Course ID: CSE-506
Total Points 100

Assignment 3

Deadline: Nov 27, 2021

Instructions

1. The assignment is to be attempted in pairs.
2. Programming Language: Python
3. For Plagiarism, institute policy will be followed
4. You need to submit the readme.pdf, Code files and Model files.
5. Report, models and code in .py format should be submitted in the classroom in a zip folder with the name A3_RollNumber1_RollNumber2.zip.
6. Attach a Report.pdf including all your hypothesis, procedure, steps and outcomes in the zip file.
7. You **can use any library** for pre-processing, training, doing experiments and post-processing in all questions.
8. One member should submit on google classroom while other member can mark turn in without the attachment.
9. In case of doubts, please comment on the classroom.

Dataset: Attached in the assignment post.

Q1: (40 points) Using any library, Perform any three clustering modelling and one Gaussian based clustering modelling and report the following for each model-

1. Centroid/representative object/prototype of each cluster for every model.
2. Visualization of the clusters. (You can use lesser data points/ dimensions for visualizations).
3. Compare your cluster distribution with the true label count.
4. Compare the cluster formation of the gaussian based method with the other three clustering methods and report your observations on the results.

Q2: (60 points) The points will be given based on the performance on the test set. You need to modify the code snippet attached to produce output for the test set, and return the output labels. You are allowed to use any clustering techniques from any library and any optimization technique for improving your model's performance.

Evaluation Metrics : Balanced F1
Create your own train and validation set and measure your performance against it. The column "target" is the true label and are for your performance benchmarks.

Rank	Points
Top 20%	100%
Next 30%	90%
Next 30%	80%
Last 20%	70%

Table 1: Points according to rank.

PS: If the code doesn't run or gives any error, 0 points will be awarded.

Best of luck for the Assignment and Final exam