4.2.4.7 Bisecting *K*-Means Clustering

Bisecting *K*-means clustering [47] is a divisive hierarchical clustering method which uses *K*-means repeatedly on the parent cluster *C* to determine the best possible split to obtain two child clusters C_1 and C_2. In the process of determining the best split, bisecting *K*-means obtains uniform-sized clusters. The algorithm for bisecting *K*-means clustering is given in Algorithm 17.

Algorithm 17 Bisecting *K*-Means Clustering

1: **repeat**
2: Choose the parent cluster to be split *C*.
3: **repeat**
4: Select two centroids at random from *C*.
5: Assign the remaining points to the nearest subcluster using a prespecified distance measure.
6: Recompute centroids and continue cluster assignment until convergence.
7: Calculate inter-cluster dissimilarity for the 2 subclusters using the centroids.
8: **until** *I* iterations are completed.
9: Choose those centroids of the subclusters with maximum inter-cluster dissimilarity.
10: Split *C* as C_1 and C_2 for these centroids.
11: Choose the larger cluster among C_1 and C_2 and set it as the parent cluster.
12: **until** *K* clusters have been obtained.

In line 2, the parent cluster to be split is initialized. In lines 4–7, a 2-means clustering algorithm is run *I* times to determine the best split which maximizes the Ward's distance between C_1 and C_2. In lines 9–10, the best split obtained will be used to divide the parent cluster. In line 11, the larger of the split clusters is made the new parent for further splitting. The computational complexity of the bisecting *K*-means is much higher compared to the standard *K*-means.

4.2.4.8 Kernel *K*-Means Clustering

In Kernel *K*-means clustering [44], the final clusters are obtained after projecting the data onto the high-dimensional kernel space. The algorithm works by initially mapping the data points in the input space onto a high-dimensional feature space using the kernel function. Some important kernel functions are polynomial kernel, Gaussian kernel, and sigmoid kernel. The formula for the SSE criterion of kernel *K*-means along with that of the cluster centroid is given in Equation (4.15). The formula for the kernel matrix *K* for any two points $x_i, x_j \in C_k$ is also given below.

$$SSE(C) = \sum_{k=1}^{K} \sum_{x_i \in C_k} ||\phi(x_i) - c_k||^2 \tag{4.15}$$

$$c_k = \frac{\sum_{x_i \in C_k} \phi(x_i)}{|C_k|} \tag{4.16}$$

$$K_{x_i x_j} = \phi(x_i) \cdot \phi(x_j) \tag{4.17}$$

The difference between the standard *K*-means criteria and this new kernel *K*-means criteria is only in the usage of projection function ϕ. The Euclidean distance calculation between a point and the centroid of the cluster in the high-dimensional feature space in kernel *K*-means will require the knowledge of only the kernel matrix *K*. Hence, the clustering can be performed without the actual individual projections $\phi(x_i)$ and $\phi(x_j)$ for the data points x_i, $x_j \in C_k$. It can be observed that the