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

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

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

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: B

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: C

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: D

7. Lasso and Ridge regression techniques belong to_____?
a)Cross validation   B) Removing outliers
C) SMOTE   D) Regularization
Ans: B

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

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: a

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

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: a,b and c

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

13. Explain the term regularization?

   Ans: Regularisation is used for tuning the function by adding the penalty term in the error

   function.Regularization is the process of adding information in order to solve or to prevent

   overfitting.The additional term controls the excessively fluctuating function such that

   the coefficients don't take extreme values.

   Regularizations are techniques used to reduce the error by fitting a function

   appropriately on the given training set and avoid overfitting.

While training a machine learning model, the model can easily be overfitted or underfitted. To avoid this, we use regularization in machine learning to properly fit a model onto our test set. Regularization techniques help reduce the chance of overfitting and help us get an optimal model.

The commonly used regularization techniques are :

1) L1 regularization and

2) L2 regularization

14. Which particular algorithms are used for regularization?

Ans: Particular algorithms used for regularization are also called regularization techniques. The commonly used regularization techniques are:

1) L1 regularization and

2) L2 regularization

There is also one more regularization techniques which is not used normally and it is "Elastic - Net regression .

A regression model which uses L1 Regularization technique is called LASSO(Least Absolute Shrinkage and Selection Operator) regression.

A regression model that uses L2 regularization technique is called Ridge regression.

Lasso Regression adds "absolute value of magnitude" of coefficient as penalty term to the loss function(L).

Lasso method is a type of method which does not give importance to the data which has no relationship with the label.

Ridge method is the method which treats feature according to lits importance i.e, its strength of relationship with the label.

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

Ans: In a linear regression model over the time, the error term is the difference between the expected price at a particular time and the price that was actually observed.

The error term in a regression equation represents the effect of the variables that were removed from the equation.The error term is also known as the residual, disturbance, or remainder term.

An error term is a residual variable produced by a statistical model, 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 appears to indicate the uncertainty in the model and is a residual variable that accounts for a lack of perfect goodness of fit.

Heteroskedastic refers to a condition in which the variance of the residual term, or error term, in a regression model varies widely.