Question 11 is compulsory.

Attempt any 5 of the first 10

Question 1)

Which of the following sentence is FALSE regarding regression? (

A) It relates inputs to outputs.

B) It is used for prediction.

C) It may be used for interpretation.

D) It discovers causal relationships.

Answer: D

Question 2)

You observe the following while fitting a linear regression to the data: As you increase the amount of training data, the test error decreases and the training error increases. The train error is quite low (almost what you expect it to), while the test error is much higher than the train error. What do you think is the main reason behind this behavior. Choose the most probable option.

A) High variance

B) High model bias

C) High estimation bias

D) None of the above

Answer: A

Question 3)

The most widely used metrics and tools to assess a classification model are:

1. Confusion matrix
2. Cost-sensitive accuracy
3. Area under the ROC curve
4. All of the above

Answer: D

Question 4)

Which of the following is a good test dataset characterstic?

 Large enough to yield meaningful results

 Is representative of the dataset as a whole

 Both A and B

 None of the above

Answer: C

Question 5)

Which of the following is a disadvantage of decision trees?

 Factor analysis

 Decision trees are robust to outliers

 Decision trees are prone to be overfit

 None of the above

Answer: C

Question 6)

What is the purpose of performing cross-validation?

 To assess the predictive performance of the models

 To judge how the trained model performs outside the sample on test data

 Both A and B

Answer: C

Question 7)

Which of the following statements about regularization is not correct?

 Using too large a value of lambda can cause your hypothesis to underfit the data.

 Using too large a value of lambda can cause your hypothesis to overfit the data.

 Using a very large value of lambda cannot hurt the performance of your hypothesis.

 None of the above

Answer: D

Question 8)

How can you prevent a clustering algorithm from getting stuck in bad local optima?

 Set the same seed value for each run

 Use multiple radom initializations

 Both A and B

 None of the above

Answer: B

Question 9)

In which of the following cases will K-means clustering fail to give good results? 1) Data points with outliers 2) Data points with different densities 3) Data points with nonconvex shapes

 1 and 2

 2 and 3

 1, 2, and 3

 1 and 3

Answer: C

Question 10)

Which of the following methods do we use to find the best fit line for data

in Linear Regression?

A) Least Square Error

B) Maximum Likelihood

C) Logarithmic Loss

D) Both A and B

Answer: A

Question 11)

This dataset contains information about used cars listed on www.cardekho.com This data can be used for a lot of purposes such as price prediction to exemplify the use of linear regression in Machine Learning. The columns in the given dataset is as follows:

1. Car\_Name
2. Year
3. Selling\_Price
4. Present\_Price
5. Kms\_Driven
6. Fuel\_Type
7. Seller\_Type
8. Transmission
9. Owner

**Predict Price – Dataset is up here:** https://drive.google.com/file/d/1V2kd9cI\_suLndqYpsyUg1KqTPAGy6lCP/view?usp=sharing