

## 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
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
3. A line falls from left to right if a slope is\_\_\_\_\_?  
A) Positive B) **Negative**  
C) Zero D) Undefined
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
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
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
7. Lasso and Ridge regression techniques belong to\_\_\_\_\_?  
A) Cross validation B) Removing outliers  
C) SMOTE D) **Regularization**
8. To overcome with imbalance dataset which technique can be used?  
A) Cross validation B) Regularization  
C) Kernel 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
10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.  
A) True B) **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

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.
-

## **MACHINE LEARNING**

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

**13. Explain the term regularization?**

Answer-

Regularization is a technique used in machine learning to prevent overfitting, which occurs when a model learns the training data too well and fails to generalize to new, unseen data. Regularization involves adding a penalty term to the model's objective function during training. This penalty term discourages the model from fitting the noise in the training data too closely and encourages it to find simpler patterns that generalize better to new data. In essence, regularization helps in controlling the complexity of the model and reducing the risk of overfitting.

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

Answer-

There are several algorithms used for regularization, but two of the most commonly used ones are:

- A) Ridge Regression (L2 regularization): It adds the square of the magnitude of coefficients to the loss function during training. This encourages the model to keep the coefficients small and prevents them from becoming too large, which helps in reducing overfitting.
- B) Lasso Regression (L1 regularization): It adds the absolute value of coefficients to the loss function during training. Lasso Regression not only helps in reducing overfitting but also has the additional benefit of performing feature selection by shrinking some coefficients to zero, effectively removing less important features from the model.

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

Answer-

The error present in the linear regression equation refers to the difference between the actual values of the dependent variable (target) and the predicted values generated by the linear regression model. The goal is to minimize this error by finding the best-fit line that represents the relationship between the independent and dependent variables.

---