ASSIGNMENT – 39

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 1 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 2 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 3 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 4 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 5 C) Low bias and high 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 6 B) Predictive modal

7. Lasso and Ridge regression techniques belong to _____?

A) Cross validation
B) Removing outliers
C) SMOTE
D) Regularization

Ans 7 D) Regularization

8. To overcome with imbalance dataset which technique can be used?

A) Cross validation
B) Regularization
C) Kernel
D) SMOTE

Ans 8 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 9 A) TPR and FPR

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

A) True
B) False

Ans 10 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

Ans 11 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.

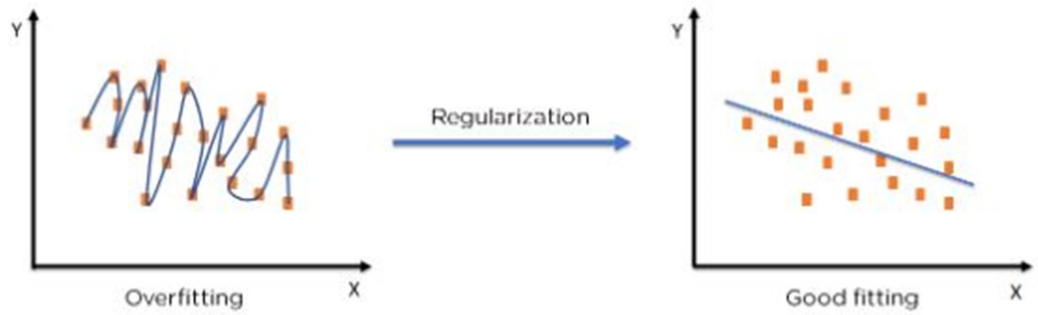
Ans 12 A) We don't have to choose the learning rate.
B) It becomes slow when number of features is very large.
C) We need to iterate.

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

13. Explain the term regularization?

Ans 13

Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting. Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.



14. Which particular algorithms are used for regularization?

Ans 14

There are two main types of regularization techniques: Ridge Regularization and Lasso Regularization.

Ridge Regression modifies the over-fitted or under fitted models by adding the penalty equivalent to the sum of the squares of the magnitude of coefficients.

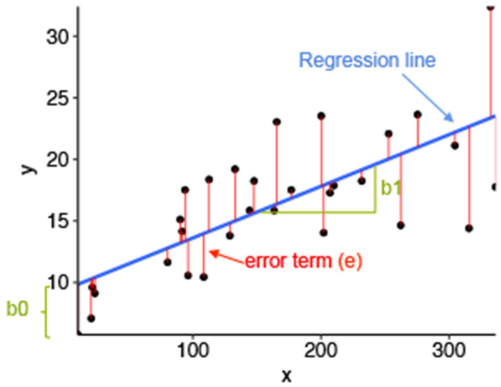
Lasso Regularization modifies the over-fitted or under-fitted models by adding the penalty equivalent to the sum of the absolute values of coefficients.

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

Ans 15

A Linear Regression model's main aim is to find the best fit linear line and the optimal values of intercept and coefficients such that the error is minimized.

Error is the difference between the actual value and Predicted value and the goal is to reduce this difference.



In the above diagram:

- **x is our dependent variable which is plotted on the x-axis and y is the dependent variable which is plotted on the y-axis.**
- **Black dots are the data points i.e the actual values.**
- **bo is the intercept which is 10 and b1 is the slope of the x variable.**
- **The blue line is the best fit line predicted by the model i.e the predicted values lie on the blue line.**

The vertical distance between the data point and the regression line is known as error or residual