MACHINE LEARNING – WORKSHEET II (CLUSTERING) SOLUTIONS

1. d
2. e
3. a
4. a
5. b
6. b
7. a
8. d
9. d
10. a
11. f
12. E

13.

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. here are other clustering algorithms out there that are less sensitive to outliers. To counter this we use algorithms like K-medoids.

14.

K-means is efficient in terms of computing than rest of the algorithms which have better features. Also we can ensure definite converge using K-means algorithm.

15.

Deterministic Algorithms are algorithms which could results similar outputs every time executed on same data.

The basic k means clustering algorithm is an non-deterministic algorithm. This means that every time you run the algorithm you could get different results on same data. The non-deterministic nature of K-Means is due to its random selection of data points as initial centroids