

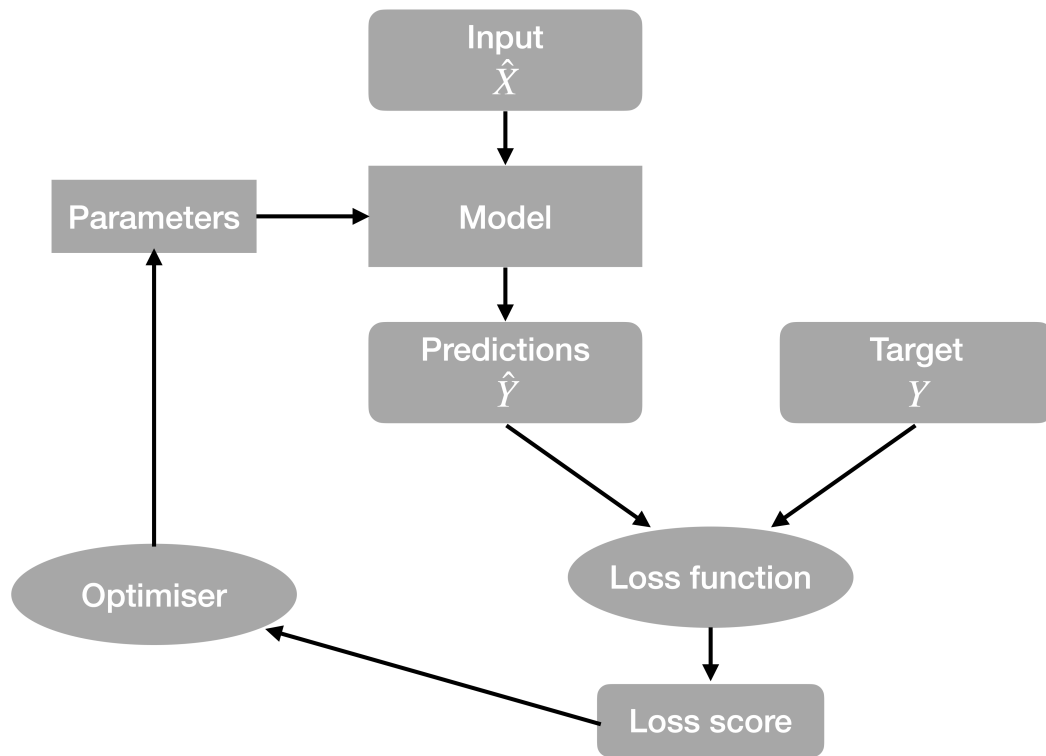


Applied Geodata Science I

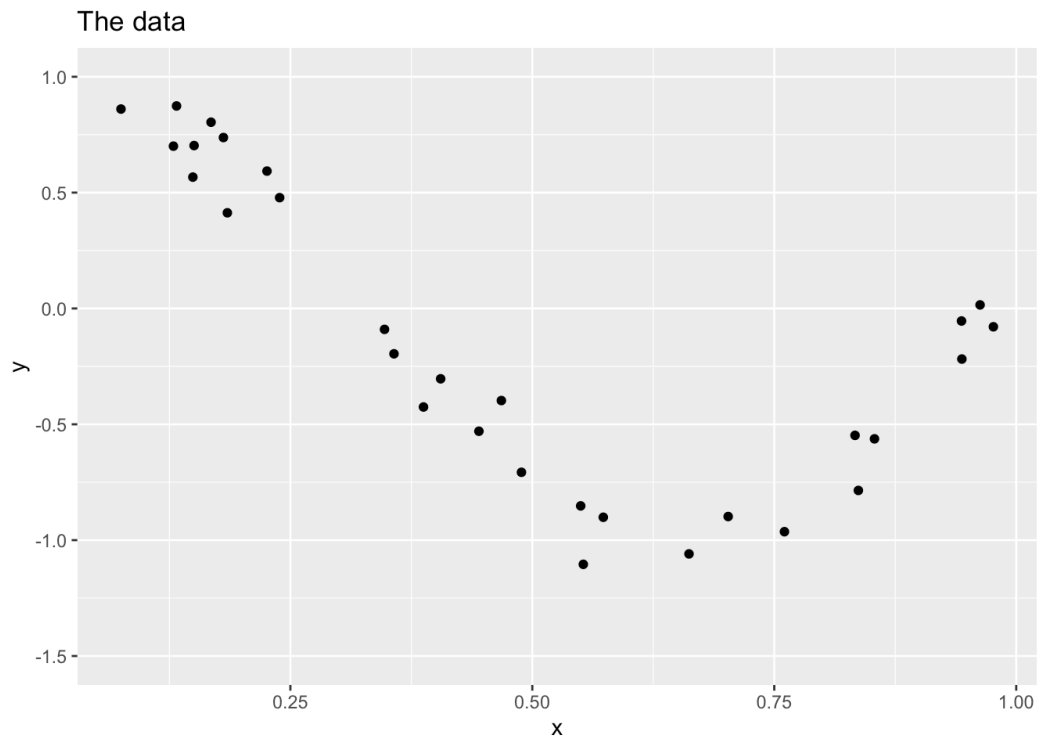
Session 11

Prof. Dr. Benjamin Stocker
Spring semester 2023





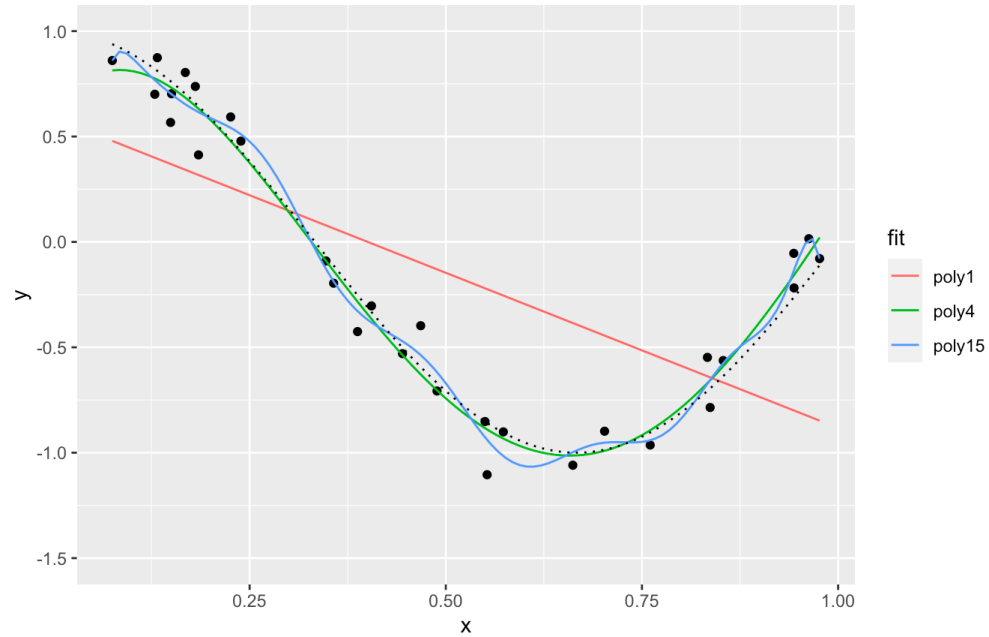
Example data



$$y = \sum_{n=0}^N a_n x^n$$

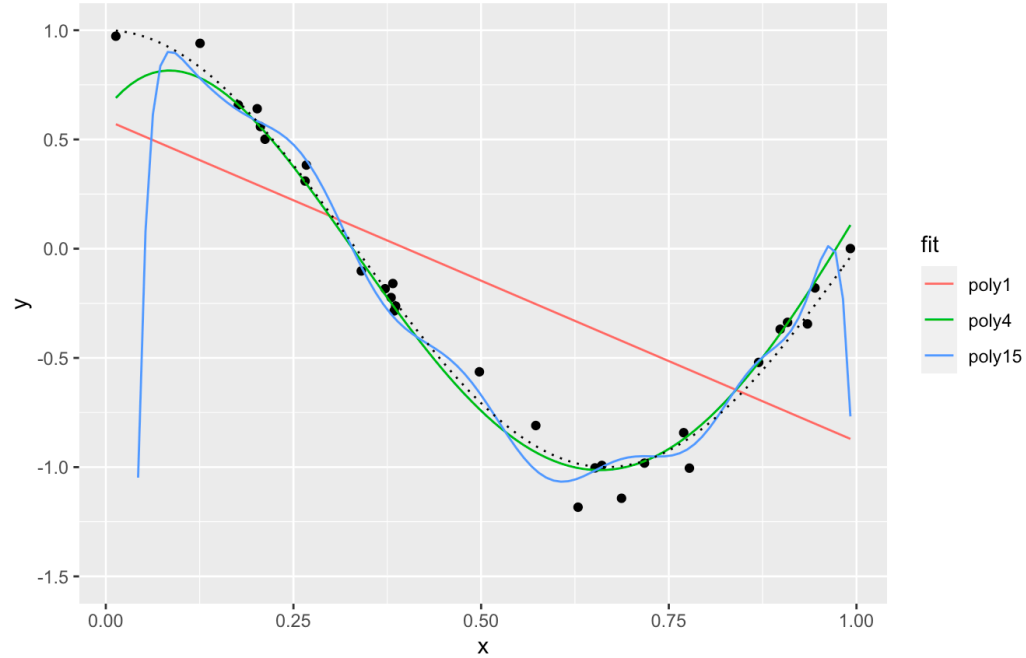
Training

RMSE: poly1 = 0.48 , poly4 = 0.11 , poly15 = 0.094



Validation

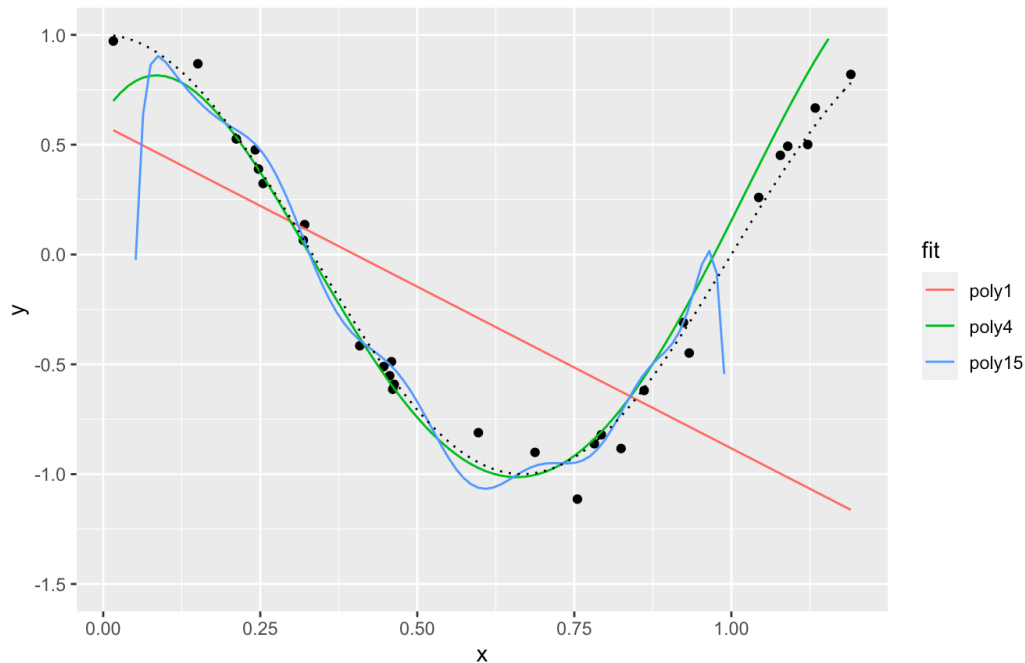
RMSE: poly1 = 0.45 , poly4 = 0.11 , poly15 = 2.8



Validation (with extrapolation)

Validation (with extrapolation)

RMSE: poly1 = 0.79 , poly4 = 0.15 , poly15 = 161



Testing vs. validation set

Initial split

training set

testing set

Testing vs. validation set

Initial split

training set

testing set

Resample

training set

validation
set

Testing vs. validation set

Initial split

training set

testing set

Resample

training set

validation
set

- Hyperparameter tuning
- Feature selection

K-fold cross validation

Initial split

training set

testing set

Folds

training set

validation
set

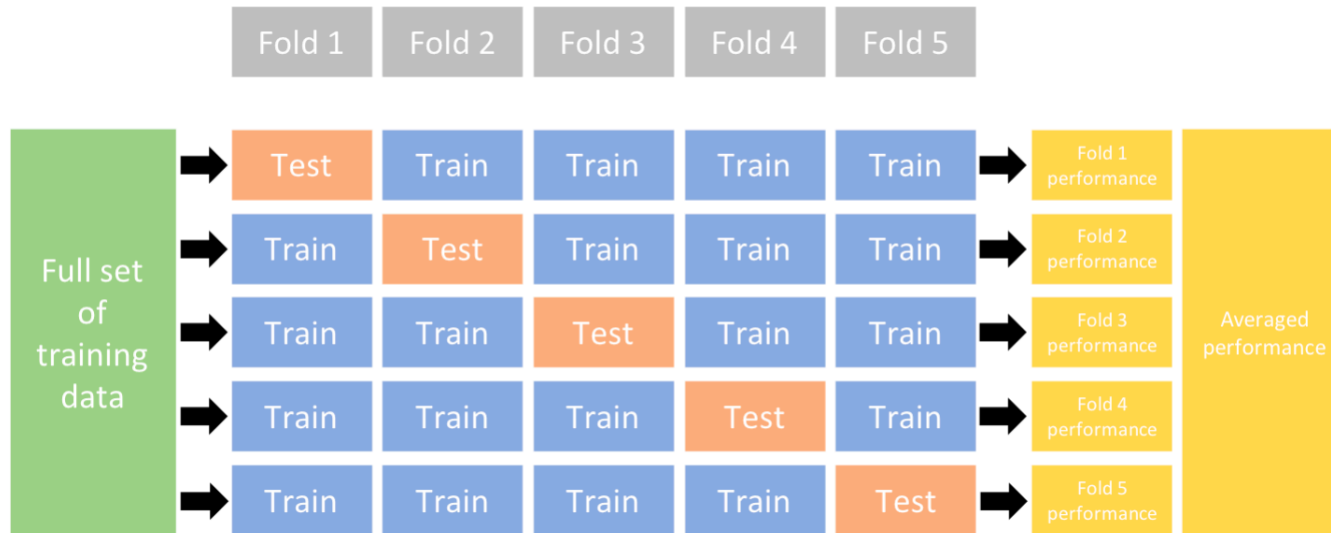
training set

validation
set

training set

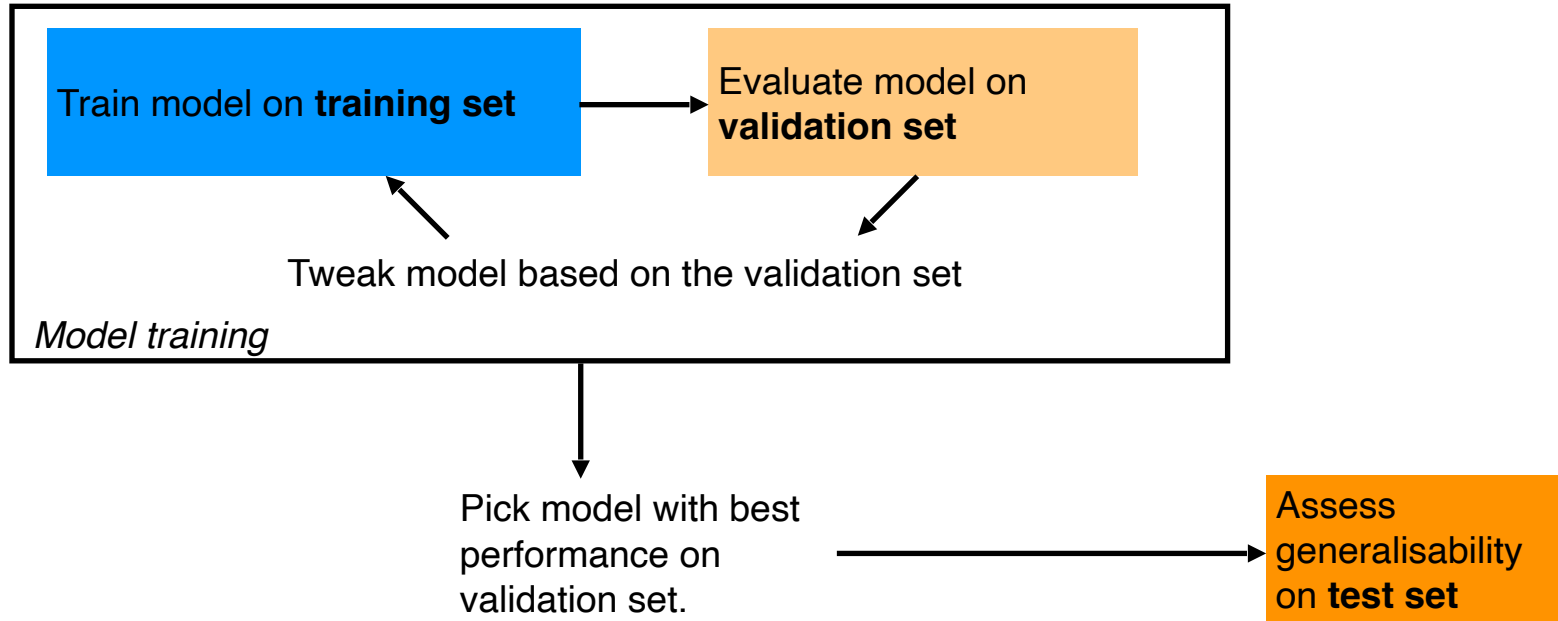
...

Resampling



Boehmke & Greenwell (2019) *Hands On Machine Learning in R*

Workflow of model training and testing



Workflow of model training and testing

