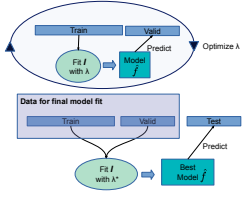
Introduction to Machine Learning

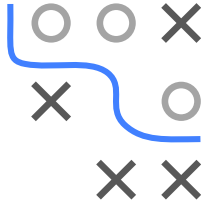
Training - Validation - Test



Learning goals

- Understand how to fulfill the untouched test set principle by a 3-way split of the data
- Understand how thereby the tuning step can be seen as part of a more complex training procedure

TUNING PROBLEM



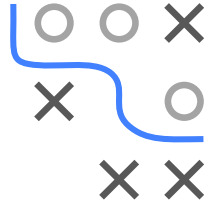
Remember:

We need to

- **select an optimal learner**
- without compromising the **accuracy of the performance estimate** for that learner

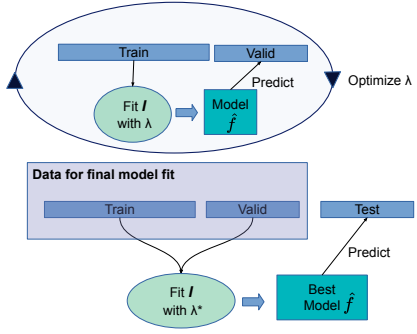
– for that we need an **untouched test set**!

TRAIN - VALIDATION - TEST

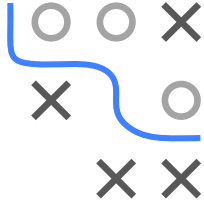


Simplest method to achieve this: a 3-way split

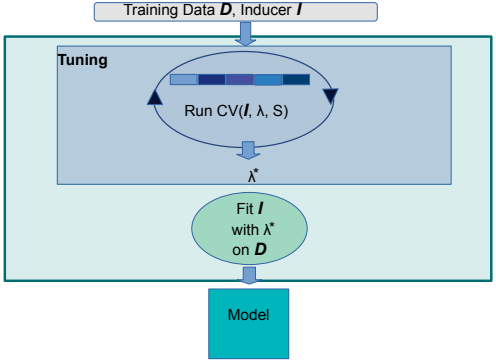
- During tuning, a learner is trained on the **training set**, evaluated on the **validation set**
- After the best model configuration $\boldsymbol{\lambda}^*$ has been selected, we re-train on the joint (training+validation) set and evaluate the model's performance on the **test set**.



TUNING AS PART OF MODEL BUILDING



- Effectively, the tuning step is now simply part of a more complex training procedure.
- We could see this as removing the hyperparameters from the inputs of the algorithm and making it “self-tuning”.



TUNING AS PART OF MODEL BUILDING / 2

More precisely: the combined training & validation set is actually the training set for the “self-tuning” endowed algorithm.