

**MACHINE LEARNING**

In Q1 to Q11, only one option is correct, choose the correct option:

1. 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

**ANS: - (A) Least Square Error**

2. Which of the following statement is true about outliers in linear regression?
- A) Linear regression is sensitive to outliers
  - B) linear regression is not sensitive to outliers
  - C) Can't say
  - D) none of these

**ANS: - (A) Linear regression is sensitive to outliers**

3. A line falls from left to right if a slope is\_\_\_\_\_?
- A) Positive
  - B) Negative
  - C) Zero
  - D) Undefined

**ANS: - (B) Negative**

4. Which of the following will have symmetric relation between dependent variable and independent variable?
- A) Regression
  - B) Correlation
  - C) Both of them
  - D) None of these

**ANS: - (B) Correlation**

5. Which of the following is the reason for over fitting condition?
- A) High bias and high variance
  - B) Low bias and low variance
  - C) Low bias and high variance
  - D) none of these

**ANS: - (C) Low bias and high variance**

6. If output involves label then that model is called as:
- A) Descriptive model
  - B) Predictive modal
  - C) Reinforcement learning
  - D) All of the above

**ANS: - (B) Predictive modal**

7. Lasso and Ridge regression techniques belong to\_\_\_\_\_?
- A) Cross validation
  - B) Removing outliers
  - C) SMOTE
  - D) Regularization

**ANS: - (D) Regularization**

**MACHINE LEARNING**

8. To overcome with imbalance dataset which technique can be used?
- A) Cross validation
  - B) Regularization
  - C) Kernel
  - D) SMOTE

**ANS: - (D) SMOTE**

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses \_\_\_\_\_ to make graph?
- A) TPR and FPR
  - B) Sensitivity and precision
  - C) Sensitivity and Specificity
  - D) Recall and precision

**ANS: - (A) TPR and FPR**

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

**ANS: - (A) True**

11. Pick the feature extraction from below:
- A) Construction bag of words from a email
  - B) Apply PCA to project high dimensional data
  - C) Removing stop words
  - D) Forward selection

**ANS: - (B) Apply PCA to project high dimensional data**

**In Q12, more than one options are correct, choose all the correct options:**

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?
- A) We don't have to choose the learning rate.
  - B) It becomes slow when number of features is very large.
  - C) We need to iterate.
  - D) It does not make use of dependent variable.

**ANS: - (A) We don't have to choose the learning rate.**  
**(B) It becomes slow when number of features is very large.**

---

## **MACHINE LEARNING**

**Q13 and Q15 are subjective answer type questions, Answer them briefly.**

13. Explain the term regularization?

**ANS: -**

Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting. There are 2 techniques to use regularization (1) LASSO (2) Ridge. This happens because our model is trying too hard to capture the noise in our training dataset. By noise we mean the data points that don't really represent the true properties of our data but random chance. Learning such data points makes our model more flexible at the risk of overfitting.

14. Which particular algorithms are used for regularization?

**ANS: -**

There are 3 algorithms to use regularization (1) LASSO (2) Ridge (3) Dropout

1. LASSO

It adds absolute value of magnitude of coefficient as penalty term to the loss function.

2. Ridge

It is one of the type of liner regression in which small amount of bias is introduced so that we can get better long predictions.

3. Dropout

Dropout is a regularization algorithms used in neural networks It prevents complex co-adaptations from neurons. Dropout decreases overfitting by avoiding training all the neurons on the complete training data in one go.

Among this three mostly the LASSO and Ridge algorithms used for regularization.

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

**ANS: -**

Error term is a residual variable produced by a statistical or mathematical model which is created when the model does not fully represent the actual relationship between the independent variables and the dependent variables. An error term appears in a statistical model like a regression model to indicate the uncertainty in the model.

Within a linear regression model tracking a stock's price over time, the error term is the difference between the expected prices at a particular time and the price that was actually observed. In instances where the price is exactly what was anticipated at a particular time, the price will fall on the trend line and the error term will be zero.

---