School of Computer Science and Engineering, VIT Chennai.

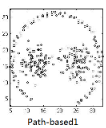
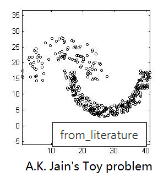
BCSE209L Machine Learning

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

Faculty : Dr. R. Jothi

Submit your python code (Jupyter notebook)

Q1. Consider the following 2-D datasets Path-based and Toy datasets.

Apply K-means, hierarchical clustering algorithms (use Sklearn) to segment the above datasets into k clusters (as given in ground truth values of datasets). Show the scatterplot for the original clusters (as given in dataset file) as well as the clusters obtained by your implementation.

Q2. 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.
* Extract principal components (PCA) form the dataset, and apply the algorithms and re-evaluate the performance.