ML Assignment Answers
Batch DS2312
15/01/2024

Answer1 - A) Least Square Error

Answer2 - A) LR is sensitive to outliers

Answer3 - B) Negative

Answer4 - B) Correlation

Answer5 - A) High bias and High Variance

Answer6 - B) Predictive model

Answer7 - A) Cross Validation

Answer8 - D) SMOTE

Answer9 - A) TPR&FPR

Answer10 - B) False

Answer11 - A) Construction bag of words from a email

Answer12 - A) We don't have to choose the learning rate & B) It becomes slow when number of features is very large

Answer13-

In machine learning, regularization is a technique used to prevent overfitting and improve the generalization performance of a model. Overfitting occurs when a model fits the training data too closely, capturing noise and fluctuations that are not representative of the true underlying patterns in the data. Regularization introduces a penalty term to the model's objective function, discouraging overly complex models with large coefficients. Regularization is commonly applied to linear regression, logistic regression, and other models to improve their robustness and generalization performance on unseen data.

Answer14-

There are two common types of regularization used in machine learning:

1. L1 Regularization (Lasso):
    - In L1 regularization, the penalty term is proportional to the absolute values of the model coefficients. It adds the sum of the absolute values of the coefficients to the cost function.

    - L1 regularization encourages sparsity in the model, meaning it tends to set some coefficients to exactly zero. This can be useful for feature selection, as some features may have negligible impact on the model.

2. L2 Regularization (Ridge):
    - In L2 regularization, the penalty term is proportional to the square of the model coefficients. It adds the sum of the squared values of the coefficients to the cost function.

- L2 regularization tends to shrink the coefficients towards zero but does not typically set them exactly to zero. It helps to prevent the model from becoming too sensitive to individual data points.

Answer15-

In linear regression, the term "error" refers to the difference between the predicted values generated by the linear regression model and the actual observed values in the dataset. These errors are also known as residuals. The linear regression equation models the relationship between the independent variable(s) and the dependent variable, and the errors represent the unexplained variability or noise in the data that the model is unable to capture.