

# Machine Learning

## ***WORKSHEET SET – 2 SOLUTIONS***

### Questions 1 to 11 Answers

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

### **Subjective Questions**

12. Is K sensitive to outliers?

Yes K-means is sensitive to outliers k-means can be used as outlier detection. In K-means, using the symmetric distance measure which is the key component to define the samples that belongs to the same cluster. symmetric distance measurement gives similar weight to each dimension (feature) this may not always be the case for defining outliers.

For example:

The mean of 2,2,2,3,3,3,4,4,4 is 3

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, we take the mean a lot, and wind up with a lot of outlier-sensitive calculations.

That's why we have the k-medians algorithm. It just uses the median rather than the mean and is less sensitive to outliers.

13. Why is K means better?

Advantages of K-means

1. Relatively simple to implement. 2. Scales to large data sets 3. Guarantees convergence 4. Easily adapts to new examples 5. Can Generalize to clusters of different shapes and sizes, such as elliptical clusters.

14. Is K means a deterministic algorithm?

No K-means is not a 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.

A deterministic function always returns the same results if given the same input values. A nondeterministic function may return different results every time it is called, even when the same input values are provided