

Machine Learning Worksheet – 2

1. d) 2 and 3
2. d) 1, 2 and 4
3. a) True
4. a) 1 only
5. b) 1
6. a) Yes
7. a) Yes
8. d) All of the above
9. a) K-means clustering algorithm
10. a) 1 only
11. d) All of the above
12. The k-means algorithm is sensitive to the outliers. In this paper, we propose a robust two-stage k-means clustering algorithm based on the observation point mechanism, which can accurately discover the cluster centers without the disturbance of outliers.
13. Guarantees convergence. Can warm-start the positions of centroids. Easily adapts to new examples. Generalizes to clusters of different shapes and sizes, such as elliptical clusters
14. The non-deterministic nature of K-Means is due to its random selection of data points as initial centroids. Method: We propose an improved, density based version of K-Means, which involves a novel and systematic method for selecting initial centroids.