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

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

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

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

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

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

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

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

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

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 LinearRegression?

A) We don't have to choose the learning rate.
B)
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.

MACHINE LEARNING

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

13. Explain the term regularization?

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

In a linear equation, we do not want huge weights/coefficients as a small change in weight can make a large difference for the dependent variable (Y). So, regularization constraints the weights of such features is required to avoid overfitting.

14. Which particular algorithms are used for regularization?

Ans. There are different type of regularization i.e. L1, L2 and elastic-net regression.
Here, L1 = Lasso
L2 = Ridge
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.

In this technique, the cost function is altered by adding the penalty term (shrinkage term),

which multiplies the lambda with the squared weight of each individual feature. Therefore,

the optimization function(cost function) becomes:

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}\beta_j^2=\mathrm{RSS}+\lambda\sum_{j=1}^{p}\beta_j^2$$

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**.

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}|\beta_j|=\mathrm{RSS}+\lambda\sum_{j=1}^{p}|\beta_j|.$$

In this technique, the L1 penalty has the effect of forcing some of the coefficient estimates to be exactly equal to zero which means there is a complete removal of some of the features for model evaluation when the tuning parameter λ is sufficiently large.

Let us take an example of Ridge:

Suppose there are three features f1, f2 and f3 and one label. If f1 is not contributing much to predict the label then in that case Ridge will give very less importance to feature f1. Similarly, with the other features.

In case of Lasso, if there are feature f1, f2 and f3 and one label. If f1 is not contributing much than it will eliminate that feature. It will make completely 0. If in case l3 adds much weightage than it will add some penalty to l3 to balance the model so that model wouldn't get biased.

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

Ans: Suppose if a Linear regression model tracking a price of a used car, the error term is the difference between the expected price a used car and the price that was actually quoted. The error term stands for any influence being exerted on the price variable.

Mathematical, error is

r = y – (mx + b)