

# MACHINE LEARNING

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

- 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
- 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
- A line falls from left to right if a slope is \_\_\_\_\_?  
A) Positive B) **Negative**  
C) Zero D) Undefined
- 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
- 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
- If output involves label then that model is called as:  
A) Descriptive model B) **Predictive modal**  
C) Reinforcement learning D) All of the above
- Lasso and Ridge regression techniques belong to \_\_\_\_\_?  
A) Cross validation B) Removing outliers  
C) SMOTE D) **Regularization**
- To overcome with imbalance dataset which technique can be used?  
A) Cross validation B) Regularization  
C) Kernel D) **SMOTE**
- 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
- In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.  
A) True B) **False**
- 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?
14. Which particular algorithms are used for regularization?
15. Explain the term error present in linear regression equation?

### **13. Explain the term regularization?**

Answer: When we use regression models to train some data, there is a good chance that the model will overfit the given training data set. Regularization helps sort overfitting problem by restricting the degree of freedom of a given equation. i.e. simply reducing the number of degrees of polynomial function by reducing their corresponding weights.

In a linear equation we do not want huge weights/coefficients as a small change in weight can make a large difference for the dependant variable (Y). So, the regularization constraints the weight of such features to avoid overfitting.

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

Answer: Ridge Regression, Lasso (Least Absolute Shrinkage and Selection Operator) and Elastic-Net Regression.

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

Linear regression most often uses mean-square error (MSE) to calculate the error of the model. MSE is calculated by:

measuring the distance of the observed y-values from the predicted y-values at each value of x;

squaring each of these distances;

calculating the mean of each of the squared distances.

Linear regression fits a line to the data by finding the regression coefficient that results in the smallest MSE.

---