Machine learning Assignment

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?

1. A) Least Square Error
2. B) Maximum Likelihood
3. C) Logarithmic Loss
4. D) Both A and B

Answer:

1. A

2. Which of the following statement is true about outliers in linear regression?

1. A) Linear regression is sensitive to outliers

2. B) Linear regression is not sensitive to outliers
3. C) Can’t say
4. D) None of these

Answer:

1. A

3. A line falls from left to right if a slope is _____?

1. A) Positive
2. B) Negative
3. C) Zero
4. D) Undefined

Answer:

2. B

4. Which of the following will have symmetric relation between dependent variable and independent variable?

1. A) Regression
2. B) Correlation
3. C) Both of them

4. D) None of these

Answer:

2. B

5. Which of the following is the reason for overfitting condition?

1. A) High bias and high variance
2. B) Low bias and low variance
3. C) Low bias and high variance
4. D) None of these

Answer:

3. C

6. If output involves label then that model is called as:

1. A) Descriptive model
2. B) Predictive model
3. C) Reinforcement learning
4. D) All of the above

Answer:

2. B

7. Lasso and Ridge regression techniques belong to _____?

1. A) Cross validation
2. B) Removing outliers
3. C) SMOTE
4. D) Regularization

Answer:

4. D

8. To overcome with imbalance dataset which technique can be used?

1. A) Cross validation
2. B) Regularization
3. C) Kernel
4. D) SMOTE

Answer:

4. D

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary

classification problems. It uses _____ to make graph?

1. A) TPR and FPR
2. B) Sensitivity and precision
3. C) Sensitivity and Specificity
4. D) Recall and precision

Answer:

1. A

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.

1. A) True
2. B) False

Answer:

2. B

11. Pick the feature extraction from below:

1. A) Construction bag of words from an email

2. B) Apply PCA to project high dimensional data
3. C) Removing stop words
4. D) Forward selection

Answer:

2. B

In Q12, more than one option is correct, choose all the correct options:

12. Which of the following is true about the Normal Equation used to compute the coefficient of the Linear Regression?

1. A) We don’t have to choose the learning rate.
2. B) It becomes slow when the number of features is very large.
3. C) We need to iterate.
4. D) It does not make use of dependent variable.

Answers:

1. A

2. B

13. Explain the term regularization?

Regularization is a way to make sure a machine learning model doesn’t become too complicated and only works well on the training data. It adds a small penalty to the model to keep it simple and more likely to work well on new data.

14. Which particular algorithms are used for regularization?

Some common algorithms for regularization are:

- **Lasso (L1 Regularization)**: Adds a penalty to make some model coefficients zero, which can help in feature selection.

- **Ridge (L2 Regularization)**: Adds a penalty to make the coefficients smaller, but not zero.
- **Elastic Net**: Combines both Lasso and Ridge penalties.

15. Explain the term error present in linear regression equation?

In linear regression, the error is the difference between the actual value and the value predicted by the model. It shows how far off the model's prediction is from the actual data. The goal is to make these errors as small as possible.