COMP809 – Agglomerative clustering and K-means analyses

Lab 6

1. Simulate a data set using the following python code:

```
X, y = make_blobs(n_samples=100,
                  n_features=2,
                  centers=5,
                  cluster_std=1,
                  shuffle=True,
                  random_state=1)
```

   The make_blobs function can be accessed from sklearn.datasets. X contains the features and y the cluster number.

   a. Perform an agglomerative cluster analysis. How many clusters would you recommend? Justify your answer.
   b. Increase the number of features to 10? How many clusters would you recommend? Justify your answer.
   c. What can you conclude from the results in a) and b).
   d. Plot your cluster results in a scatter plot for both data sets. Comment on it.
2. Analise the simulated data generated in question through K-means.
   a. How many clusters would you recommend when there are 2 features? Justify your answer.
   b. How many clusters would you recommend when there are 10 features? Justify your answer.
   c. What can you conclude from the results in a) and b).
   d. Plot your cluster results in a scatter plot for both data sets. Comment on it.