**Lab Week 6**

**Unsupervised Learning**

**Part A: Hierarchical Clustering**

Topic 1: Agglomerative

1. Import libraries:
   1. Numpy
   2. Matplotlib
   3. Agglomerative Clustering
   4. Dendrogram
   5. Linkage
2. Set both X and y values based on the following table:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 4 | 5 | 10 | 4 | 3 | 11 | 14 | 6 | 10 | 12 | 9 | 14 | 12 | 11 | 11 | 10 |
| y | 21 | 19 | 24 | 17 | 16 | 25 | 24 | 22 | 21 | 21 | 18 | 16 | 17 | 17 | 18 | 16 |

1. Store both x and y value into a list and print each data point coordinate based on x and y values.
2. Visualization: Plot all the data points using scatter plot.
3. Build AgglomerativeClustering models based on the following parameters:
   1. n\_clusters = (repeat: 1 to 10)
   2. metric = ‘euclidean’
   3. linkage = ‘ward’
4. Produce cluster groups based on each model trained using different number of clusters.
5. For each model in (5), create scatter plot graph.
6. Create a linkage model using linkage() with the following parameters:
   1. y = data
   2. method = ‘ward’
   3. metric = ‘euclidean’
7. Plot a dendrogram using the linkage produced from (8).
8. Based on the dendrogram, determine suitable threshold limit to form a cluster by plotting a horizontal line on the dendrogram.
9. Without setting the number of clusters, create a new agglomerative clustering model by setting the distance threshold parameter from (10).
10. Produce the clustering result.

**Part B: Case study based on price data**

Explore on the ETF (Exchange Traded Fund) price data of S&P Global data set (Standard and Poor 500). Perform data analysis and clustering analysis on the dataset and determine whether which equity can be grouped together. You may use any clustering algorithms (K-means, Agglomerative and divisive hierarchical clustering). Produce the clustering results.