

Machine Learning
Regression Decision Trees
Assignment
Due Date: April 30, 2020

We used decision trees for classifying data samples. In the model we used, each internal node tested a condition (value of a feature was tested) to follow a branch down until a leaf node was reached. Leaf nodes correspond to classes and a decision is made when we reach a leaf node.

In addition, during training stage, we decided to split a node based on its impurity value (using Entropy or Gini metrics).

It is also possible to use decision trees for predictions (regression). The attached document expresses the concept in details.

Yet it is possible to split our data and have multiple decision trees for better accuracy. In this case, we will have a random forest for regression.

Use the flight delay data to create a regression random forest. In this case you are not classifying a flight as **with delay** or **without delay**, but predicting the amount of expected delay.

Implement your model and determine its accuracy.

Submit your code, with a report describing how you reached to this solution (the theory behind your solution), and your code.

Submit both your report and the code as a single pdf file.