Assignment 3 Q3

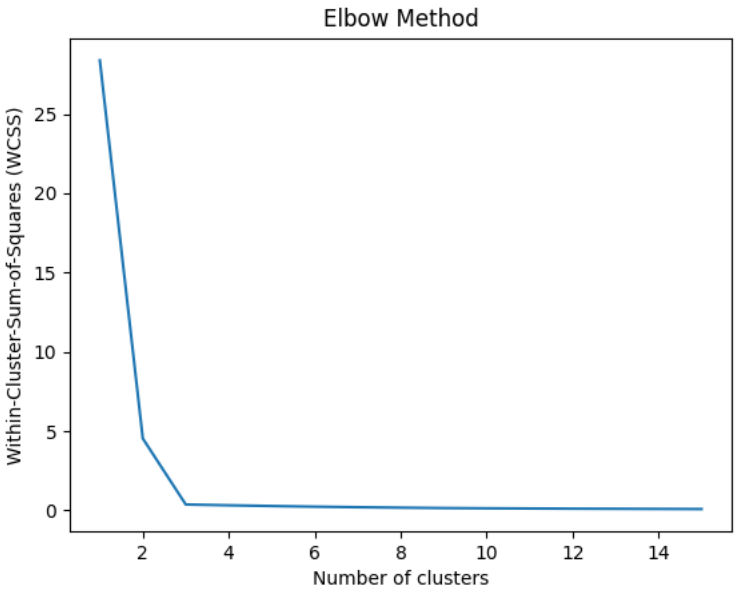

Image 1: Elbow plot for 2-d dataset

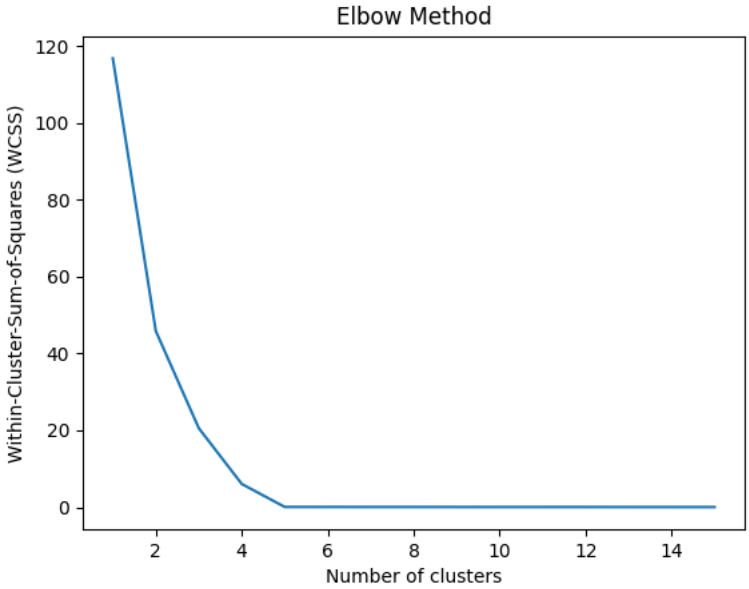

Image 2: Elbow plot for 3-d dataset

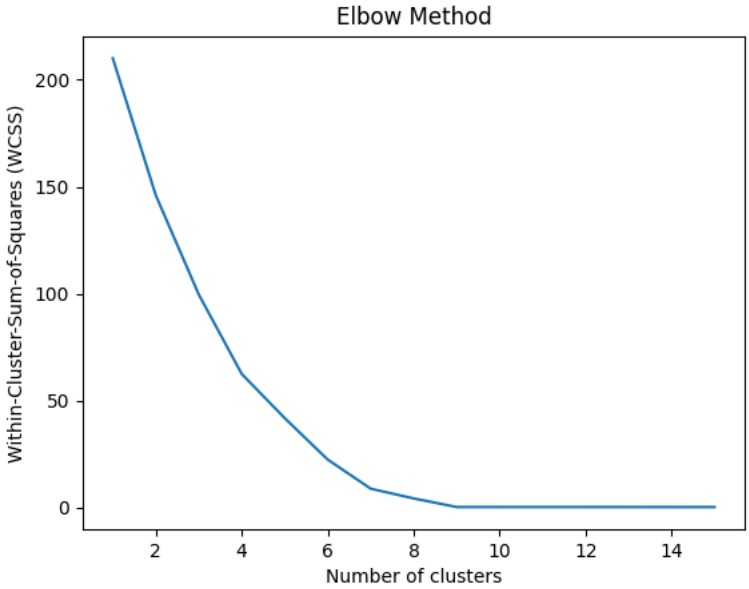

Image 3: Elbow plot for 4-d dataset

Explanation: As we can see from the images above that initially there is a steep decline in within cluster sum of square distances with the increase in number of clusters. We observe a elbow shape with a knee point at a specific value of k. Beyond the knee point as K increases we do not observe significant decrease in within cluster sum of square distances. This is because beyond the knee point the groups of datapoints are already compact (having low intra cluster distance) thus creating a cluster. This knee point specifies the number of clusters in our dataset.

For 2-D dataset, we observe knee point at K=3, indicating that there are 3 clusters.

For 3-D dataset, we observe knee point at K=5, indicating that there are 5 clusters.

For 4-D dataset, we observe knee point at K=9, indicating that there are 9 clusters.

There are multiple metrices to evaluate the quality of clusters formed such as –

1. Silhouette index
2. Dunn's index