MACHINE LEARNING ASSIGNMENT 2

1. a)
2. d)
3. a)
4. a)
5. b)
6. b)
7. a)
8. d)
9. a)
10. d)
11. d)
12. Yes, K mean 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.
13. K means is better because it is relatively simple to implement. It scales to large data sets and even guarantees convergence. K means can warm start the positions of centroids and easily adapts to new examples.
14. No, K means is not a deterministic algorithm. This means that running the algorithm several times on the same data, could give different results.