School of Computer Science and Engineering, VIT Chennai.

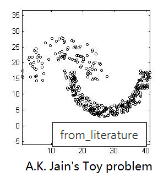
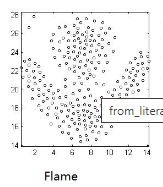
BCSE209L Machine Learning

**Lab-10 K-means Clustering Algorithm**

Faculty : Dr. R. Jothi

Submit your python code (Jupyter notebook)

Q1. Consider the following 2-D datasets flame (flame.csv) and toy (toy.csv).



Implement K-means clustering algorithm to segment the above datasets into 2 clusters. Show the scatterplot for the original clusters as well as the clusters obtained by your implementation.

Q2. Verify results obtained in Q1 using Sklearn library. Change the algorithm parameters such as distance metric, number of iterations, initialization (seed) method etc. and plot the results.

Q3. Identifying valid customer groups of a retailer shops / malls is an important business problem. Find attached a dataset (shop.csv) having customer information such as gender, age, annual income and spending score. Based on these information you need to identify a set of customer groups having similar purchase patterns. The number of groups is unknown (use elbow method to find optimal number clusters). Report your clustering performance metrics (Rand index, Silhouette index). Use sklearn library for clustering as well as metrics.