Tutorial 12 - Tree-based Models

AUTHOR
Victoria Hünewaldt

Tutorial 12

This tutorial will cover decision trees.

You will learn:

- how to run a regression decision tree
- how to run a classification decision tree
- how to visualise decision trees
- how to evaluate its performance on test & training data

Exercises

Use the airbnbsmall data set. You will need to use the "rpart" and "rpart.plot" library. Create a .qmd-file and solve the tasks there. Store it in the JupyterHub folder "Session 12".

Regression Decision Tree

Run a regression decision tree explaining the variable "price" (= endogenous variable). Use all other variables as potential predictor variables (i.e. specify a full model).

Print and plot the tree.

Classification Decision Tree

Split your data into test and training data. Run a classification decision tree explaining the variable "high_rating" (= endogenous variable) on the training data. Use all other variables as potential predictor variables (i.e. specify a full model).

Create a confusion matrix for the training data and one for the test data.

Compare your results to the results obtained by logistic regression in tutorial 10.