1. Which of the following methods do we use to find the best fit line for data in Linear Regression?
   Least Square Error
2. Which of the following statement is true about outliers in linear regression?
   Linear regression is sensitive to outliers.
3. A line falls from left to right if a slope is _____?
   Negative
4. Which of the following will have symmetric relation between dependent variable and independent variable?
   Correlation
5. Which of the following is the reason for over fitting condition?
   Low bias and high variance
6. If output involves label, then that model is called as:
   Predictive modal
7. Lasso and Ridge regression techniques belong to _____?
   Regularization
8. To overcome with imbalance dataset which technique can be used?
   SMOTE
9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?
   TPR and FPR
10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
    False
11. Pick the feature extraction from below:
    Apply PCA to project high dimensional data
    Forward selection
12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?
    It becomes slow when number of features is very large.
    We need to iterate

13. Explain the term regularization?
    Many times, it has been observed that our machine learning model performs well on our training data but does not perform well on the unseen or test data. This is termed as introduction of noise in the output. By noise we mean those data points that don't really represent true properties of our data. Such type of model is called as overfitted model. Regularization technique prevents the model from overfitting by adding extra information to it. It is a form of regression which shrinks the coefficient estimates to Zero. In other words, this technique forces us not to learn more complex or flexible model to avoid the problem of overfitting. For regression problems the increase in flexibility of the model is represented by its coefficient, which are calculated from the regression line. In regularization technique we reduce the magnitude of the independent variable by keeping same number of dependant variable.

14. Which algorithms are used for regularization?

Below are to two algorithms we use for regularization.

- **Ridge Regression:**
  Ridge Regression is one of the types of linear regression in which we introduce bias, known as Ridge Regression Penalty, so that we can better long term predictions. In Statistics it is called as L2 Norm. When we have the independent variable which is having high collinearity between then at that time the polynomial regression will fail, so to solve such problems we can use Ridge Regression.

- **Lasso Regression:**
  Lasso Regression is another variant of the regularization technique used to reduce complexity in the model. It stands for Least Absolute and Selection Operator.In statistics it is know as L1 Norm. In this technique, the L1 penalty has the effect of forcing some of the coefficient estimates to be exactly equal to zero, which means that there is complete removal of some features for model evaluation.

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

The Error Term is the residual variable produced by a statistical or mathematical model, which is created when a model does not fully represent the actual relationship between the independent and dependent variables. The error term appears in statistical model to indicate the uncertainty of the model. It refers to as an sum of the deviations within the regression line, which provides the difference between the theoretical and the observed values of the model.

Error Term Formula is -

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

where: α, β=Constant parameters

X, ρ=Independent variables

ε=Error term