# MACHINE LEARNING

**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
B) Maximum Likelihood
C) Logarithmic Loss
D) Both A and B
**ANS: A) Least Square Error**

2. 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
**ANS: A) Linear regression is sensitive to outliers**

3. A line falls from left to right if a slope is _____?
A) Positive
B) Negative
C) Zero
D) Undefined
**ANS: B) Negative**

4. 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
**ANS: C) Both of them**

5. 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
**ANS: B) Low bias and low variance**

6. If output involves label then that model is called as:
A) Descriptive model
B) Predictive modal
C) Reinforcement learning
D) All of the above
**ANS: B) Predictive modal**

7. Lasso and Ridge regression techniques belong to _____?
A) Cross validation
B) Removing outliers
C) SMOTE
D) Regularization
**ANS: D) Regularization**

8. To overcome with imbalance dataset which technique can be used?
A) Cross validation
B) Regularization
C) Kernel
D) SMOTE
**ANS: D) SMOTE**

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
B) Sensitivity and precision
C) Sensitivity and Specificity
D) Recall and precision
**ANS: C) Sensitivity and Specificity**

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
A) True
B) False
**ANS: A) True**

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
**ANS: B) Apply PCA to project high dimensional data**

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

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

13. Explain the term regularization?

**ANS:** Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting.
This technique can be used in such a way that it will allow to maintain all variables or features in the model by reducing the magnitude of the variables. Hence, it maintains accuracy as well as a generalization of the model.

The commonly used regularization techniques are:

- **L1 Regularization:**
  A regression model which uses L1 Regularization technique is called LASSO (Least Absolute Shrinkage and Selection Operator) regression.
  Lasso Regression adds *"absolute value of magnitude"* of coefficient as penalty term to the loss function.
  Since it takes absolute values, hence, it can shrink the slope to 0, whereas Ridge Regression can only shrink it near to 0.
  The Lasso regression can help us to reduce the overfitting in the model as well as the feature selection.

- **L2 regularization:**
  A regression model that uses L2 regularization technique is called Ridge regression.
  Ridge regression adds "*squared magnitude*" of coefficient as penalty term to the loss function.
  A general linear or polynomial regression will fail if there is high collinearity between the independent variables, so to solve such problems, Ridge regression can be used.
  It helps to solve the problems if we have more parameters than samples.

- **Dropout:**
  Dropout is a regularization technique used in neural networks. It prevents complex co-adaptations from other neurons.
  Using dropout, you can drop connections with *1-p* probability for each of the specified layers. Where *p* is called keep probability parameter.
  Dropout decreases overfitting by avoiding training all the neurons on the complete training data in one go. It also improves training speed and learns more robust internal functions that generalize better on unseen data.
  Along with Dropout, neural networks can be regularized also using L1 and L2 norms. Apart from that, if you are working on an image dataset, image augmentation can also be used as a regularization method.

14. Which particular algorithms are used for regularization?

**ANS:** The general idea behind these algorithms is that to try to minimize and even prevent overfitting.
Lasso has the effect of forcing some coefficient estimates to be exactly zero when hyper parameter θ is sufficiently large. Therefore, one can say that Lasso performs variable selection producing models much easier to interpret than those produced by Ridge Regression. Basically, it is reducing the variability and improving the accuracy of linear regression models.
Lasso is a regularization technique for performing linear regression.

Lasso is one alternative method to stepwise regression and other model selection and dimensionality reduction techniques.

LASSO works well for feature selection in case we have a huge number of features (it reduce redundant features and identify the important ones).

It shrinks coefficients to zero (compare to Ridge which adds "squared magnitude" of coefficient as penalty term to the loss function).

If group of predictors are highly correlated, lasso picks only one of them and shrinks the others to zero.

Other methods like cross-validation, stepwise regression work fairly well for reducing overfitting and perform feature selection. However, they mainly work with a small amount of features. Ridge and LASSO work well with a large amount of features.

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

**ANS:** An error term is a residual variable produced by a statistical or mathematical model, which is created when the model does not fully represent the actual relationship between the independent variables and the dependent variables. As a result of this incomplete relationship, the error term is the amount at which the equation may differ during empirical analysis.

An error term represents the margin of error within a statistical model; it refers to the sum of the deviations within the regression line, which provides an explanation for the difference between the theoretical value of the model and the actual observed results. The regression line is used as a point of analysis when attempting to determine the correlation between one independent variable and one dependent variable.

Within a linear regression model tracking a stock's price over time, the error term is the difference between the expected price at a particular time and the price that was actually observed. In instances where the price is exactly what was anticipated at a particular time, the price will fall on the trend line and the error term will be zero.

The two data points with the greatest distance from the trend line should be an equal distance from the trend line, representing the largest margin of error.

If a model is heteroskedastic, a common problem in interpreting statistical models correctly, it refers to a condition in which the variance of the error term in a regression model varies widely.