Machine Learning-Assignment Answers

1.
(A) Least Square Error

2.
(A) Linear regression is sensitive to outliers

3.
(B) Negative

4.
(B) Correlation

5.
(C) Low bias and high variance

6.
(B) Predictive model

7.
(D) Regularization

8.
(D) SMOTE

9.
(A) TPR and FPR

10.
(B) False

11.
(B) Apply PCA to project high dimensional data

12.
(A) We don’t have to choose the learning rate.
(B) It becomes slow when number of features is very large.

13.
Answer:

Regularization

Regularization is a technique used in machine learning to prevent models from becoming too specialized in the training data (overfitting) and improve the generalization performance of a model. It adds extra constraints or penalties during the training process to encourage simpler and more generalized models. The goal of regularization is to improve the model's ability to make accurate predictions on new, unseen data.

Overfitting

Overfitting occurs when a model becomes too complex and starts to memorize the training data instead of learning the underlying patterns. This can lead to poor performance when applied to new, unseen data.

When we train a machine learning model, we want it to learn patterns and make predictions accurately. However, sometimes the model can learn too much from the training data, including noise or random variations that are not useful for making predictions in general. Regularization techniques help us avoid this problem by adding extra rules or penalties during the training process. These rules encourage the model to be simpler and less likely to rely too much on specific details in the training data.

By promoting simplicity, regularization helps the model focus on the most important patterns and generalize better to new situations. In other words, it helps the model to be more flexible and accurate when making predictions on data it has never seen before.

Features of Regularization Technique:

● Regularization adds a penalty term to the loss function of the model, discouraging overly complex or extreme parameter values.
● The penalty term helps to control the model's complexity by shrinking the parameter values towards zero, making the model less sensitive to individual data points.
● There are different types of regularization techniques, such as L1 regularization (Lasso), L2 regularization (Ridge), and Elastic Net regularization.

● L1 regularization adds the absolute values of the parameter coefficients to the loss function, encouraging sparsity and feature selection.
● L2 regularization adds the squared values of the parameter coefficients to the loss function, promoting small and distributed parameter values.
● Elastic Net regularization combines L1 and L2 regularization to leverage the benefits of both techniques.
● The regularization strength or hyperparameter determines the balance between reducing the model's complexity and fitting the training data.

So, by applying regularization, models can generalize better to unseen data, reduce the risk of overfitting, and improve overall performance and stability.

14.
Answer:

Regularization Algorithms

In machine learning, several regularization algorithms are used to prevent overfitting and improve generalization of model. Some popular regularization algorithms that commonly used are:

- Ridge Regularization Algorithm
- Lasso Regularization Algorithm
- Elastic Net Regularization Algorithm

Ridge Regularization Algorithm

Ridge regularization is also known as L2 regularization. Ridge regularization technique is used in regression models to prevent overfitting and improve the generalization of the model. Ridge regularization adds a penalty term proportional to the square of the magnitude of the coefficients to the loss function.
Loss function = Sum of squared residuals + alpha * (sum of squared coefficients)
Or
Loss function = $||y - Xw||^2 + \alpha * ||w||^2$
Here
X: input values
y: output values
w: coefficients
α: regularization strength

$||y - Xw||^2$: represents the ordinary least squares (OLS) loss, which measures the difference between the predicted values and the true values.
$\alpha * ||w||^2$: represents the ridge penalty term, which is the L2 norm (squared magnitude) of the coefficients multiplied by the regularization strength (alpha).

Here, alpha is the regularization strength that controls the amount of regularization applied. Higher alpha values result in greater regularization.
The ridge penalty term shrinks the coefficients towards zero but does not set them exactly to zero.
Ridge regularization is particularly useful when dealing with multicollinearity (high correlation) between the features, as it helps to reduce the impact of correlated features on the model.
Ridge regression tends to provide a more stable and robust solution compared to ordinary least squares regression when dealing with high-dimensional data.

Lasso Regularization Algorithm

Lasso regularization is also known as L1 regularization. Lasso regularization technique is also used in regression models to prevent overfitting and improve the generalization of the model. Lasso regularization adds a penalty term proportional to the absolute value of the coefficients to the loss function.
Loss function = Sum of squared residuals + alpha * (sum of absolute values of coefficients)
Or
Loss function =$||y - Xw||^2 + \alpha * ||w||$
Here
X: input values
y: output values
w: coefficients
α: regularization strength
$||y - Xw||^2$: represents the ordinary least squares (OLS) loss, which measures the difference between the predicted values and the true values.
$\alpha * ||w||$: represents the lasso penalty term, which is the L1 norm (absolute values of coefficients) of the coefficients multiplied by the regularization strength (alpha).

Similar to ridge regularization, the alpha parameter controls the amount of regularization applied, with higher values resulting in stronger regularization.
The key difference between Lasso and Ridge is that Lasso has the ability to shrink some coefficients exactly to zero, effectively performing feature selection.
Lasso regularization encourages sparsity by eliminating irrelevant or less important features, which can lead to a more interpretable and simpler model.
Lasso regression is particularly useful when dealing with a large number of features or when feature selection is desired.

Elastic Net Algorithm

Elastic Net combines L1 and L2 regularization and is effective when there are multiple correlated features. It adds both the absolute values and squared values of the parameter coefficients to the loss function, providing a balance between L1 and L2 regularization.

15.
Answer:

Error present in linear regression equation

In linear regression, the term "error" refers to the difference between the actual values of the dependent variable (also known as the target variable or the response variable) and the predicted values obtained from the linear regression equation. These errors are also known as residuals.
In linear regression, the goal is to find a line that best fits the given data points. The linear regression equation estimates the relationship between the dependent variable and one or more independent variables by calculating the coefficients that minimize the sum of squared errors.

The linear regression equation can be represented as:
$\mathbf{Y = a + b_1X_1 + b_2X_2 + \ldots + b_nX_n + \varepsilon}$

Here:
Y: represents the dependent variable
$X_1, X_2, \ldots, X_n$: represent the independent variables
a: represents intercept

b_1, b_2, ..., b_n: represent the coefficients (also known as regression coefficients or weights) corresponding to each independent variable
ε: represents the error term or residual

The objective of linear regression is to minimize these errors or residuals. The most common method for estimating the coefficients is called the least squares method, which aims to minimize the sum of the squared errors. By minimizing the errors, the linear regression model finds the best-fitting line or hyperplane that represents the relationship between the variables.

Analyzing the errors in linear regression is important for assessing the goodness of fit of the model and understanding the variability that is not explained by the model. Various statistical metrics, such as the Mean Absolute Error (MAE), Mean Squared Error (MSE) root mean squared error (RMSE) or the coefficient of determination (R-squared), are used to evaluate the performance of a linear regression model and determine how well it captures the relationship between the variables.

By minimizing the errors, the linear regression model aims to find the best-fit line that represents the relationship between the independent variables and the dependent variable.