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?
   1. Least Square Error B) Maximum Likelihood

C) Logarithmic Loss D) Both A and B

1. Which of the following statement is true about outliers in linear regression?
   1. Linear regression is sensitive to outliers B) linear regression is not sensitive to outliers

C) Can’t say D) none of these

1. A line falls from left to right if a slope is \_\_\_\_\_\_?
   1. Positive B) Negative C) Zero D) Undefined
2. Which of the following will have symmetric relation between dependent variable and independent variable?
   1. Regression B) Correlation C) Both of them D) None of these
3. Which of the following is the reason for over fitting condition?
   1. High bias and high variance B) Low bias and low variance

C) Low bias and high variance D) none of these

1. If output involves label then that model is called as:
   1. Descriptive model B) Predictive modal

C) Reinforcement learning D) All of the above

1. Lasso and Ridge regression techniques belong to \_\_\_\_\_\_\_\_\_?
   1. Cross validation B) Removing outliers

C) SMOTE D) Regularization

1. To overcome with imbalance dataset which technique can be used?
   1. Cross validation B) Regularization

C) Kernel D) SMOTE

1. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses \_\_\_\_\_ to make graph?
   1. TPR and FPR B) Sensitivity and precision

C) Sensitivity and Specificity D) Recall and precision

1. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
   1. True B) False
2. Pick the feature extraction from below: A) Construction bag of words from a email
   1. Apply PCA to project high dimensional data
   2. Removing stop words
   3. Forward selection

In Q12, more than one options are correct, choose all the correct options:

1. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?
   1. We don’t have to choose the learning rate.
   2. It becomes slow when number of features is very large.
   3. We need to iterate.
   4. It does not make use of dependent variable.

ASSIGNMENT – 39

MACHINE LEARNING

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

1. Explain the term regularization?
2. Which particular algorithms are used for regularization?
3. Explain the term error present in linear regression equation?

Explain the term regularization?

Regularization is a technique used in machine learning and statistical modeling to prevent overfitting by adding a penalty to the loss function. This penalty discourages the model from becoming too complex by constraining the magnitude of the model coefficients. Regularization techniques ensure that the model generalizes better to unseen data by smoothing or shrinking the coefficients, which helps in improving the model's performance on new, unseen data.

There are two common types of regularization used in linear regression:

1. **L1 Regularization (Lasso Regression)**:
   * Adds a penalty equal to the absolute value of the magnitude of coefficients.
   * The loss function becomes: Loss=Sum of Squared Errors+λ∑∣wi∣\text{Loss} = \text{Sum of Squared Errors} + \lambda \sum |w\_i|Loss=Sum of Squared Errors+λ∑∣wi​∣.
   * Can shrink some coefficients to exactly zero, thus performing feature selection.
2. **L2 Regularization (Ridge Regression)**:
   * Adds a penalty equal to the square of the magnitude of coefficients.
   * The loss function becomes: Loss=Sum of Squared Errors+λ∑wi2\text{Loss} = \text{Sum of Squared Errors} + \lambda \sum w\_i^2Loss=Sum of Squared Errors+λ∑wi2​.
   * Tends to shrink coefficients but usually doesn't eliminate them completely.

**Elastic Net** combines both L1 and L2 regularization.

Regularization helps to:

* Reduce model complexity.
* Prevent overfitting by keeping the model weights small.
* Improve the generalization of the model to new, unseen data.

In summary, regularization is a critical technique in machine learning for building robust and generalizable models.

1. Which particular algorithms are used for regularization?

The particular algorithms used for regularization in machine learning include:

1. **Lasso Regression (Least Absolute Shrinkage and Selection Operator)**:
   * Uses L1L1L1 regularization, which adds a penalty equal to the absolute value of the magnitude of coefficients.
   * Can shrink some coefficients to zero, effectively performing feature selection.
2. **Ridge Regression**:
   * Uses L2L2L2 regularization, which adds a penalty equal to the square of the magnitude of coefficients.
   * Tends to shrink coefficients but usually does not eliminate them completely.
3. **Elastic Net**:
   * Combines both L1L1L1 and L2L2L2 regularization.
   * The loss function includes both the absolute value and the square of the coefficients' magnitudes.
4. **Regularized Logistic Regression**:
   * Can use L1L1L1, L2L2L2, or Elastic Net regularization.
   * Applied to classification problems to prevent overfitting.
5. **Support Vector Machines (SVM) with Regularization**:
   * SVMs can include a regularization parameter (often denoted as CCC) to control the trade-off between maximizing the margin and minimizing classification error.

These algorithms incorporate regularization terms in their loss functions to penalize large coefficients and thus prevent overfitting, resulting in more robust and generalizable models.

Explain the term error present in linear regression equation?

In the context of a linear regression equation, the term "error" refers to the difference between the observed value and the predicted value of the dependent variable. It is also known as the residual. The general form of a linear regression equation is:

y=β0+β1x+ϵy = \beta\_0 + \beta\_1 x + \epsilony=β0​+β1​x+ϵ

where:

* yyy is the dependent variable (the outcome or target variable).
* xxx is the independent variable (the predictor or feature).
* β0\beta\_0β0​ is the y-intercept of the regression line (the value of yyy when x=0x = 0x=0).
* β1\beta\_1β1​ is the slope of the regression line (the change in yyy for a one-unit change in xxx).
* ϵ\epsilonϵ (epsilon) is the error term (the residual).

**Types of Errors**

**Residual/Error (ϵ\epsilonϵ)**:

* + Represents the deviation of the observed value (yyy) from the predicted value (y^\hat{y}y^​).
  + For an observation iii, the residual is calculated as: ϵi=yi−y^i\epsilon\_i = y\_i - \hat{y}\_iϵi​=yi​−y^​i​.

**Mean Squared Error (MSE)**:

* + A common measure to evaluate the overall error in a regression model.
  + Calculated as the average of the squared residuals: MSE=1n∑i=1n(yi−y^i)2\text{MSE} = \frac{1}{n} \sum\_{i=1}^n (y\_i - \hat{y}\_i)^2MSE=n1​∑i=1n​(yi​−y^​i​)2.

**Root Mean Squared Error (RMSE)**:

* + The square root of the Mean Squared Error.
  + Provides a measure of the average magnitude of the residuals: RMSE=MSE\text{RMSE} = \sqrt{\text{MSE}}RMSE=MSE​.

**Mean Absolute Error (MAE)**:

* + The average of the absolute values of the residuals: MAE=1n∑i=1n∣yi−y^i∣\text{MAE} = \frac{1}{n} \sum\_{i=1}^n |y\_i - \hat{y}\_i|MAE=n1​∑i=1n​∣yi​−y^​i​∣.

**Interpretation of Error**

* **Low Error**: Indicates that the predicted values are close to the observed values, meaning the model has a good fit.
* **High Error**: Indicates that the predicted values deviate significantly from the observed values, meaning the model has a poor fit.

**Sources of Error**

* **Model Error**: Due to the assumptions and limitations of the linear regression model itself.
* **Measurement Error**: Due to inaccuracies in the data collection process.
* **Omitted Variable Error**: Due to the exclusion of relevant variables that influence the dependent variable.

Understanding and minimizing error is crucial for improving the accuracy and reliability of a linear regression model.