WORKSHEET 2 - MACHINE LEARNING

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) The K-means clustering technique is sensitive to outliers because a mean is easily influenced by extreme values,. K-medoids clustering is a K-means variation that is more resistant to noise and outliers. Instead of using the mean point as the centre of a cluster, K-medoids represents it with an actual point within the cluster. Medoid is the cluster's most centrally situated object, with the shortest sum of distances to other locations. Mean is heavily influenced by the outlier and hence cannot accurately reflect the cluster centre, but medoid is resilient to the outlier and accurately represents the cluster centre.

13)
- 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.

14) The basic k-means clustering algorithm is non-deterministic. This means that executing the algorithm on the same data multiple times may get different results. The elbow approach is a heuristic used in cluster analysis to determine the number of clusters in a data set. The method entails charting the explained variation as a function of the number of clusters and selecting the curve's elbow as the number of clusters to utilise.

Furthermore, the compiler is unable to solve the problem in polynomial time and is unsure about the next step. This is due to the fact that some problems are extremely random.
These algorithms usually have 2 steps —
1)Guessing step
2)Assignment step
The K-Means algorithm divides the data space into K clusters such that the total variance of all data points with respect to the cluster mean is minimised.