

- 1) This may not be good as the distribution of data is assumed to be multivariate normal distribution, but it may not fit the situation of the distribution of the clustered data. K-means only consider Euclidean distance during clustering but didn't take variety of dependency and scale of data into consideration. Thus, the classification by computing covariance may not have good accuracy.
- 2) K-means algorithm clusters data by iterative process of cluster assignment and centroid adjustment while average-link clustering clusters data by continuously joining up smaller cluster to bigger one (or reverse).

Similarity: both algorithms are clustering iteratively by the average instances within a cluster

Difference: K-means have k clusters in every iteration while average link clustering has decreasing (or increasing if agglomerative clustering) number of clustering in ongoing iterations

- 3) We can remove outliers, which are instances that far away from other and have low probability, by local outlier factor first. For example, the instance (denoted by x) that have $LOF(x) > 1$ can be removed before applying into k-means.