← Go Back to Machine Learning

Course Content

Weekly Quiz - K-means Clustering

Type	:	Graded Quiz
Attempts	:	1/1
Questions	:	10
Time	:	45m
Due Date	:	Jul 01, 4:30 AM EDT
Your Marks	:	10/10

Instructions

Attempt History

Attempt #1
Jun 30, 8:34 PM

Marks: 10

Q No: 1 Correct Answer

Marks: 1/1

When doing K-means clustering, what will be the Euclidean distance of a point A(4,0) be from the centroid of the cluster which has two data points (3,3) and (5,5)?

- 2
- 3
- 4 — You Selected
- 16

Cluster centroid C1 = {((3+5)/2) , ((3+5)/2)} = {4,4}

Distance between the point A{4,0} and centroid C1{4,4}

Euclidean distance = $\sqrt{(4-4)^2 + (4-0)^2}$ = 4

Q No: 2 Correct Answer

Marks: 1/1

For K-means clustering, what will be the cluster centroids for the following 2 clusters?

C1: {(3,5),(5,4), (4,6)}

C2: {(6,0),(8,1), (7,2}

- C1: {5,4} C2: {7,1}
- C1: {4,5} C2:{7,1} — You Selected
- C1: {60,120} C2: {168,2}
- C1: {12, 15} C2: {21, 3}

C1 = {(3+5+4)/3 , (5+4+6)/3} = {4,5}

C2 = {(6+8+7)/3, (0+1+2)/3} = {7,1}

Q No: 3 Correct Answer

Marks: 1/1

While using K-means clustering, we scale the variables before we do clustering. This is done primarily to

- treat missing values to make the data more robust for analysis
- convert the data to same scale hence variables which are of different units are given equal importance — You Selected
- avoid multicollinearity among the variables
- make the model less susceptible to outliers

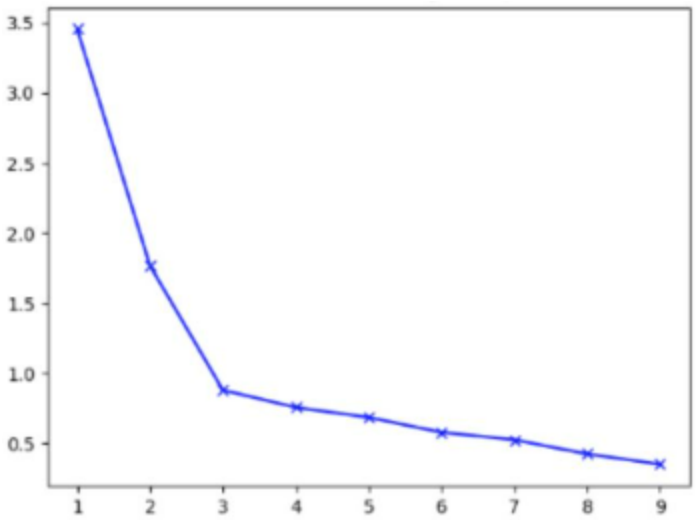
Scaling the data brings all the attributes to similar scale which makes equal importance to all the attributes while performing clustering

---

Q No: 4 Correct Answer

Marks: 1/1

Consider the following elbow plot:



While performing K-means clustering, what is the ideal value of K to choose based on the above plot?

- 4
- 2
- 1
- 3 — You Selected

As the slope decreases drastically from 2 to 3 and at cluster point 3, the graph takes a sharp turn, 3 is considered to be the ideal no. of clusters.

Q No: 5 Correct Answer

Marks: 1/1

What does the predict() function of the sklearn KMeans class return?

- The closest cluster to which a data point belongs — You Selected
- The distance between each data point and the cluster centroids.
- Gives the position of cluster centroids
- Number of clusters

The predict()Links to an external site. function is used to predict the closest cluster to which a data point belongs.

Q No: 6 Correct Answer

Marks: 1/1

Which of the following is considered to be the weakness of K-means clustering?

1. Finding out the ideal value of K is complex and time-consuming
2. Susceptible to the curse of dimensionality
3. Not sensitive to starting positions of the initial centroid
4. Not sensitive to outliers

- ○ 3 and 4
- ◉ 1 and 2 — You Selected
- ○ 2 and 4
- ○ 1 and 3

1. Finding an ideal value of K requires multiple iterations with different values of K to see which value of K has the lowest within-cluster sum of squared errors.
2. K-means clustering is considered to be affected by the curse of dimensionality. As the no. of dimensions increases, the computational complexity of K-means clustering increases.
3. K-means clustering is considered to be sensitive to the starting position as this determines the position of the centroids of the clusters.
4. K-means clustering is sensitive to outliers. Outliers significantly affect the position of the centroid in K-means clustering.

Q No: 7 Correct Answer

Marks: 1/1

Which of the following is NOT true in the case of K-means clustering?

- (●) The data points that are the farthest from a centroid will create a cluster centered around that centroid — You Selected
- ( ) It requires the number of clusters to be specified
- ( ) K-means clusters data by separating data points into group based on distance from cluster centroid.
- ( ) Choosing different starting points can result in different clusters

The data points that are the closest to a centroid will create a cluster. If we're using the Euclidean distance between data points and every centroid, a straight line is drawn between two centroids, then a perpendicular bisector (boundary line) divides this line into two clusters

---

Q No: 8 Correct Answer

Marks: 1/1

In K-means clustering, suppose the number of clusters is equal to the number of data points (observations). Then what will be the sum of squared errors within each group (or cluster)?

- (●) 0 — You Selected
- ( ) Approaches infinity (very large number)
- ( ) 1
- ( ) Cannot be determined

With increase in the no of clusters, ideally, the within group sum of squared errors decreases. As the no of clusters increase and is equal to the no of observations the within group sum of squared errors becomes zero.

When there is only one data point in a cluster, the data point itself becomes the centroid. Distance from the point to the centroid is always zero

---

Q No: 9 Correct Answer

Marks: 1/1

Which of the following are Unsupervised Learning techniques?

1. Hierarchical Clustering
2. Random Forests
3. K-means Clustering
4. Logistic Regression

- 3 and 4
- 2 and 4
- 1 and 3 — You Selected
- 1 and 2

Both forms of clustering (K-means and Hierarchical) are considered to be forms of Unsupervised Learning as we don't categorize the data into dependent and independent variables before clustering.

As for Random Forests and Logistic Regression, we separate the data into dependent and independent variables before applying the algorithms. So, they are Supervised Learning techniques.

Q No: 10 Correct Answer

Marks: 1/1

What is the default value of n_clusters in sklearn.cluster.KMeans, the K-means clustering class in Scikit-learn?

- 6
- 1
- 2
- 8 — You Selected

Previous

.cluster.KMeansLinks to an external site., the n_clusters is an optional parameter that takes an integer value specifying the number of clusters to form as well as the number of centroids to generate. The default value is 8.

Next