CLL788 - Process Data Analytics

Assignment 3

Deadline: 10th April 11:55 PM

Question : SVM

1. Plot the training data (Data1.xlsx) to get an idea of the data distribution.Plot the points with variable 1 on x-axis and variable 2 on y-axis. Now color the coordinates/points of class 0 with blue and class 1 with red. Report your visual observations.
2. Apply SVM and KMeans on training data (Data1.xlsx) to find decision boundaries. Plot training data along with decision boundary.
3. Now apply SVM with“modified optimization problem”on Data2.xlsx and try out different values of C and report your observations along with plots of the decision boundary.

Note: *For SVM modeling you can use numerical packages available. Experiment with kernel C and other hyper parameters and report the results*

Question: K means

1. Manually perform K Means clustering on Manual_Data.xlsx. There are 10 data points given and you have to separate them into 2 clusters.

Question: GMM

1. Carbon and nitrogen emission tests of 2 different types of vehicles were done. Test results are provided in Excel sheets. Your task is to identify the two groups of vehicles from the data.
   a) Plot the data (Data.xlsx) to get an idea of the data distribution. Plot Result 1 on x-axis and Result 2 on y-axis. Report your visual observations.
   b) Apply K-Means clustering on the data to find out the 2 clusters. Make appropriate plots.
   c) Plot the data (Data_GMM.xlsx) to get an idea of the data distribution. Plot Result 1 on x-axis and Result 2 on y-axis. Report your visual observations.
   d) Apply Gaussian Mixture Model on the Data_GMM.xlsx to find out the 2 clusters. Make appropriate plots.
   e) Compare the two methods used.

Note Use of GMM and K-means libraries are strictly not allowed

Submission Details

1. Submit two files, first is a pdf report describing your work with all the graphs and analysis included, second is a zip file containing codes on Moodle.
2. Name the files <EntryNumber>_report_assgn2.pdf and < EntryNumber >_codes_assgn2.zip respectively. Only MATLAB(.m) & python (.py or .ipynb) are allowed for this assignment.
3. The deadline for the submission is 10th April 11:55 PM.