

Question 1

Which of the following is an example of clustering?

- a. Separate the data into distinct groups by similarity
- b. Creating a new representation of the data with fewer features
- c. Compress elongated clouds of data into more spherical representations
- d. Accumulate data into groups based on labels

Question 2

In clustering evaluation process, "elbow point" is where the rate of accuracy increase sharply, when we run clustering multiple times, increasing k in each run.

- a. True
- b. False

Question 3

In which type of dissimilarity calculation between clusters, we find the longest distance between points in each cluster?

- a. Single-Linkage Clustering
- b. Centroid Linkage Clustering
- c. Complete-Linkage Clustering
- d. Average-Linkage Clustering

Question 4

Which of the following are the characteristics of density-based clustering? Select all that apply.

- a. Density-based clustering algorithms have no notion of outliers.
- b. Density-based clustering algorithms are proper for arbitrary shape clusters.
- c. Density-based clustering algorithms locate regions of high density that are separated from one another by regions of low density.

Question 5

Any initialization of the centroids in k-means is just as good as any other.

- a. True
- b. False

Question 6

Which of the following statements are true? Select all that apply.

- a. For some datasets, the "right" or "correct" value of K (the number of clusters) can be ambiguous, and hard even for a human expert looking carefully at the data to decide.
- b. Since K-Means is an unsupervised learning algorithm, it cannot overfit the data, and thus it is always better to have as large a number of clusters as is computationally feasible.
- c. If we are worried about K-means getting stuck in bad local optima, one way to ameliorate (reduce) this problem is if we try using multiple random initializations.
- d. The standard way of initializing K-means is setting all the centroids locations to be equal to a vector of zeros.

Question 7

Which of the following is a reasonable way to select the number of principal components k ? (Note that m is the number of input examples.)

- a. Choose to use the elbow method.
- b. Choose k to be 99% of m (i.e., $k = 0.99*m$, rounded to the nearest integer).
- c. Choose k to be the largest value so that at least 99% of the variance is retained.
- d. Choose k to be the smallest value so that at least 99% of the variance is retained.

Question 8

Which of the following statements are true? Check all that apply.

- a. Given an input vector x , PCA compresses it to a lower-dimensional vector z .
- b. If the input features are on very different scales, it is a good idea to perform feature scaling before applying PCA.
- c. PCA can be used only to reduce the dimensionality of data by 1 (such as 3D to 2D, or 2D to 1D).
- d. Feature scaling is not useful for PCA, since the eigenvector calculation takes care of this automatically.

Question 9

PCA components are always orthogonal

- a. True
- b. False

Question 10

Which of the following are good/recommended applications of PCA? Select all that apply.

- a. To visualize high-dimensional data (by choosing $k = 2$ or $k = 3$)
- b. To reduce the dimension of the input data so as to speed up a learning algorithm
- c. Instead of using regularization, use PCA to reduce the number of features to reduce overfitting
- d. To compress the data so it takes up less computer memory/disk space