

🎉 Congratulations! You passed!

Grade
received 90%

Latest Submission
Grade 90%

To pass 80% or
higher

Go to next item

1. Which of the following statements correctly describe key aspects of k-means? Select all that apply.

0.5 / 1 point

- ☐ K-means organizes data into clusters by creating a logical scheme to make sense of it.
- ☒ The clustering process has four steps that repeat until the model disperses evenly.

❌ This should not be selected
Review [the video that introduces k-means](#) ↗.

- ☒ K-means groups unlabeled data into k clusters based on their similarities.

✔ Correct

- ☒ Poor clustering can be caused by local minima, which means the model has converged in a sub-optimal way.

✔ Correct

2. Which of the following is NOT a step of the k-means algorithm?

1 / 1 point

- ☐ Initiate k centroids
- ☐ Assign all points to their nearest centroid
- ☒ Calculate the mean number of points per centroid
- ☐ Repeat steps two and three until the model converges

✔ Correct

3. Fill in the blank: In order to evaluate the ____ space in a k-means model, a data professional uses the inertia metric. This is the sum of the squared distances between each observation and its nearest centroid.

1 / 1 point

- ☐ midpoint
- ☐ intercluster
- ☐ converged
- ☒ intracluster

✔ Correct

4. Which of the following statements accurately describe agglomerative clustering? Select all that apply.

1 / 1 point

- ☒ There are numerous hyperparameters available for agglomerative clustering.

✔ Correct

- ☒ Agglomerative clustering works by first assigning every point to its own cluster, then progressively combining clusters based on intercluster distance.

✔ Correct

- ☒ The algorithm will stop when the specified number of clusters is met.

✔ Correct

- ☐ The algorithm will stop before an intercluster distance threshold is reached.

5. Which type of linkage determines whether to merge clusters by considering the distance between each cluster's centroid and the other clusters' centroids?

1 / 1 point

- ☒ Average
- ☐ Single
- ☐ Complete
- ☐ Ward

✔ Correct

6. A data analyst creates a k-means model. They examine the silhouette coefficient of an observation and find it to have a value close to negative one. What conclusion should they draw in this scenario?

1 / 1 point

- ☒ The observation may be in the wrong cluster.
- ☐ The observation is on the boundary between clusters.
- ☐ The observation is in the correct cluster.
- ☐ The observation is suitably within its own cluster and well separated from other clusters.

✔ Correct

7. How would a data professional use inertia to evaluate the number of clusters in their data?

1 / 1 point

- ☐ Plot the silhouette score for different values of k to determine where the elbow is
- ☐ Choose the number of clusters that results in the lowest inertia
- ☒ Plot the inertia for different values of k to determine where the elbow is
- ☐ Choose the number of clusters that results in the highest inertia

✔ Correct

8. Which of the following statements accurately describe the elbow method? Select all that apply.

1 / 1 point

- ☒ The sharpest bend in the curve is usually the model that will provide the most meaningful clustering of data.

✔ Correct

- ☒ The elbow method uses a line plot to visually compare the inertias of different models.

✔ Correct

- ☐ When using the elbow method, data professionals aim to find the smoothest part of the curve.

☒ There is not always an obvious elbow.

✔ Correct

9. You want to use an algorithm to group these data points. Which of the following statements are true? Select all that apply.

0.75 / 1 point



☒ DBSCAN would probably perform well to cluster this data, because the DBSCAN algorithm uses data density to determine cluster membership, not Euclidean distance from centroids.

✔ Correct

☐ Running a k-means model with $k=4$ would result in lower inertia than a model with $k=3$.

☒ Running a k-means model with $k=3$ would result in a greater silhouette score than a model with $k=2$.

✔ Correct

☒ Using k-means to cluster this data could be sub-optimal because it works using distance from centroids, and therefore is best used on clusters that are round.

✔ Correct

You didn't select all the correct answers

10. Which of the following scenarios make use of unsupervised learning? Select all that apply.

0.75 / 1 point

☒ A marketing department uses a clustering model to identify groups of similar customers.

✔ Correct

☐ A scientist builds a model that uses historical weather data to predict the amount of rainfall.

☐ Developers at Google build a text completion model to suggest the next word you'll type in the search engine.

☒ A data scientist develops a recommendation system to serve music on a streaming service.

✔ Correct

You didn't select all the correct answers