MACHINE LEARNING

Answer 1. A) Least Square Error

Answer 2. A) Linear regression is sensitive to outliers

Answer 3. B) Negative

Answer 4. B) Correlation

Answer 5. C) Low bias and high variance

Answer 6. B) Predictive model

Answer 7. D) Regularization

Answer 8. D) SMOTE

Answer 9. A) TPR and FPR

Answer 10. B) False

Answer 11. A) Construction bag of words from an email

Answer 12. A) We don't have to choose the learning rate. AND C) We need to iterate

Answer 13. Regularization in machine learning is aimed at preventing overfitting, a phenomenon where a model excessively tailors itself to the training data, failing to generalize to new data effectively. It involves adding a penalty term to the model loss function during training. This penalty discourages the model from assigning overly large weights to certain features or parameters, which could lead to capturing noise in the training data. Common regularization techniques include L1 regularization, which encourages sparsity by penalizing the absolute values of parameters, and L2 regularization, which penalizes the squared values of parameters, discouraging large weights. Regularization promotes the selection of more meaningful features and avoids over-reliance on noise. It helps create simpler, more interpretable models that generalize better to unseen data, enhancing model reliability and performance.

Answer 14. Common algorithms using regularization include Ridge and Lasso regression, which apply L2 and L1 penalties respectively, promoting smaller parameter values and feature selection. Elastic Net combines both penalties for balance. Logistic regression integrates L1 or L2 regularization. Support Vector Machines utilize regularization to manage the trade-off between margin width and classification accuracy. Neural networks employ dropout, randomly disabling nodes during training to prevent over-reliance. Pruning in decision trees curbs excessive growth. Even K-Nearest Neighbors can benefit from regularization to avoid noise sensitivity. These techniques collectively mitigate overfitting, enhancing model generalization and reliability.

Answer 15. In the linear regression equation, the "error" refers to the discrepancy between the predicted values generated by the model and the actual observed values in the training data. It represents the inherent variability and noise present in real-world data that the model cannot fully capture. Minimizing these errors, typically through techniques like least squares optimization, is the goal of linear regression. The goal is to find the line that best fits the data by minimizing the sum of squared errors, resulting in a more accurate representation of the underlying relationship between variables.