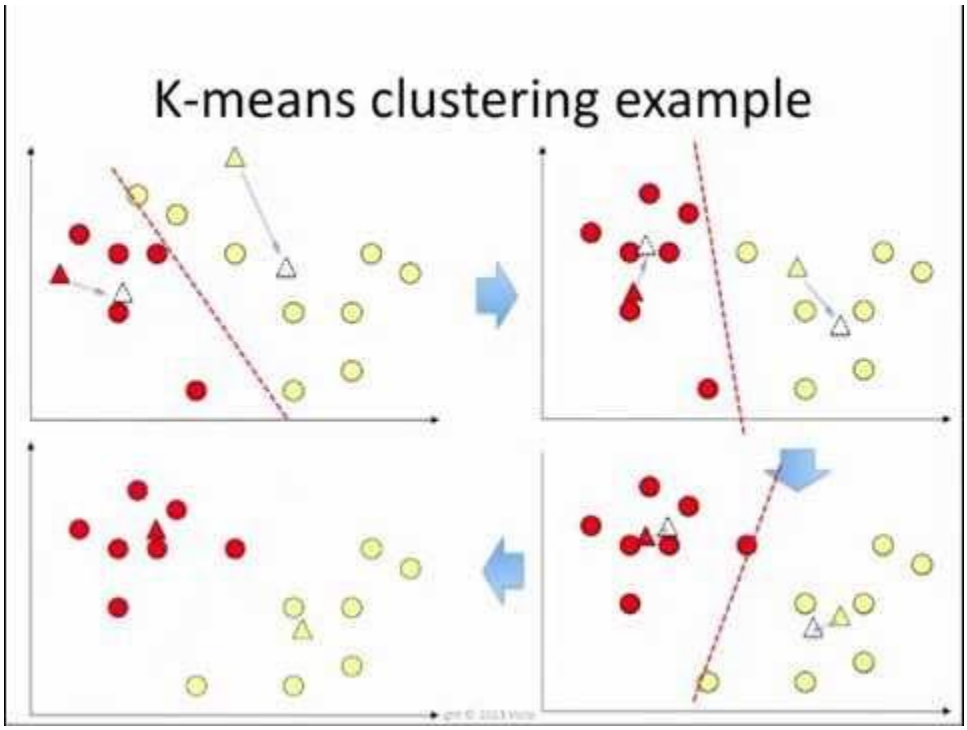
Q1) Describe different methods of clustering and give some examples as well as illustrations.

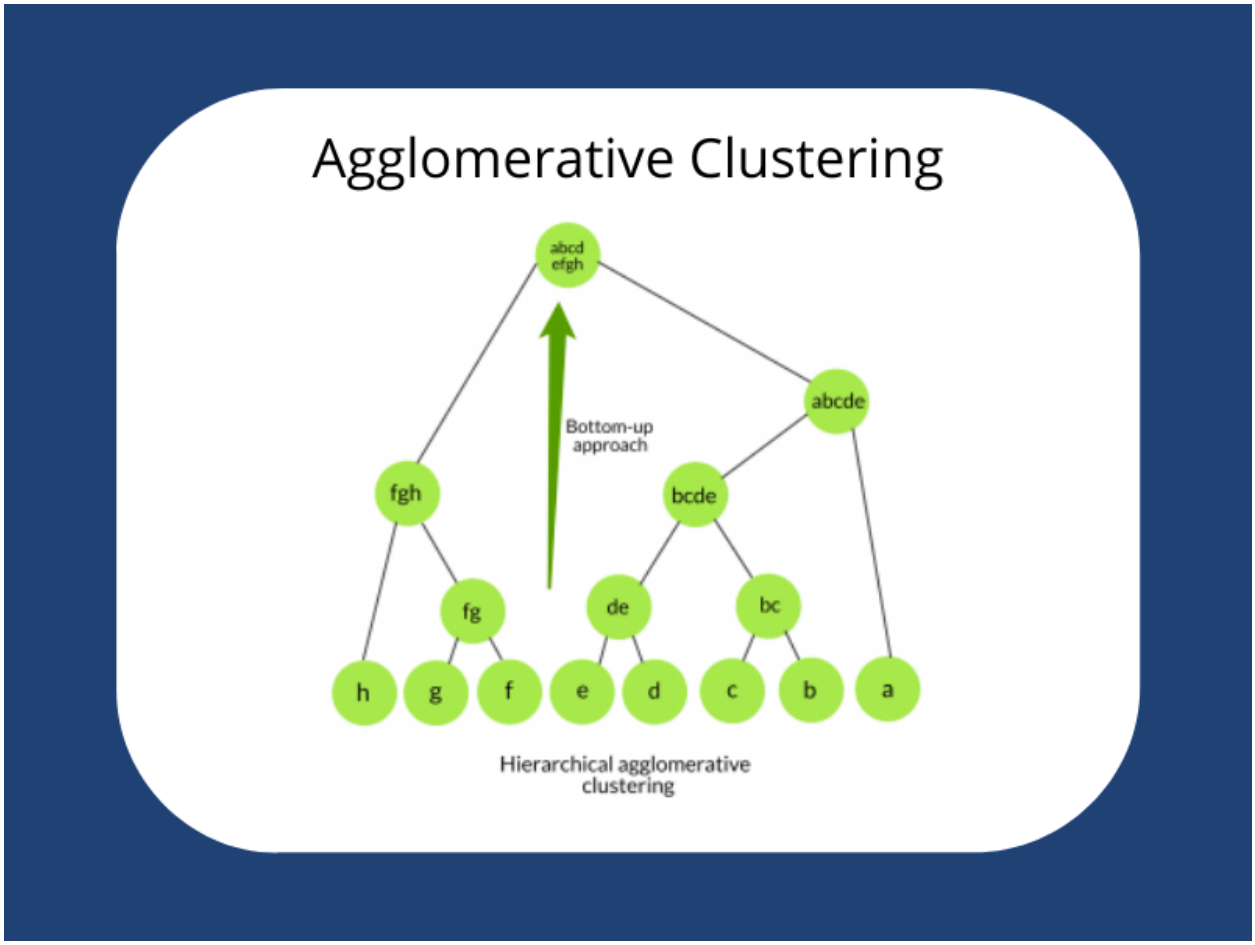
Clustering is a machine learning technique, specifically unsupervised machine learning, which groups data points together based on their similarities. There are several clustering methods, each with its own approach and use cases. Here are some of the most commonly used clustering methods along with examples:

- K-Means: this is a method of partitioning in which the data is divided into what are called “K clusters”. These clusters are organized based on the average/mean/’centroid’ of the data points within each cluster. An example of a good use of K-Means clustering would be using it on a store’s commerce data in order to create sets of similar customers which you can use for specific targeting of ads, promotions, etc.



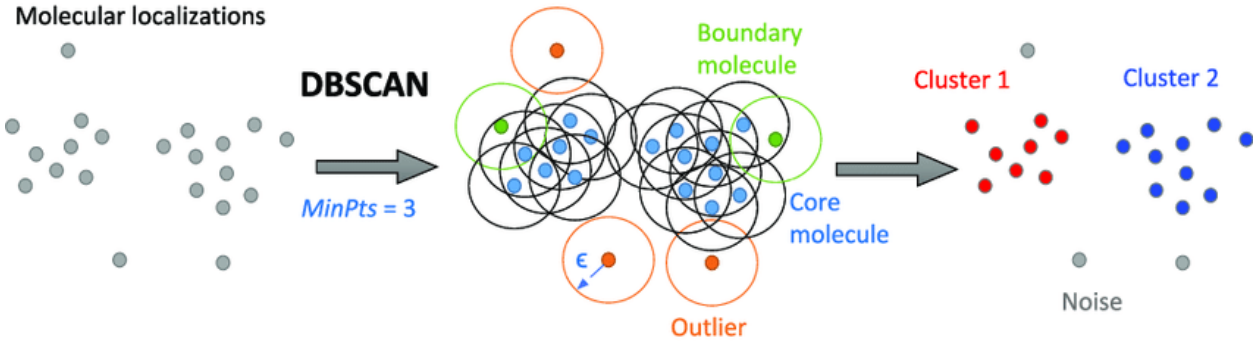
- Hierarchical: hierarchical clustering works by building a tree-like structure of clusters nested within each other by successively merging or splitting data points based on their perceived similarities. It essentially iterates through the data determining if each point is alike enough to group together or split apart. Hierarchical clustering works especially well with smaller data sets. A couple examples of a good use of hierarchical clustering would be grouping animals according to their biological features to reconstruct phylogeny trees, or forming a

hierarchy of employees based on salary.



- DBSCAN: DBSCAN clustering, which stands for ‘Density-Based Spatial Clustering of Applications with Noise’, is a method of clustering which works by judging whether or not data points should be clustered together by their density. It essentially groups together data points in areas with a high density of data, and separates the data points in areas of lower density of data. The measure of high/low density is of course relative to the current dataset. Of all the common clustering algorithms out there, DBSCAN is one of the algorithms that makes the fewest assumptions about the shape of your clusters. That means that DBSCAN can be used to detect clusters that are oddly or irregularly shaped. An example of a good use for DBSCAN clustering would be identifying ‘hotspots’ within

datasets.



There are several other methods of clustering that I won't delve into as deeply right now such as:

1. Agglomerative clustering, which is a type of hierarchical clustering that starts with individual data points as separate clusters and then merges the closest clusters iteratively.
2. Mean Shift clustering, which is a non-parametric clustering method that shifts data points towards the mode (peak) of the density function.
3. Spectral clustering, which transforms data into a lower-dimensional space and performs clustering in that space. It is effective for complex data structures.

Q2) Write code for different clustering approaches, plot graphs, and explain. [Code is in hw5.py. Code output, including the graph I printed for hierarchical clustering using matplotlib.pyplot, are included in images below.]

```
nayatrov@ariahas-MBP AI-HW % python3 hw5.py
Data:
[[ 1.   2. ]
 [ 5.   8. ]
 [ 1.5  1.8]
 [ 8.   8. ]
 [ 1.   0.6]
 [ 9.  11. ]
 [ 3.   2.5]
 [ 7.   9. ]
 [ 2.   2. ]
 [ 7.5  7. ]]
/usr/local/lib/python3.11/site-packages/sklearn/cluster/_kmeans.py:1412: FutureWarning: The default valu
e of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the
 warning
  super()._check_params_vs_input(X, default_n_init=10)

K-Means Clustering
        X


      X
    X  X







XX


Cluster Labels: [0 1 0 1 0 1 0 1 0 1]
Centroid Coordinates: [[1.7  1.78]
 [7.3  8.6 ]]

Hierarchical Clustering

DBSCAN Clustering
        X


      X
    X  X







XX


Cluster Labels: [ 0 -1  0  1  0 -1  0  1  0  1]
```

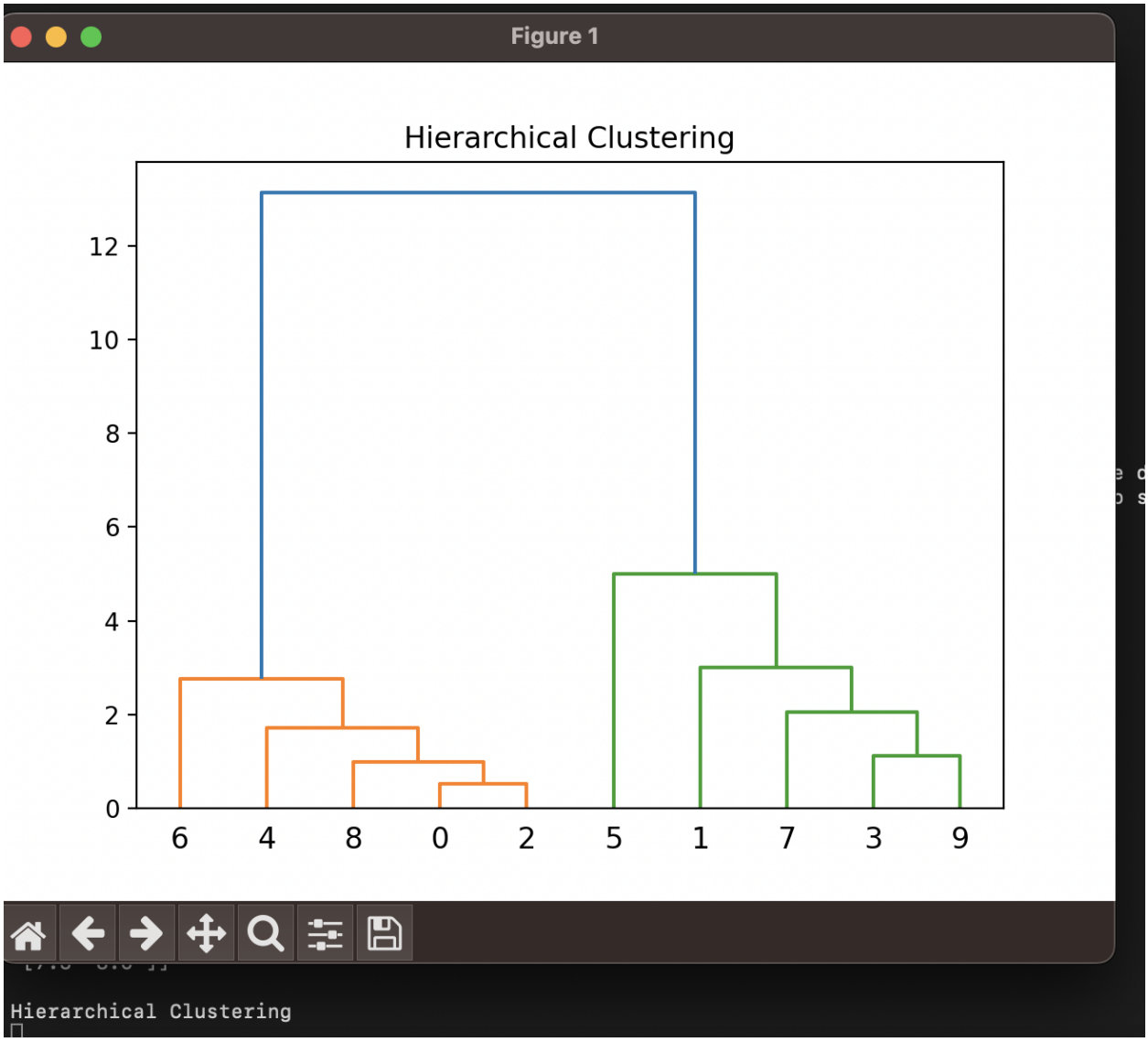


Figure 1
Hierarchical Clustering
12
10
8
6
4
2
0
6
4
8
0
2
5
1
7
3
9
Hierarchical Clustering