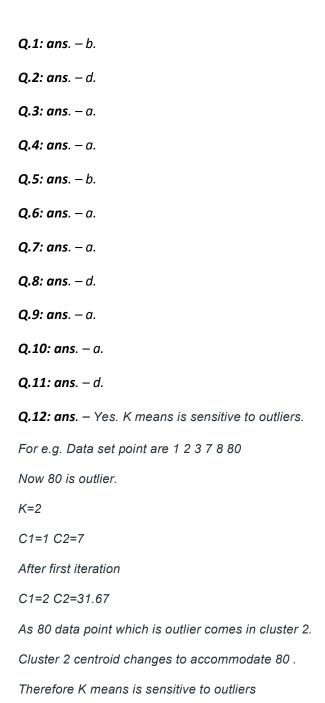
WORKSHEET-II



Q.13: ans. – The Balance Iterative Reducing and Clustering using Hierarchies (BIRCH) algorithm works better on large data sets than the k-means algorithm. It breaks the data into little summaries that are clustered instead of the original data points..

Q.14: ans. – 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.