Toulouse
School of
Economics

# Worksheet 3 – Classification and Regression Trees (solution)

### High-Dimensional Data Analysis and Machine Learning

Camille Mondon

January 19, 2026

## 1 CART on Insurance Data

**Exercise 1.** The dataset contains **2220 observations** of young drivers with car insurance contracts. The variables are TYPE, VALUE, SEX, AGEV, and AGEI. Below is a step-by-step solution in **R**.

1. Load the dataset and describe it.

2. Split the dataset into training and test samples (80%-20%).

3. Use the CART algorithm in order to explain the variable TYPE by the variables VALUE, SEX, AGEV and AGEI.

> 💡 Remark
>
> *Remark* 1.1. Ensure TYPE is a factor for classification trees.

4. Prune the tree using cross-validation (x-error) and the 1-SE rule.

5. Predict the TYPE on test set and assess prediction quality.

6. Compute and visualize ROC curves.

7. Interpret the results and compare them with Linear Discriminant Analysis.