School of Computer Science and Engineering, VIT Chennai.

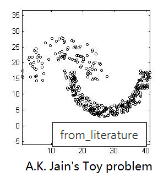
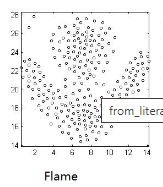
CSE4020 Machine Learning

**Lab-10 Cluster Analysis**

Faculty : Dr. R. Jothi

**Due Date : 26/10/2022**

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



Implement a partitonal 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. Mention the algorithm parameters such as distance metric, number of iterations, etc.

B. Use scikit library functions for partitional as well as hierarchical agglomerative clustering (with all linkage criterions) to perform cluster analysis and report the results in a plot.

C. Identifying valid customer groups of a retailer shops / malls is an important business problem. Find attached a dataset 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) for both partitional as well as hierarchical agglomerative clustering.