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) Leat Square Error

2. Which of the following statement is true about outliers in linear regression?

B) 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 an Independent variable?

A) Regression B) Correlation
C) Both of them D) None of these

Ans. B) Correlation

5. Which of the following is the reason for over fitting condition?

A) High bias and high variance
B) Low bias and lowvariance
C) Low bias andhigh variance
D) none of these

Ans. 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 theabove

Ans. B) Predictive model

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. 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. B) False

11. Pick the feature extraction from below

A) Construction bag of words from a email C) Removing stop words
B) Apply PCA to project high dimensional data 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. A) We don't have to choose the learning rate and B) It becomes slow when number of features is very l arge

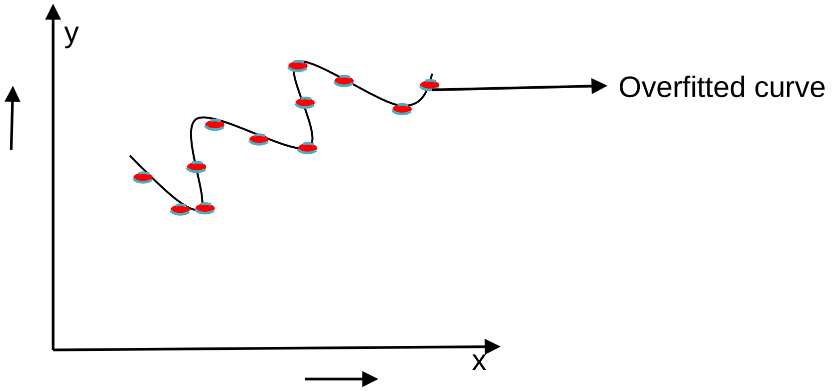
Q13 and Q15 are subjective answer type questions, Answer them briefly.

1. Explain the term regularization?

Ans.

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**.

a. Over fitting model



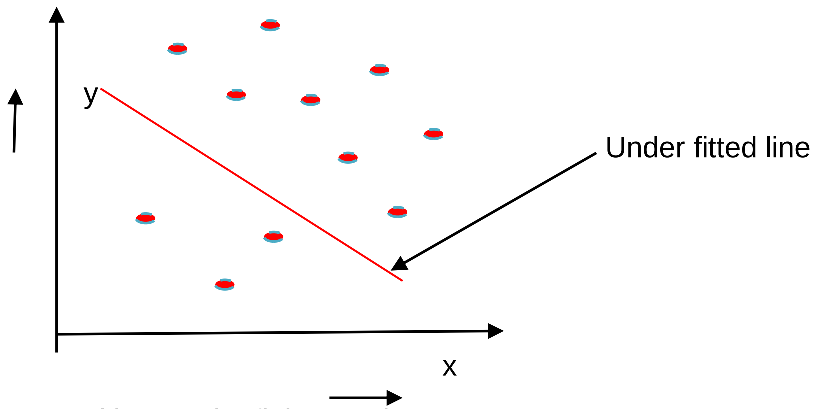
a) Over fitting model

where the machine learning model tries to learn from the details along with the noise in the data and tries to fit each data point on the curve is called Over fitting.

Reasons for over fitting:

- Data used for training is not cleaned and contains noise (garbage values) in it
- The model has a high variance
- The size of the training dataset used is not enough
- The model is too complex

b. Under fitting model



b) Under fitting model

where a machine learning model can neither learn the relationship between variables in the testing data nor predict or classify a new data point is called Under fitting.

Reasons for under fitting:

- Data used for training is not cleaned and contains noise (garbage values) in it
- The model has a high bias
- The size of the training dataset used is not enough
- The model is too simple

2. Which particular algorithms are used for regularization?

Ans. There are three main regularization techniques, namely:

1. Ridge Regression (L2 Norm)
2. Lasso (L1 Norm)
3. Dropout

Ridge Regression (L2 Regularization)

Ridge regression is also called L2 norm or regularization.

When using this technique, we add the sum of weight’s square to a loss function and thus create a new loss function which is denoted thus:

$$\text{Loss} = \sum_{j=1}^{m}\left(Yi - Wo - \sum_{i=1}^{n} Wi\,Xji\right)^{2} + \lambda \sum_{i=1}^{n} Wi^{2}$$

λ is the parameter that needs to be tuned using a cross-validation dataset. When you use λ=0, it returns the residual sum of square as loss function which you chose initially. For a very high value of λ, loss will ignore core loss function and minimize weight’s square and will end up taking the parameters’ value as zero.

Lasso Regression (L1 Regularization)

Also called lasso regression and denoted as below:

$$\text{Loss} = \sum_{j=1}^{m}\left(Yi - Wo - \sum_{i=1}^{n} Wi\,Xji\right)^{2} + \lambda \sum_{i=1}^{n} |Wi|$$

λ is again a tuning parameter and behaves in the same as it does when using ridge regression.

As loss function only considers absolute weights, optimization algorithms penalize higher weight values.

In ridge regression, loss function along with the optimization algorithm brings parameters near to zero but not actually zero, while lasso eliminates less important features and sets respective weight values to zero. Thus, lasso also performs feature selection along with regularization.

Dropout

Dropout is a regularization technique used in neural networks. It prevents complex co-adaptations from other neurons.

Dropout decreases over fitting 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. However, it is important to note that Dropout takes more epochs to train compared to training without Dropout (If you have 10000 observations in your training data, then using 10000 examples for training is considered as 1 epoch).

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

Ans:-

Understanding an Error Term

1. 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.

Error Term Use in a Formula

An error term essentially means that the model is not completely accurate and results in differing results during real-world applications. For example, assume there is a multiple linear regression function that takes the following form:

$Y=\alpha X+\beta\rho+\epsilon$

where:

α,β=Constant parameters

X,ρ=Independent variables

ϵ=Error term

When the actual Y differs from the expected or predicted Y in the model during an empirical test, then the error term does not equal 0, which means there are other factors that influence Y.

What Do Error Terms Tell Us?

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.

Points that do not fall directly on the trend line exhibit the fact that the dependent variable, in this case, the price, is influenced by more than just the independent variable, representing the passage of time. The error term stands for any influence being exerted on the price variable, such as changes in market sentiment.

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.

Linear Regression, Error Term, and Stock Analysis

Linear regression is a form of analysis that relates to current trends experienced by a particular security or index by providing a relationship between a dependent and independent variables, such as the price of a security and the passage of time, resulting in a trend line that can be used as a predictive model.

A linear regression exhibits less delay than that experienced with a moving average, as the line is fit to the data points instead of based on the averages within the data. This allows the line to change more quickly and dramatically than a line based on numerical averaging of the available data points.

The Difference Between Error Terms and Residuals
Although the error term and residual are often used synonymously, there is an important formal difference. An error term is generally unobservable and a residual is observable and calculable, making it much easier to quantify and visualize. In effect, while an error term represents the way observed data differs from the actual population, a residual represents the way observed data differs from sample population data.