

The code to plot the clustered labels and the cluster centers should be executed in the same notebook. The plot of clusters made by the K-means algorithm is shown in Figure 25-4.

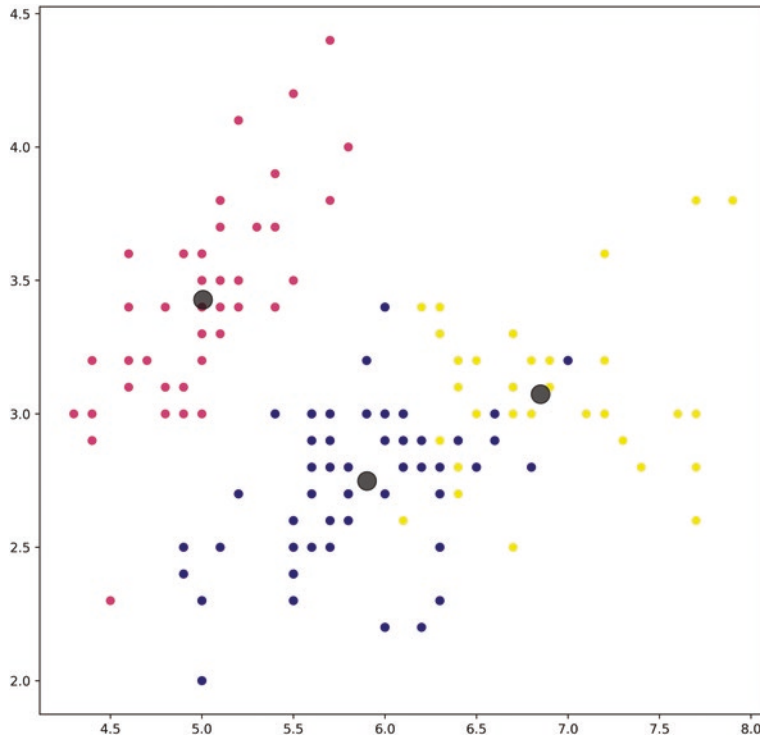


Figure 25-4. Plot of K-means clusters and their cluster centers

Hierarchical Clustering

Hierarchical clustering is another clustering algorithm for finding homogeneous sub-groups or classes within a dataset. However, as opposed to k -means, we do not need to make an a priori assumption of the number of clusters in the dataset before running the algorithm.

The two main techniques for performing hierarchical clustering are

- Bottom-up or agglomerative
- Top-down or divisive

In the bottom-up or agglomerative method, each data point is initially designated as a cluster. Clusters are iteratively combined based on homogeneity that is determined by some distance measure. On the other hand, the divisive or top-down approach starts with a cluster and subsequently splits into homogeneous sub-groups.

Hierarchical clustering creates a tree-like representation of the partitioning called a dendrogram. A dendrogram is drawn somewhat similar to a binary tree with the root at the top and the leaves at the bottom. The leaf on the dendrogram represents a data sample. The dendrogram is constructed by iteratively combining the leaves based on homogeneity to form clusters moving up the tree. An illustration of hierarchical clustering is shown in Figure 25-5.

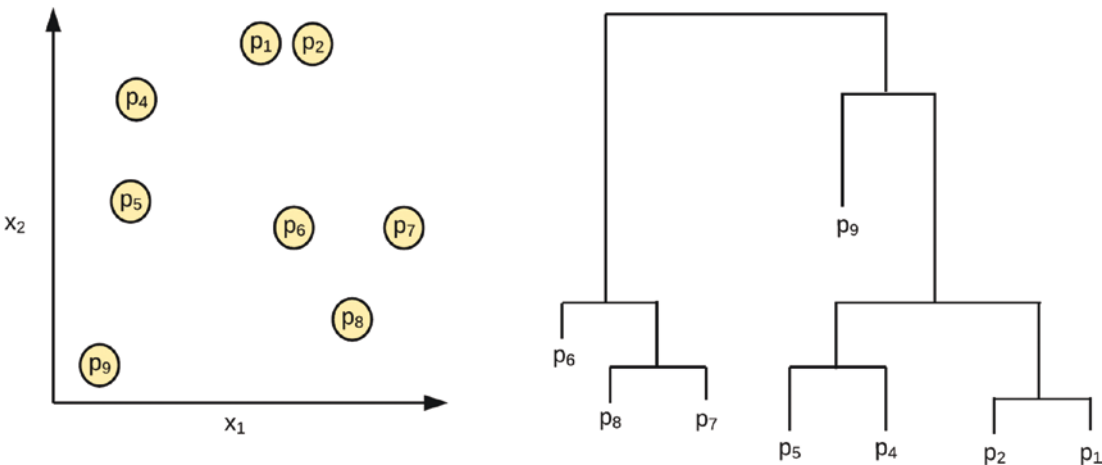


Figure 25-5. An illustration of hierarchical clustering of data points in a 2-D feature space. Left: The spatial representation of points in 2-D space. Right: A hierarchical cluster of points represented by a dendrogram.

How Are Clusters Formed

Clusters are formed by computing the nearness between each pair of data points. The notion of nearness is most popularly calculated using the Euclidean distance measure. Beginning at the leaves of the dendrogram, we iteratively combine those data points that are closer to one another in the multi-dimensional vector space until all the homogeneous points are placed into a single group or cluster.