DLI Accelerated Data Science Teaching Kit

Module 15 Quiz:
Machine Learning: Clustering and Dimensionality Reduction

1. The rate of convergence of KMeans may be affected by the algorithm's initialization.

A. True
B. False

Answer: A

2. DBSCAN is a popular clustering technique because it uses many hyperparameters.

A. True
B. False

Answer: B

3. In DBSCAN, if a point A is density-reachable from point B, then B is also density-reachable from A.

A. True
B. False

Answer: B

4. Hyperparameters in t-SNE are not important because t-SNE can produce stable, consistent results over a wide range of hyperparameter values.

A. True
B. False

Answer: B

5. Cluster sizes in t-SNE is meaningful.

A. True
B. False

Answer: B

6. When using t-SNE, it is a good idea to generate one plot, as long as the algorithm runs until convergence.

A. True
B. False

Answer: B

7. Unsupervised learning needs no labels during training process.

A. True
B. False

Answer: A

Explanation: See lecture 15.1 slides

8. For KMeans clustering, K value can be learned automatically.

A. True
B. False

Answer: B

Explanation: See lecture 15.2 slides

9. Which of following strategies can be used as convergence criterion?

A. No (or minimum) re-assignments of data points to different clusters.
B. No (or minimum) change of centroids.
C. Minimum decrease in the sum of squared error (SSE).
D. All samples have been assigned to a centroid.

Answer: ABC

Explanation: See lecture 15.2 slides

10. Which method can be used to process missing values?

A. Dummy substitution
B. Mean substitution
C. Supervised learning
D. Frequent substitution

Answer: ABD

Explanation: See lecture 3.2 slides

11. What are the strengths of KMeans?

A. Easy to understand
B. Interpretable
C. Low time complexity
D. High performance on big data

Answer: ABC

Explanation: See lecture 15.2 slides