source code, data availability at Github, and Google classroom submission of the markdown pdf (do not submit any other format on Google classroom).
5. You can use any library, no need to code from scratch.

Submission date and time

April 25, 2021, by 6 PM.

Please follow the steps below to complete your assignment:

1. You need to download "Wine" data from the kaggle; ref is given below. [2]
2. Follow assignment-1 steps 2 to 5 to perform data processing and write inferences. [3]
3. Perform at least five clustering methods with varying cluster sizes. Find correct cluster numbers for each method and show with line plot; how you finalized this cluster number. [5]
4. Compare all the clustering methods, and write how you compared, what method you used and why? [5]
5. Do steps 3 and 4 again after dimension reduction, and show whether it is helpful; if not, why? [5]

Ref:

1. https://www.kaggle.com/harrywang/wine-dataset-for-clustering

Note: References can be used only for learning purposes, not like copy-paste.