Decision Trees

Introduction

Decision Trees are versatile and easy-to-understand models used in both classification and regression tasks in *Machine Learning*. While Decision Trees are commonly associated with classification problems, they can also be effectively used for regression tasks by predicting continuous values instead of discrete classes.

Structure of a Decision Tree

1. **Nodes:** In a Decision Tree, nodes represent questions or conditions based on features of the dataset.
2. **Edges:** Edges of the tree show the outcome of a split based on a particular attribute.
3. **Leaves:** Terminal nodes of the tree are called leaves and contain the final output or prediction.

Regression with Decision Trees

When using *Decision Trees* for regression, the algorithm aims to predict a continuous value for the target variable instead of a class label. The prediction at each leaf node is usually the average (or median) of the target values of the training instances that fall into that leaf.

Splitting Criteria

1. **Mean Squared Error (MSE):** Commonly used criterion for regression trees, where the model tries to minimize the average squared difference between the predicted values and the actual values.

2. **Mean Absolute Error (MAE):** Another criterion used for regression trees, where the model tries to minimize the average absolute difference between the predicted values and the actual values.

Tree Pruning

Decision Trees are prone to overfitting, especially when the tree grows too large and captures noise in the training data. Pruning techniques such as setting a minimum number of samples required to split a node or limiting the maximum depth of the tree can help prevent overfitting.

Advantages of Decision Trees for Regression

1. **Interpretability:** Decision Trees provide clear insights into the decision-making process, making them easy to interpret and explain.
2. **Non-linear relationships:** Decision Trees can capture non-linear relationships between features and the target variable.
3. **Robustness to outliers:** Decision Trees are robust to outliers in the data compared to some other regression algorithms.

Conclusion

In conclusion, Decision Trees are powerful tools for regression tasks, offering simplicity, interpretability, and the ability to handle non-linear relationships in the data. By effectively splitting the feature space based on informative criteria, Decision Trees can produce accurate predictions for continuous target variables.

Producer: Elham Jafari

Computer Engineering