Clustering

TOTAL POINTS 15

1. Which statement is **NOT TRUE** about k-means clustering? 3 points

- [ ] k-means divides the data into non-overlapping clusters without any cluster-internal structure.
- [ ] The objective of k-means, is to form clusters in such a way that similar samples go into a cluster, and dissimilar samples fall into different clusters.
- [x] As k-means is an iterative algorithm, it guarantees that it will always converge to the global optimum.

2. Which of the following are characteristics of DBSCAN? Select all that apply. 3 points

- [x] DBSCAN can find arbitrarily shaped clusters.
- [x] DBSCAN can find a cluster completely surrounded by a different cluster.
- [x] DBSCAN has a notion of noise, and is robust to outliers.
- [x] DBSCAN does not require one to specify the number of clusters such as k in k-means

3. Which of the following is an application of clustering? 3 points

- [ ] Customer churn prediction
- [ ] Price estimation
- [x] Customer segmentation
- [ ] Sales prediction

4. Which approach can be used to calculate dissimilarity of objects in clustering?

3 points

- ○ Minkowski distance
- ○ Euclidian distance
- ○ Cosine similarity
- ◉ All of the above

5. How is a center point (centroid) picked for each cluster in k-means?

3 points

- ☑ We can randomly choose some observations out of the data set and use these observations as the initial means.
- ☑ We can create some random points as centroids of the clusters.
- ☐ We can select it through correlation analysis.