

- 1) b) 1& 2
- 2) d)1,2 &4
- 3) a) True
- 4) a) 1 only
- 5) b)1
- 6) b) No
- 7) a) Yes
- 8) d) All of the above
- 9) a) K-means clustering algorithm
- 10) d) All of the above
- 11) d) All of the above
- 12) k-means can be quite sensitive to outliers in your data set. The reason is simply that k-means tries to optimize the sum of squares. And thus a large deviation (such as of an outlier) gets a lot of weight.

The k-means algorithm updates the cluster centers by taking the average of all the data points that are closer to each cluster center. When all the points are packed nicely together, the average makes sense. However, when you have outliers, this can affect the average calculation of the whole cluster. As a result, this will push your cluster center closer to the outlier.

- 13) The K-means clustering algorithm is used to find groups which have not been explicitly labeled in the data. This can be used to confirm business assumptions about what types of groups exist or to identify unknown groups in complex data sets.

Guarantees convergence. Can warm-start the positions of centroids. Easily adapts to new examples. Generalizes to clusters of different shapes and sizes, such as elliptical clusters.

K-Means for Clustering is one of the popular algorithms for this approach. Where K means the number of clustering and means implies the statistics mean a problem. It is used to calculate code-vectors (the centroids of different clusters).

- 14) The basic k-means clustering is based on a non-deterministic algorithm. This means that running the algorithm several times on the same data, could give different results.

The non-deterministic nature of K-Means is due to its random selection of data points as initial centroids. ... The key idea of the algorithm is to select data points which belong to dense regions and which are adequately separated in feature space as the initial centroids.