Machine Learning Worksheet 2

1. a) 2 Only
2. d) 1, 2 and 4
3. a) True
4. a) 1 only
5. b) 1
6. b) No
7. a) Yes
8. d) All of the above
9. a) K-means clustering algorithm
10. d) All of the above
11. d) All of the above

12. K-Means clustering algorithm is most sensitive to outliers as it uses the mean of cluster data points to find the cluster centre.

13. K means is better because advantages of k-means are:
    - Relatively simple to implement.
    - Scales to large data sets.
    - Guarantees convergence.
    - Can warm-start the positions of centroids.
    - Easily adapts to new examples.
    - Generalises to clusters of different shapes and sizes, such as elliptical clusters.
    - Choosing manually.
    - Being dependent on initial values.

14. No, k-means clustering is based on a non-deterministic algorithm. This means that running the algorithm several times on the same data, could give different results.