# MACHINE LEARNING

1. **A.**Least Squared error
2. **A.**Linear regression is sensitive to outlier
3. **B.**Negative
4. **A.**Regression
5. **B.**Low bias and Low variance
6. **B.**Predictive model
7. **A.**Cross validation
8. **D.**SMOTE
9. **A.**FPR and TPR
10. **B.**FALSE
11. **B.**Apply PCA to project high dimensional data
12. **A & B**

## 13 Regularization.

- It is one of the most important concepts of machine learning. This technique prevents the model from overfitting by adding extra information to it.
- Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting.
- This technique forces us not to learn a more complex or flexible model, to avoid the problem of overfitting.
- In the Regularization technique, we reduce the magnitude of the independent variables by keeping the same number of variables.
- In simple linear regression, our optimization function or loss function is known as the residual sum of squares (RSS).

## 14. Algorithms are used for regularization & Techniques

• Ridge Regression
• Lasso Regression

## 1.Ridge Regression:

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

## Usage of Ridge Regression:

- When we have the independent variables which are having high collinearity between them, at that time general linear or polynomial regression will fail so to solve such problems, Ridge regression can be used.

- If we have more parameters than the samples, then Ridge regression helps to solve the problems.

## Limitation of Ridge Regression:

- Not helped in Feature Selection: It decreases the complexity of a model but does not reduce the number of independent variables since it never leads to a coefficient being zero rather only minimizes it. Hence, this technique is not good for feature selection.

- Model Interpretability: Its disadvantage is model interpretability since it will shrink the coefficients for least important predictors, very close to zero but it will never make them exactly zero. In other words, the final model will include all the independent variables, also known as predictors.

## 2. Lasso Regression

Lasso regression is another variant of the regularization technique used to reduce the complexity of the model. It stands for Least Absolute and Selection Operator.

## Limitation of Lasso Regression:

- **Problems with some types of Datasets:** If the number of predictors is greater than the number of data points, Lasso will pick at most n predictors as non-zero, even if all predictors are relevant.
- **Multicollinearity Problem:** If there are two or more highly collinear variables then LASSO regression selects one of them randomly which is not good for the interpretation of our model.

## 15. Error present in linear regression equation?

**Error term:** 1. An error term in statistics is a value which represents how observed data differs from actual population data. It can also be a variable which represents how a given statistical model differs from reality.

- The error term shows the difference between the predicted value and actual value given by the model.
- Linear regression model involves finding the best fitting linear model for observed data that shows the relationship between two variables.
- The distance between each point and the linear graph (shown as black arrows on the below graph) is our error term**.**
- The error term includes everything that separates your model from actual reality. This means that it will reflect nonlinearities, unpredictable effects, measurement errors.