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

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

3. A line falls from left to right if a slope is ______?

   A) Positive B) Negative C) Zero D) Undefined

   Answer: 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

   Answer: C

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

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

7. Lasso and Ridge regression techniques belong to _________?

A) Cross validation B) Removing outliers

C) SMOTE D) Regularization

Answer: D

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

A) Cross validation B) Regularization

C) Kernel D) SMOTE

Answer: 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

Answer: A

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

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

Answer: B

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 and D

MACHINE LEARNING

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

13. Explain the term regularization?

Regularization:

While training a machine learning model, the model can easily be overfitted or under fitted. 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.

What Are Overfitting and Underfitting?

To train our machine learning model, we give it some data to learn from. The process of plotting a series of data points and drawing the best fit line to understand the relationship between the variables is called Data Fitting. Our model is the best fit it can find all necessary patterns in our data and avoid the random data points and unnecessary patterns called Noise

A scenario 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 **Overfitting**

A scenario 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 Underfitting.

14. Which particular algorithms are used for regularization?

Regularization Techniques:

There are two main types of regularization techniques: **Ridge Regression** and **Lasso**

Ridge Regression:

Ridge regression is a powerful technique used in linear regression to combat overfitting and improve model generalization. It's essentially a regularized version of ordinary least squares (OLS) regression, meaning it penalizes the complexity of the model to prevent it from memorizing the training data too closely.

Here's how it works:

The penalty term: In OLS regression, we minimize the sum of squared errors (SSE) to find the best-fitting line. Ridge regression adds a penalty term to the SSE that penalizes the magnitude of the coefficients. This means that models with smaller coefficients are favored, even if they have slightly higher SSE

L2 regularization: This penalty term uses the L2 norm (sum of squares) of the coefficients, hence the name ridge regression. As this term increases, the penalty on large coefficients also increases, effectively shrinking them towards zero. This reduces the model's complexity and makes it less sensitive to noise in the data.

Benefits of Ridge Regression:

Reduces overfitting: By shrinking coefficients, ridge regression avoids the model memorizing the training data, leading to better performance on unseen data.

Improves model stability: When there are highly correlated features, OLS regression can become unstable. Ridge regression stabilizes the model by reducing the influence of such features.

Feature selection: While not as strong as Lasso, ridge regression can still shrink some coefficients to zero, effectively performing some feature selection.

Lasso Regression:

Lasso regression, also known as Least Absolute Shrinkage and Selection Operator, is another regularization technique used in linear regression to combat overfitting and improve model interpretability. Unlike ridge regression, which uses the L2 norm (sum of squares) of the coefficients as a penalty term, Lasso uses the L1 norm (sum of absolute values).

Here's how it works:

Penalty term: Similar to ridge regression, Lasso adds a penalty term to the sum of squared errors (SSE). This penalty term is based on the L1 norm of the coefficients.

L1 regularization: The L1 norm simply adds up the absolute values of all the coefficients. As this term increases, the penalty on coefficients with larger absolute values also increases, forcing them towards zero. This can lead to some coefficients becoming exactly zero, effectively removing the corresponding features from the model.

Benefits of Ridge Regression:

Reduces overfitting: By setting some coefficients to zero, Lasso prevents the model from memorizing the training data, leading to better generalization.

Improves model interpretability: The sparse nature of the model makes it easier to understand which features are most important and how they affect the dependent variable.

Feature selection: Lasso can automatically select relevant features by setting their coefficients to zero.

Elastic Net Regression:

Elastic Net Regression combines the strengths of both L1 (Lasso) and L2 (Ridge) regularization techniques to create a more robust and versatile model. It aims to address some of the limitations of each individual method, providing several advantages

Key Concepts:

L1 penalty (Lasso): Shrinks coefficients towards zero, potentially setting some to zero for feature selection.

L2 penalty (Ridge): Shrinks coefficients towards zero but doesn't eliminate them, stabilizing model coefficients and improving performance when features are highly correlated.

Elastic Net: Blends L1 and L2 penalties, using a parameter called alpha to control the balance between them.

Here's how it works:

Penalty Term: Like Lasso and Ridge, Elastic Net adds a penalty term to the standard least squares objective function.

Combined Penalties: This penalty term includes both L1 and L2 components, weighted by alpha:

alpha = 0: Pure Ridge regression.

alpha = 1: Pure Lasso regression.

0 < alpha < 1: Blend of both.

Coefficient Shrinkage and Selection: The optimization process minimizes both model error and the combined penalty, leading to:

Shrinkage of coefficients, reducing overfitting.

Potential removal of less important features (coefficients set to zero), performing feature selection.

Benefits of Elastic Net Regression:

Balanced Regularization: Offers a flexible approach by adjusting the balance between L1 and L2 penalties.

Robust to Correlated Features: Better handles scenarios with highly correlated features, where Lasso might arbitrarily select one and ignore others.

Feature Selection: Retains the feature selection ability of Lasso, identifying important features.

Improved Performance: Often outperforms both Lasso and Ridge in terms of prediction accuracy and model generalization.

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

In linear regression equation, the error term represents the difference between the actual observed values of the dependent variable (y) and the values predicted by the regression line. It's a crucial component because it acknowledges that there's inherent variability in the data that the model can't perfectly capture.

Sources of Error:

Unmeasured variables: The model might not include all factors influencing the dependent variable.

Measurement errors: Imperfections in data collection can introduce errors.

Random variation: Natural randomness in the data can lead to deviations from the model.

Importance of Error:

Assessment of model fit: The magnitude of the errors helps evaluate how well the model fits the data. Smaller errors indicate better fit.

Statistical inference: The distribution of errors is used to make statistical inferences about the coefficients and predictions.

Residual analysis: Examining the patterns of errors can reveal potential model misspecifications or violations of assumptions.