Machine Learning Assignment 2

1. Option a
2. Option d
3. Option a
4. Option a
5. Option b
6. Option b
7. Option a
8. Option d
9. Option a
10. Option d
11. Option d

12. The K-means clustering algorithm is sensitive to outliers, because a mean is easily influenced by extreme values. **For example.** The mean of 2,2,2,3,3,3,4,4,4 is 3 if we take the integer value, if we add a single 23 to that, the mean becomes 5 which is larger than any of the other values. Since in k-means, you will be taking the mean ,so you wind up a lot of outlier sensitive calculations.

13. K means is better because it is
    - Relatively simple to implement.
    - Scales to large data sets.
    - Guarantees convergence.
    - Can warm-start the positions of centroids.
    - Easily adapts to new examples.

14. No, k-means is not a deterministic algorithm, it is non-deterministic in nature. K-Means starts with a random set of data points as initial centroids. This random selection influences the quality of the resulting clusters. It gives the different output for the same dataset each time when the algorithm is run.