

Exercises – Topic5 – Regression and Regression Trees

1. Use three methods: i) regression, ii) regression tree, iii) model tree on the weather dataset: <https://archive.ics.uci.edu/ml/datasets/SML2010> that we also used in Topic2 Naïve Bayes (chapter 4), and forecast a) temperature (attribute 5), b) the amount of precipitation (in attribute 12). How do the three methods compare (in terms of mean absolute error MAE)?

Note: The data from the repository is in .txt rather than .csv format. You can use:

read.table instead of **read.csv** when importing the file to RStudio

Also install package: **rpart.plot**

2. Use the Titanic dataset <https://www.kaggle.com/c/titanic/> and decide what decision tree based approach works best for predicting survival of passengers.

You will need a Kaggle user, and you can find multiple approaches and inspiration on the subject through the website.