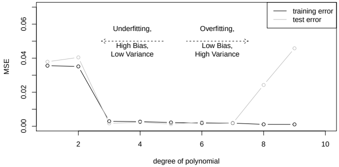
Introduction to Machine Learning

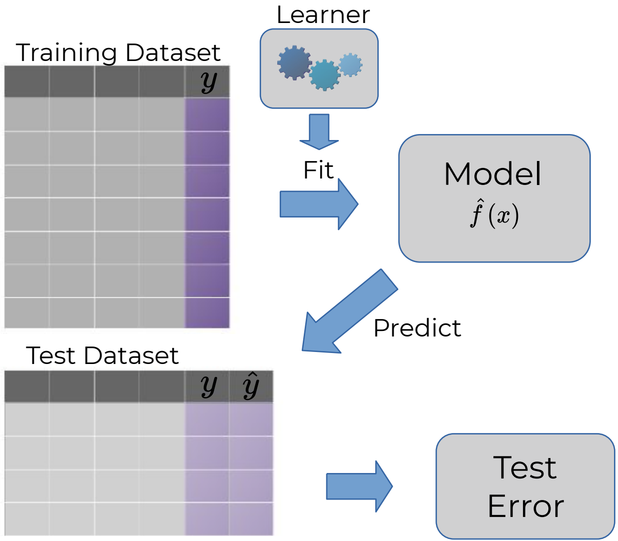
Evaluation: Test Error



Learning goals

- Understand the definition of test error
- Understand how overfitting can be seen in the test error

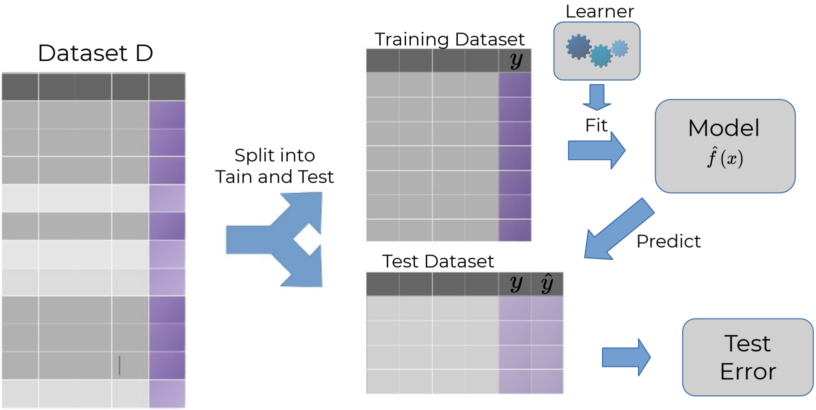
TEST ERROR



Learner
Training Dataset
y
Fit
Model
$\hat{f}(x)$
Predict
Test Dataset
y $\hat{y}$
Test
Error

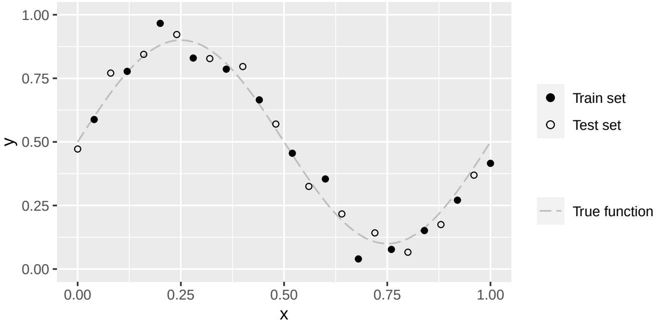
TEST ERROR AND HOLD-OUT SPLITTING

- Split data into 2 parts, e.g., 2/3 for training, 1/3 for testing
- Evaluate on data not used for model building



TEST ERROR

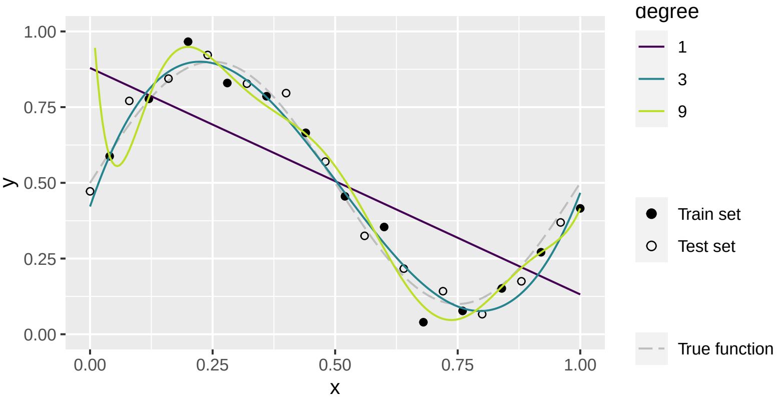
Let's consider the following example:
Sample data from sinusoidal function $0.5 + 0.4 \cdot \sin(2\pi x) + \epsilon$



Try to approximate with a d^{th}-degree polynomial:

$$f(\mathbf{x} \mid \boldsymbol{\theta}) = \theta_0 + \theta_1 x + \cdots + \theta_d x^d = \sum_{j=0}^{d} \theta_j x^j.$$

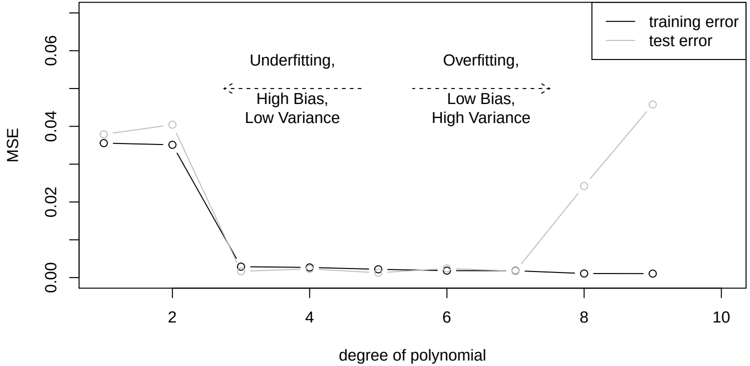
TEST ERROR



- d=1: MSE = 0.038: Clear underfitting
- d=3: MSE = 0.002: Pretty OK
- d=9: MSE = 0.046: Clear overfitting

TEST ERROR

Plot evaluation measure for all polynomial degrees:



Increase model complexity (tendentially)

- decrease in training error
- U-shape in test error
  (first underfit, then overfit, sweet-spot in the middle)

TEST ERROR PROBLEMS

- Test data has to be i.i.d. compared to training data.
- Bias-variance of hold-out:
  - The smaller the training set, the worse the model $\rightarrow$ biased estimate.
  - The smaller the test set, the higher the variance of the estimate.
- If the size of our initial, complete data set $\mathcal{D}$ is limited, single train-test splits can be problematic.

TEST ERROR PROBLEMS

A major point of confusion:

- In ML we are in a weird situation. We are usually given one data set. At the end of our model selection and evaluation process we will likely fit one model on exactly that complete data set. As training error evaluation does not work, we have nothing left to evaluate exactly that model.
- Hold-out splitting (and resampling) are tools to estimate the future performance. All of the models produced during that phase of evaluation are intermediate results.