# MACHINE LEARNING

**Batch 1835 (Internship 22) bhushan pise**

**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**
2. Which of the following statement is true about outliers in linear regression?
   **A) Linear regression is sensitive to outliers**
3. A line falls from left to right if a slope is_____?
   **B) Negative**
4. Which of the following will have symmetric relation between dependent variable and independent variable?
   **B) Correlation**
5. Which of the following is the reason for over fitting condition?
   **B) Low bias and low variance**
6. If output involves label then that model is called as:
   **C) Reinforcement learning**
7. Lasso and Ridge regression techniques belong to_____?
   **D) Regularization**
8. To overcome with imbalance dataset which technique can be used?
   **A) Cross validation**
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(true positive rate and false positive rate)**
10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
    **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
    (**DON'T'know**)

**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) It becomes slow when number of features is very large.**

    B) **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?

**ANS-** Regularization refers to techniques that are used to calibrate machine learning and which minimize the adjusted loss function and prevent over fitting and underfitting.

There are two main types of regularization i.e lasso and Ridge regularization.

1) In Ridge regularization – We modify underfitted and over-fitted models by adding the penalty equivalent to the sum of the square of the magnitude of coefficients.
2) In Lasso regularization – We modify over fitted values and underfitted models by adding penalty equivalent to the sum of the absolute value of coefficients.

14. Which particular algorithms are used for regularization?

**ANS**- It uses Lasso(Least Absolute Shrinkage and Selection operator) and Ridge Algorithm.

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

**ANS** – term error present represent the margin of error within a statistical model. It refers to a sum of the deviation within the regression line . In simple words it provide the difference between the theoretical value of the model and actual observed result.