Module overview

Contents

What you will learn

In the previous modules, we showed how to create, train, predict, and even evaluate a predictive model. However, we did not change the models' parameters that can be given when creating an instance. For example, for k-nearest neighbors, we initially used this default parameter: `n_neighbors=5` before trying other model parameters.

These parameters are called **hyperparameters**: they are parameters used to control the learning process, for instance the parameter `k` of the k-nearest neighbors. Hyperparameters are specified by the user, often manually tuned (or by an exhaustive automatic search), and cannot be estimated from the data. They should not be confused with the other parameters that are inferred during the training process. These parameters define the model itself, for instance `coef_` for the linear models.

In this module, we will first show that the hyperparameters have an impact on the performance of the model and that default values are not necessarily the best option. Subsequently, we will show how to set hyperparameters in scikit-learn model. Finally, we will show strategies allowing to pick-up a combination of hyperparameters that maximizes model's performance.

Before getting started

The required technical skills to carry on this module are:

- skills acquired during the "The Predictive Modeling Pipeline" with basic usage of scikit-learn;
- skills related to using the cross-validation framework to evaluate a model.

Objectives and time schedule

The objective in the module are the following:

- understand what is a model hyperparameter;
- understand how to get and set the value of a hyperparameter in a scikit-learn model;
- be able to fine tune a full predictive modeling pipeline;
- understand and visualize the combination of parameters that improves the performance of a model.

The estimated time to go through this module is about 3 hours.

Join the full MOOC for better learning!