



Score: 100%

No. of questions: 11

Correct answer: 11

Incorrect answer: 0

Show incorrect attempt only ☒

Question 1

1 Mark

k-NN algorithm does more computation on test time rather than train time.

A True



B False

Correct Answer: A. True

The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples. In the testing phase, a test point is classified by assigning the label which is most frequent among the k training samples nearest to the query point and hence, higher computation.

Question 2

2 Marks

Which of the following will be true about k in k-NN in terms of bias?

A When you increase the k, the bias will increase



Prev
Video

Next
Video

B (machine-learning/screen/6172) When you decrease the k, the bias will increase

(machine-learning/screen/6173)



Correct Answer: A. When you increase the k, the bias will increase

Large k means a simple model and a simple model always leads to high bias.

Question 3

3 marks

Suppose you are working with a dataset that has both continuous and categorical variables. Which distance metric can we use in this case?

A Cannot be calculated

B Sum of Hamming and Euclidean distance

C Sum of Hamming and Manhattan distance

D Both B and C



Correct Answer: D. Both B and C

Euclidean and Manhattan distance are valid only for continuous variables whereas Hamming distance is used for categorical variables. In our case, distance metric can be either sum of Hamming and Euclidean or sum of Hamming and Manhattan, since the dataset contains both continuous and categorical. Hence, option D is correct.

Question 4

1 Mark

Why is %matplotlib inline used?



Previous question
To plot graphs within the browser itself
Video

(machine-learning/screen/6172)

Next question
Video



(machine-learning/screen/6173)

B To save plots to disk



Module Test



H

C To stop plots from being displayed

Correct Answer: A. To plot graphs within the browser itself

It sets the back-end of matplotlib to inline. Hence, the plots are displayed right below the cell that produced it.

Question 5

2 Marks

Which of the following will be the Euclidean distance between the two data points?

A(1,3) and B(2,3)

A 1



B 2

C 4

D 8

Correct Answer: A. 1

Euclidean distance will be: $\sqrt{(1-2)^2 + (3-3)^2}$
 $= \sqrt{1^2 + 0^2} = 1$

Question 6

3 marks

What would be the Hamming distance for the given strings?

String1: 001010100
String2: 110101010
(machine-learning/screen/6172)

Next
Video
(machine-learning/screen/6173)



8

Module Test



H



B

9

C

1

D

2

Correct Answer: A. 8

Hamming distance between String1 and String2 is the number of places where they differ. In our case, these strings differ in 8 places. Hence, option A is correct.

Question 7

1 Mark

Which of the following is true about Manhattan distance?

A

It can be used for continuous variables



B

It can be used for categorical variables

C

It can be used for categorical as well as continuous

D

None of the above

Correct Answer: A. It can be used for continuous variables

Manhattan distance is designed for calculating the distance between real valued features.



B

Minkowski

C

Jaccard

D

All of the above



Correct Answer: D. All of the above

All of these distance metrics can be used as a distance metric for k-NN.

Question 9

3 marks

In k-NN what will happen when you increase/decrease the value of k?

A

The boundary becomes smoother with increasing value of K



B

The boundary becomes smoother with decreasing value of K

C

Smoothness of boundary doesn't depend on the value of K

D

None of the above

Correct Answer: A. The boundary becomes smoother with increasing value of K

As the value of k increases, the final prediction takes into account a larger number of data points. Hence the decision boundary would become smoother.



It can be used for classification.

Module Test



H

B

It can be used for regression.

C

It can be used in both classification and regression.



Correct Answer: C. It can be used in both classification and regression.

We can also use k-NN for regression problems. In this case, the prediction can be based on the mean or the median of the k-most similar instances.

Question 11

2 Marks

What does stratify = True in the train_test_split function do?

A

Makes sure that test data and train data have the same distribution of classes.



B

Equally divides the data into train and test subsets.

Correct Answer: A. Makes sure that test data and train data have the same distribution of classes.

Stratification is the process of splitting the dataset into training and test dataset such that the proportion of classes are preserved. This is helpful in case of class imbalance problems. Changing stratify = True is used to achieve this process.



Prev

Video

(machine-learning/screen/6172)

Next

Video



(machine-learning/screen/6173)