

Exercise 4 - deadline 16.04.2021 23:59

This exercise is a continuation of exercise 3 and the last one with the adult dataset. You should keep the same classifier and transformations as in the exercise 3. The goal is to vary the parameters of the classifier and use a validation dataset (split from the training dataset) to choose the best parameters. Once the best parameters are found, you should use the parameters to fit a classifier using the full training data.

The goals of the exercise:

1. Choose at least 2 parameters to tune, each having at least 2 different possible values.
2. Train and evaluate a decision tree for each possible combination of the chosen parameters.
3. Choose the best parameters and train the model using the full training dataset.
4. Save the model (as in the exercise 3).

The final submission should contain (at least) one script to contain all the goals mentioned before.

Optional goals:

1. Instead of (or additionally to) splitting the dataset into training and validation data, use cross-validation to tune the parameters.

Important Note:

The submission should be done individually. Please name your script/notebook accordingly. At the beginning of your script / notebook give the name(s) of all students who worked on the exercise. You may work in small groups (of 2-3) but all students must be confident with the work of others. They should understand and be able to answer detailed questions of the implementation.