WORKSHEET 2 MACHINE LEARNING

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 is sensitive to outlier because it is used to find the mean of the data points in order to locate the center of the cluster. For eg: 4,3,5,12 are data point in cluster and 12 is outlier, when it comes to taking mean if outlier 12 is considered, the mean of cluster will be 6, whereas if outlier was not present then mean would have been 4. Hence the new center would have been different for each new successive cluster.

13) K mean is better because it is simple to implement, works on large dataset and its self improving ability make it easy to adapt to new examples, to identify the unlabelled data. The output can also be easily visualized and interpreted.

14) No, kmean is non deterministic algorithm, since it randomly chooses the starting points. It also means that running the algorithm several times on same dataset can give different results.