Introduction to Data Science

Instructor: Daniel D. Gutierrez

MODULE 10 QUIZ

Question 1

Which one of the following best describes unsupervised machine learning?

(a) Provides great potential for discovering previously known insights from unlabeled data sets

(b) A primary method is called clustering

(c) Not focused with making predictions

(d) Trying to understand the structure of underlying data

(e) All of the above

Question 2

Defining "closeness" is the most important step when using clustering methods. Which one of the following describes a distance metric?

(a) The distance metric you choose is independent of the data you have

(b) A distance measure is restricted to 2-dimensions (e.g. longitude and latitude)

(c) Similarity measurements are substantively different from distance metrics

(d) A common distance metric for continuous variables is Euclidean distance

Question 3

Which one of the following characterizations describes hierarchical clustering?

(a) Hierarchical clustering is an agglomerative (bottom-up) approach toward the clustering process

(b) A defined distance measure is not required for hierarchical clustering

(c) Since its complexity is high, hierarchical clustering is typically used when the number of data points is not too great

(d) (a) and (c) only

(e) All of the above

Question 4

Which one of the following characterizations describe K-means?

(a) K-means is partitioning approach toward the clustering process

(b) Requires an initial number of clusters

(c) Requires an initial guess for initial "centroids" or cluster centers

(d) All of the above

Question 5

The following describe aspects of data visualization for clustering methods in R EXCEPT which one?

(a) Heat maps are not useful for clustering methods

(b) The `cluster` data item within the `kmeans` object is used to plot clusters

(c) The `cutree()` function can be used to visualize different clusters with hierarchical clustering

(d) The `centers` data item within the `kmeans` object is used to plot the cluster centers

(e) The `hclust` object can be used with the `plot()` function to visualize hierarchical clusters