

Assignment - 05

1. What is a Decision Tree, and how does it make decisions during test time?

Ans:-

It is an algorithm based on the supervised learning which is used for both regression and classification task. Based up on the input features values it split into subsets of data while working on model by a tree structure creation where feature is represented by each node & the decision rule is represented by each branch.

Decision Tree makes decisions during test time by:-

- Evaluating features at each node
- Directing the instance down the tree based on feature values
- Reaching a leaf node, which contains the predicted outcome

2. How does Bagging improve the performance of a Decision Tree?

Ans:-

Bagging improves Decision Tree performance in several ways:

- Reduces Overfitting: By averaging the predictions of multiple trees, Bagging reduces the variance of the model, making it less prone to overfitting.
- Improves Generalization: Bagging helps the model generalize better to new data, as each tree is trained on a slightly different subset of the data.
- Increases Accuracy: By combining the predictions of multiple trees, Bagging can improve the overall accuracy of the model.

3. In what situations might a Decision Tree overfit the training data, and how can this be mitigated?

Ans:-

High dimensional data, noise capture and low fluctuations might be overfit the training data.

Mitigation of Overfitting:

- (i) Pruning: Based up on the criteria, the branches should be removed which have low importance.
- (ii) Regularization: Add a penalty term to the loss function to discourage complex trees.
- (iii) Hyperparameter Tuning: Carefully tune hyperparameters, such as tree depth and pruning threshold, to avoid overfitting.
- (iv) Cross-Validation: The best tree parameters should be cross validated.

4. How does Random Forest differ from a single Decision Tree?

Ans:-

Random forest is a learning method where large number of decision tree is constructed during the data.

Random forest difference from decision tree:

1. Multiple Trees: Random Forest grows multiple Decision Trees (typically 100-1000) instead of a single tree.
2. Random Feature Selection: At each split, Random Forest randomly selects a subset of features to consider, reducing overfitting and improving generalization.

3. Bootstrap Sampling: Random Forest uses Bootstrap Sampling to create a new training dataset for each tree, reducing overfitting and increasing diversity.
4. Voting: Random Forest makes predictions by voting, where each tree contributes to the final prediction.

5. What is the main idea behind Boosting in ensemble methods?

Ans:-

Boosting is a technique improves the training model performance sequentially and correct the errors done the previous model with the attempts of new model.

- (i) Iteratively trains the model by adaptive training manner.
- (ii) Misclassified previous models are corrected by new model.
- (iii) The final model should be the more accurate model.