

WORKSHEET 2 – MACHINE LEARNING ANSWERS

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
10. d) All of the above
11. d) All of the above
12. The K-means clustering algorithm is sensitive to outliers, because a mean is easily influenced by extreme values. The group of points in the right form a cluster, while the rightmost point is an outlier.
13. Other clustering algorithms with better features tend to be more expensive. In this case, k-means becomes a great solution for pre-clustering, reducing the space into disjoint smaller sub-spaces where other clustering algorithms can be applied. K-means is the simplest.
14. The basic 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. However, to ensure consistent results, FCS Express performs k-means clustering using a deterministic method.