# 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?
   1. Least Square Error B) Maximum Likelihood

C) Logarithmic Loss D) Both A and B

1. Which of the following statement is true about outliers in linear regression?
   1. Linear regression is sensitive to outliers B) linear regression is not sensitive to outliers

C) Can’t say D) none of these

1. A line falls from left to right if a slope is ?
   1. Positive B) Negative

C) Zero D) Undefined

1. Which of the following will have symmetric relation between dependent variable and independent variable?
   1. Regression B) Correlation

C) Both of them D) None of these

1. Which of the following is the reason for over fitting condition?
   1. High bias and high variance B) Low bias and low variance

C) Low bias and high variance D) none of these

1. If output involves label then that model is called as:
   1. Descriptive model B) Predictive modal

C) Reinforcement learning D) All of the above

1. Lasso and Ridge regression techniques belong to ?
   1. Cross validation B) Removing outliers

C) SMOTE D) Regularization

1. To overcome with imbalance dataset which technique can be used?
   1. Cross validation B) Regularization

C) Kernel D) SMOTE

1. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses to make graph?
   1. TPR and FPR B) Sensitivity and precision

C) Sensitivity and Specificity D) Recall and precision

1. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
   1. True B) False
2. Pick the feature extraction from below:
   1. Construction bag of words from a email
   2. Apply PCA to project high dimensional data
   3. Removing stop words
   4. Forward selection

# In Q12, more than one options are correct, choose all the correct options:

1. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?
   1. We don’t have to choose the learning rate.
   2. It becomes slow when number of features is very large.
   3. We need to iterate.
   4. It does not make use of dependent variable.

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

1. Explain the term regularization?

Ans. When we use a regression model to train some data, there is a good chance that the model will overfit the given training dataset. Regularization helps sort this overfitting problem by restricting the degree of freedom of a given equation i.e simply reduces the degree of a polynomial function by reducing their corresponding rates.

Different types of regularization in regression:

1. LASSO
2. RIDGE
3. ELASTICNET ( less popular)
4. Which particular algorithms are used for regularization?

Ans. There are two main regularization techniques.

1. Ridge regression (L2 norm)
2. Lasso regression (L1 norm)

**Ridge regression:** It gives some weightagebut is almost equal to 0. It is a model-tuning method that is used to analyze any data that suffers from multicollinearity.

**Lasso regression**: it is used over regression methods for a more accurate prediction. This model uses shrinkage. The lasso procedure encourages simple, sparse models (i.e models with fewer parameters).

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

Ans. The standard error of the regression represents the average distance that the observed values fall from the regression line.

**Y=b0+bx+e.** In this equation, y is the dependent variable, x is the independent variable, b is the slope of a regression line, and b0 is the line’s intercept on the y- axis.

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