MACHINE LEARNING

Answer the Following Questions

Ans 1-(a) 2 Only

Ans 2-(d) 1, 2 and 4

Ans 3-(a) True

Ans 4-(a) 1 only

Ans 5-(b) 1

Ans 6-(b) No

Ans 7-(a) Yes

Ans 8-(d) All of the above

Ans 9-(a) K-means clustering algorithm

Ans 10-(d) All of the above

Ans 11-(d) All of the above

Ans 12-The k-means algorithm is sensitive to the outliers. In this paper, we propose a robust two-stage k-means clustering algorithm based on the observation point mechanism, which can accurately discover the cluster centers without the disturbance of outliers. In the first stage, a small subset of the original data set is selected based on a set of nondegenerate observation points. The subset is a good representation of the original data set because it only contains all those points that have a higher density of the original data set and does not include the outliers. In the second stage, we use the k-means clustering algorithm to cluster the selected subset and find the proper cluster centers as the true cluster centers of the original data set.

Ans 13-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). According to a tutorial, for any word/value/key that needs to be 'vector quantized', it is by calculating the distance from all the code vectors and assign the index of the code vector with the minimum distance to this value. For example, clustering can be applied to MP3 files, cellular phones are the general areas that use this technique

Ans 14-The non-deterministic nature of K-Means is due to its random selection of data points as initial centroids. Method: We propose an improved, density based version of K-Means, which involves a novel and systematic method for selecting initial centroids.