

WORKSHEET-II

Q.1: ans. – b.

Q.2: ans. – d.

Q.3: ans. – a.

Q.4: ans. – a.

Q.5: ans. – b.

Q.6: ans. – a.

Q.7: ans. – a.

Q.8: ans. – d.

Q.9: ans. – a.

Q.10: ans. – a.

Q.11: ans. – d.

Q.12: ans. – Yes. *K means is sensitive to outliers.*

For e.g. Data set point are 1 2 3 7 8 80

Now 80 is outlier.

K=2

C1=1 C2=7

After first iteration

C1=2 C2=31.67

As 80 data point which is outlier comes in cluster 2.

Cluster 2 centroid changes to accommodate 80 .

Therefore K means is sensitive to outliers

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.*