

Assignment_5_task_5.3

Machine Learning (WiSe 2025/2026)

Author: Suvansh Shukla

Matriculation No. 256245

Assignment 5 Task 5.3

(a)

A classification tree predicts nominal values. But, a regression tree predicts numbered values.

Classification deals with discrete outcomes, like Yes/No or any category. But, regression trees

deal with continuous values, like temperature, price etc...

They also have different evaluation metrics. Classification trees use Precision, Recall and F1-Score.

While Regression trees use Mean Squared Error, R2-Score, MAPE and RMSE.

Decision boundaries in classification trees are clearly defined while regression trees don't have distinct boundaries.

(b)

In regression our primary goal is to find the best fit line that can represent the relationship between independent variables (inputs) and a dependent variable (output). This best fit line may be linear or non-linear.

A way to judge whether our best fit line is adequate and accurate is to use an error function that can find difference between our predicted values and the real values. One such function is the Sum of Squared Errors or SSE. Using this function to minimize the error when predicting values helps us increase the accuracy and performance our classifier.

SSE can be used to compare what splits on a node would perform better and give better accuracy.

The Gini Index, commonly used to figure out how to split our decision

also uses SSE. It is calculated by subtracting the sum of squared probabilities of each class from one.

$$GiniIndex = 1 - \sum_{i=1}^n (P_i)^2$$

The Chi-Square algorithm used for finding out the statistical significance of differences between sub-nodes and parent nodes also uses SSE. It is measured by the SSE of standardized differences between observed and expected frequencies of a target variable.^[1]

(c)

There are multiple stopping criteria for building trees, some of which are the following:

1. When the decision tree reaches 100% purity, meaning all examples now belong to the same class
 2. When a pre-defined tree depth has been reached
 3. When the size of the data in the node is below a threshold
 4. When the error or information gain of objects in nodes have dropped below a threshold
-
-

1. <https://wandb.ai/sauravmaheshkar/Decision-Tree/reports/Decision-Trees-A-Guide-with-Examples--VmlldzoxMDE5Nzkw> ↵