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

Answer» 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

Answer» 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

Answer» 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

Answer» B. Correlation

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

Answer» 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

Answer» B. Predictive model

7. Lasso and Ridge regression techniques belong to_____?

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

Answer» D. Regularization

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

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

Answer» 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

Answer» 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

Answer» B. False

MACHINE LEARNING

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
    E) All of the above

Answer» E. All of the above

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.

Answer» A,B and C

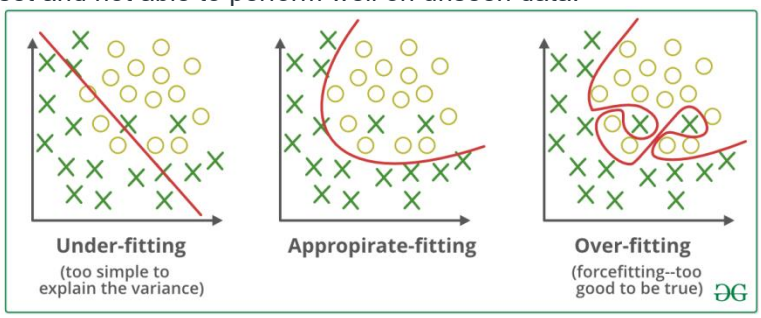
MACHINE LEARNING

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

13. Explain the term regularization?

Answer :-

Overfitting is a phenomenon that occurs when a Machine Learning model is constraint to training set and not able to perform well on unseen data.



Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting.

14.Which particular algorithms are used for regularization?

The commonly used regularization techniques are :

1. L1 regularization
2. L2 regularization
3. Dropout regularization

This article focus on L1 and L2 regularization.

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).

$$||\mathbf{w}||_1 = |w_1| + |w_2| + ... + |w_N|$$

Ridge regression adds "*squared magnitude*" of coefficient as penalty term to the loss function(L).

$$\|\mathbf{w}\|_2 = \left(|w_1|^2 + |w_2|^2 + \ldots + |w_N|^2\right)^{\frac{1}{2}}$$

NOTE that during Regularization the output function(y_hat) does not change. The change is only in the loss function.

The output function:

$$\hat{y} = w_1 x_1 + w_2 x_2 + ... + w_N x_N + b$$

The loss function before regularization:

$$Loss = Error(y, \hat{y})$$

The loss function after regularization:

$$Loss = Error(y, \hat{y}) + \lambda \sum_{i=1}^{N} |w_i|$$

MACHINE LEARNING

$$Loss = Error(y, \hat{y}) + \lambda \sum_{i=1}^{N} w_i^2$$

We define Loss function in Logistic Regression as :
L(y_hat,y) = y log y_hat + (1 - y)log(1 - y_hat)

Loss function with no regularization :
L = y log (wx + b) + (1 - y)log(1 - (wx + b))
Lets say the data overfits the above function.

Loss function with L1 regularization :
L = y log (wx + b) + (1 - y)log(1 - (wx + b)) + lambda*||w||$_1$
Loss function with L2 regularization :
L = y log (wx + b) + (1 - y)log(1 - (wx + b)) + lambda*||w||^{22}

lambda is a Hyperparameter Known as regularization constant and it is greater than zero.
lambda > 0

15 Explain the term error present in linear regression equation?

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.

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.