MACHINE LEARNING ASSIGNMENT -7

ANSWERS:

1. D) All of the above
2. A) Random forest
3. B) The regularization will decrease
4. A) It regularizes the decision tree by limiting the maximum depth up to which a tree can be grown.
5. A) It's an ensemble of weak learners., C) In case of classification problem, the prediction is made by taking mode of the class labels predicted by the component trees.
6. A) Gradient Descent algorithm can diverge from the optimal solution.
7. B) Bias will decrease, Variance increase
8. B) model is overfitting
9. Calculations:
    - Gini index:
    - The Gini index measures the impurity of the dataset and is given by: Gini = 1 - (pA)^2 - (pB)^2. Substituting the values, we get: Gini = 1 - (0.4)^2 - (0.6)^2= **0.48**.

    - Entropy:
    - The entropy measures the impurity of the dataset and is given by: Entropy = -pA * log2(pA) - pB * log2(pB), Substituting the values, we get: Entropy = -0.4 * log2(0.4) - 0.6 * log2(0.6)= **0.971.**

10. Advantages:
    - Improved accuracy: Random Forests typically have higher accuracy than a single Decision Tree because they reduce the variance and overfitting that can occur with Decision Trees.
    - Robustness to noise and outliers: Random Forests are more robust to noise and outliers in the data than a single Decision Tree.
    - Feature importance: Random Forests can provide a measure of feature importance, which can help identify the most relevant features for prediction.
    - Scalability: Random Forests can handle large datasets with a large number of features.
    - Reduced overfitting: Random Forests use techniques such as bootstrapping and feature bagging to reduce overfitting.

11. Scaling numerical features in a dataset is necessary to ensure that all features are on a comparable scale. If the features are not on the same scale, some features may dominate others in terms of their contribution to the model's predictions. This can lead to biased results and can affect the performance of machine learning models. Two Scaling Methods: Standardization, Normalization
12. Advantages of Scaling:
    - Faster convergence: Scaling the features ensures that they are on a comparable scale, which can help gradient descent converge faster.
    - Avoiding oscillations: When features have different scales, the gradient descent algorithm may oscillate between the directions of the large and small scale features.
    - Better conditioning of the optimization problem: Scaling can improve the conditioning of the optimization problem, making it easier to solve.
13. In the case of a highly imbalanced dataset for a classification problem, accuracy may not be a good metric to measure the performance of the model. This is because accuracy can be misleading when the classes are imbalanced. Example: a binary classification problem with 99% of the data belonging to class A and 1% belonging to class B. A model that simply predicts class A for all instances will achieve an accuracy of 99%, even though it is not able to identify any instances of class B. In this case, accuracy does not provide a meaningful measure of the model's performance.
14. The F-score, also called the F1-score, is a measure of a model's accuracy on a dataset. It is used to evaluate binary classification systems, which classify examples into 'positive' or 'negative'. The F-score is a way of combining the precision and recall of the model. Formula: **F1 Score = (2*(Precision*Recall))/(Precision+Recall)**
15. The fit() method helps in fitting the data into a model while the transform() method helps in transforming the data into a form that is more suitable for the model. Fit_transform() method, on the other hand, combines the functionalities of both fit() and transform() methods in one step.