

**MACHINE LEARNING**

\*\*\*\*\*

**Name:-Poonam Gawade****Batch:-1838**

\*\*\*\*\*

**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

**Answer:- 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

**Answer:- 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

**Answer:- 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

**Answer:- 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

**Answer:- 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

**Answer:-Predictive model**

**MACHINE LEARNING**

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

**Answer:- Regularization**

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

**Answer:- 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

**Answer:- 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

**Answer:- False**

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

**Answer:- 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.

**Answer:- We don't have to choose the learning rate.  
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?**

**Answer:-**

Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting.

Regularization is one of the most important concepts of machine learning. It is a technique to prevent the model from overfitting by adding extra information to it.

Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.

Sometimes the machine learning model performs well with the training data but does not perform well with the test data. It means the model is not able to predict the output when deals with unseen data by introducing noise in the output, and hence the model is called overfitted. This problem can be deal with the help of a regularization technique.

This technique can be used in such a way that it will allow to maintain all variables or features in the model by reducing the magnitude of the variables. Hence, it maintains accuracy as well as a generalization of the model.

It mainly regularizes or reduces the coefficient of features toward zero. In simple words, "In regularization technique, we reduce the magnitude of the features by keeping the same number of features."

### **14. Which particular algorithms are used for regularization?**

**Answer:-**

#### **Regularization Techniques :-**

There are two main types of regularization techniques: Ridge Regularization and Lasso Regularization.

#### **1) Ridge Regularization :**

Ridge regression is one of the types of linear regression in which a small amount of bias is introduced so that we can get better long-term predictions.

Ridge regression is a regularization technique, which is used to reduce the complexity of the model. It is also called as L2 regularization.

It modifies the over-fitted or under fitted models by adding the penalty equivalent to the sum of the squares of the magnitude of coefficients.

---

## **MACHINE LEARNING**

This means that the mathematical function representing our machine learning model is minimized and coefficients are calculated. The magnitude of coefficients is squared and added. Ridge Regression performs regularization by shrinking the coefficients present.

### **2) Lasso Regression**

Lasso regression is another regularization technique to reduce the complexity of the model. It stands for Least Absolute and Selection Operator. It is also called as L1 regularization.

It modifies the over-fitted or under-fitted models by adding the penalty equivalent to the sum of the absolute values of coefficients.

Lasso regression also performs coefficient minimization, but instead of squaring the magnitudes of the coefficients, it takes the true values of coefficients. This means that the coefficient sum can also be 0, because of the presence of negative coefficients.

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

**Answer:-**

A Linear Regression model's main aim is to find the best fit linear line and the optimal values of intercept and coefficients such that the error is minimized. Error is the difference between the actual value and Predicted value and the goal is to reduce this difference.

The error term is also known as the residual, disturbance, or remainder term, and is variously represented in models by the letters  $e$ ,  $\epsilon$ , or  $u$ .

An error term essentially means that the model is not completely accurate and results in differing results during real-world applications.

An error term represents the margin of error within a statistical model; it refers to the sum of the deviations within the regression line, which provides an explanation for the difference between the theoretical value of the model and the actual observed results. The regression line is used as a point of analysis when attempting to determine the correlation between one independent variable and one dependent variable.

$$Y = \alpha X + \beta \rho + \epsilon$$

where:  $\alpha, \beta$  = Constant parameters

$X, \rho$  = Independent variables

$\epsilon$  = Error term

Within a linear regression model tracking a stock's price over time, the error term is the difference between the expected price 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.

---