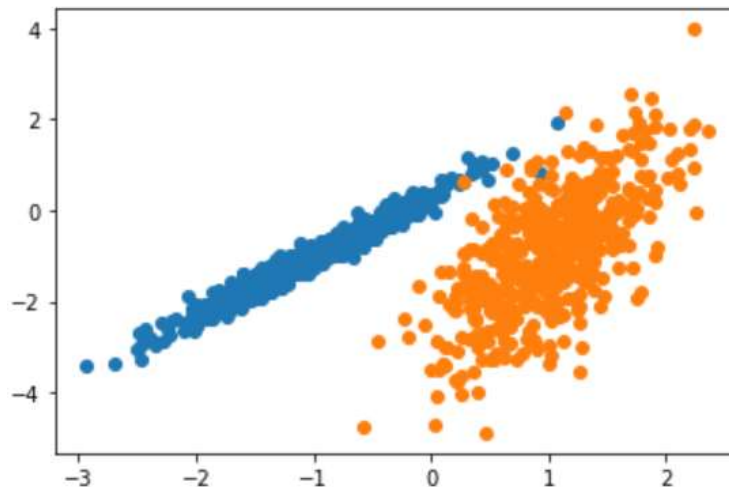


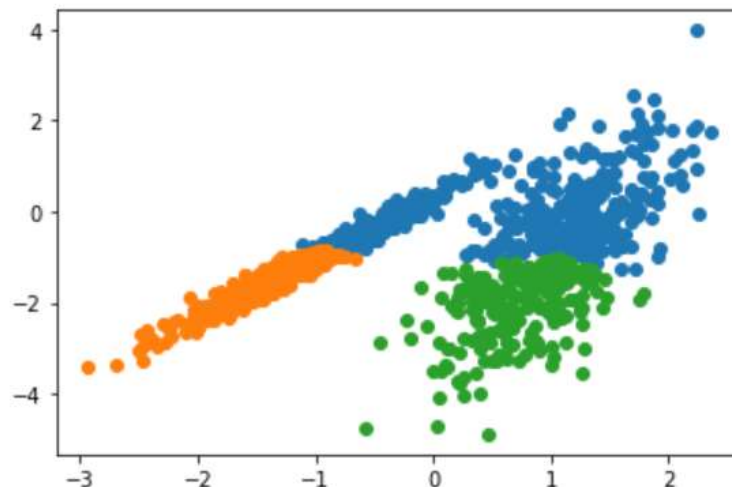
DMDW ASSIGNMENT 3 – REPORT

SRUJAN R - BT18CSE041

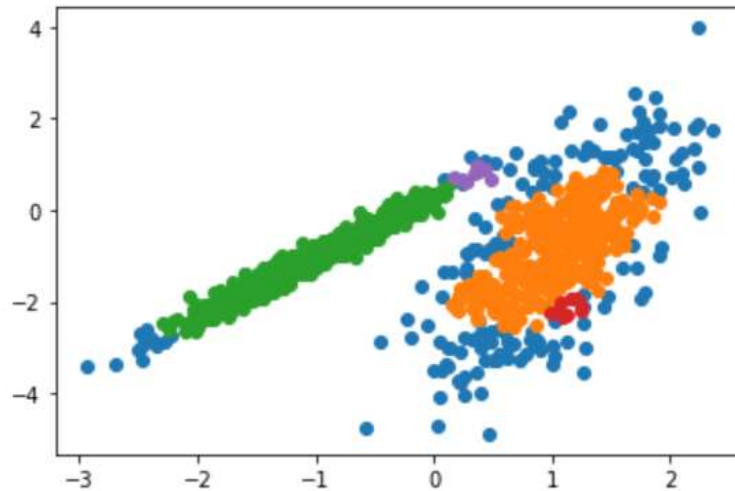
- Clustering is an unsupervised problem of finding natural groups in the feature space of input data. There are many different clustering algorithms, and no single best method for all datasets. Hence, we would have to check every algorithm to check which method works well for a given dataset.
- Used the in-built function of sklearn to get a suitable dataset for clustering, with parameters such that the clusters can be plotted on a 2-dimensional graph.
- I have loaded the in-built models of sklearn for the following clustering algorithms: AGNES, DBSCAN and K-Means.
- I am fitting the model on the training dataset which then predicts a cluster for each example in the dataset. A scatter plot is then created with points coloured by their assigned cluster.



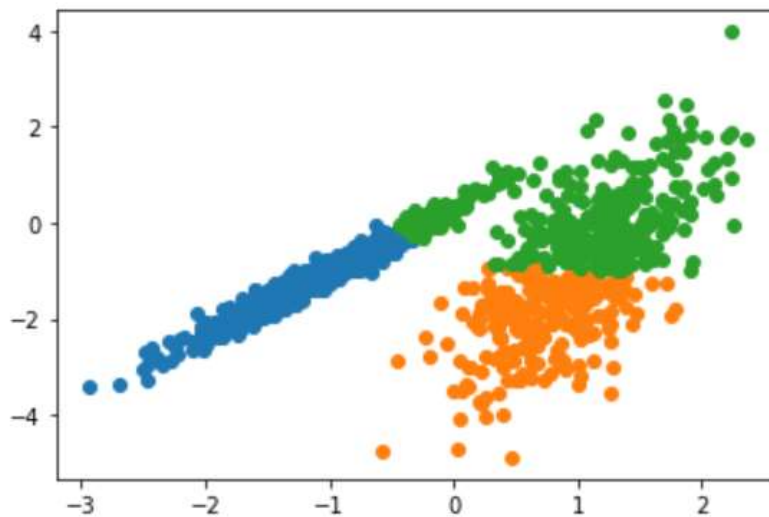
- I have initially used Agglomerative Nesting (AGNES) clustering algorithm which involves merging examples until the desired number of clusters is achieved. It is a form of hierarchical clustering method. It clearly gave 3 clusters neatly divided with an excellent grouping.



- Next, I used DBSCAN Clustering (Density-Based Spatial Clustering of Applications with Noise) which is a density-based clustering method. It involves finding high-density areas in the domain and expanding those areas of the feature space around them as clusters. In this case, a reasonable grouping is found with 5 different clusters, although more tuning is required.



- Lastly, I have used the K-Means clustering algorithm which is a partitioning based clustering method. In this case, a reasonable grouping is found, although the unequal equal variance in each dimension makes the method less suited to this dataset.



- So, we can conclude that AGNES clustering method works best for this dataset based on the results seen.