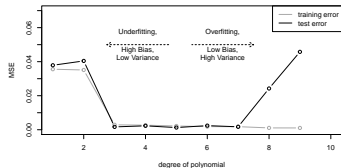
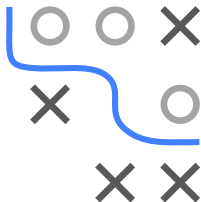


Introduction to Machine Learning

Evaluation Test Error



Learning goals

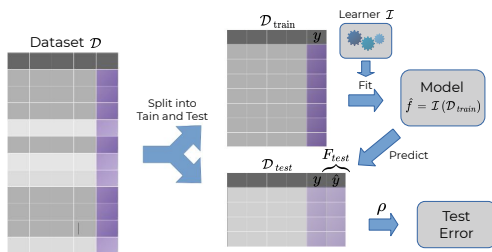
- Understand the definition of test error
- Understand that test error is more reliable than train error
- Bias-Variance analysis of holdout splitting

TEST ERROR AND HOLD-OUT SPLITTING

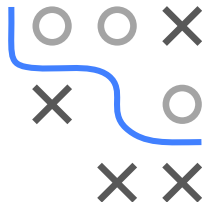
- Simulate prediction on unseen data, to avoid optimistic bias:

$$\rho(\mathbf{y}_{\text{test}}, \mathbf{F}_{\text{test}}) \text{ where } \mathbf{F}_{\text{test}} = \begin{bmatrix} \hat{f}_{\mathcal{D}_{\text{train}}}(\mathbf{x}_{\text{test}}^{(1)}) \\ \vdots \\ \hat{f}_{\mathcal{D}_{\text{train}}}(\mathbf{x}_{\text{test}}^{(m)}) \end{bmatrix}$$

- Partition data, e.g., 2/3 for train and 1/3 for test.

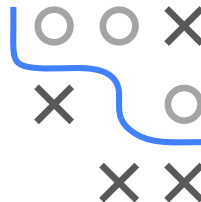
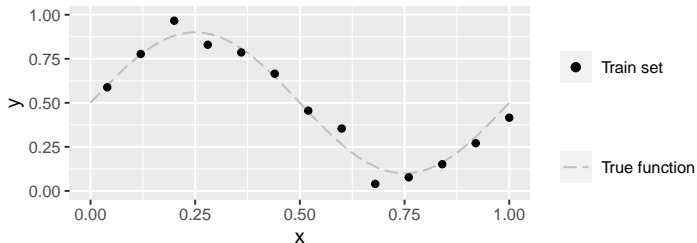


A.k.a. holdout splitting.



EXAMPLE: POLYNOMIAL REGRESSION

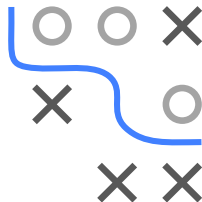
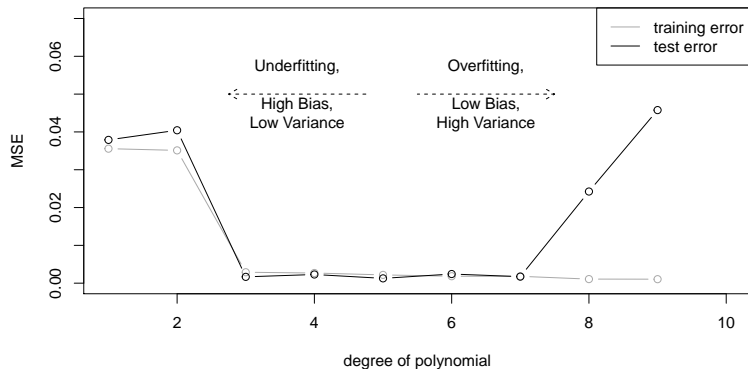
Previous example:



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

TEST ERROR

Let's plot train and test MSE for all d :



Increasing model complexity tends to cause

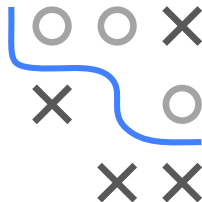
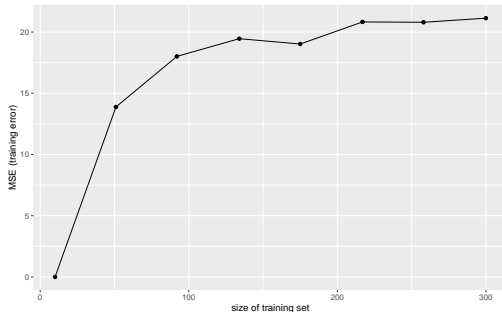
- a decrease in training error, and
- a U-shape in test error
(first underfit, then overfit, sweet-spot in the middle).

TRAINING VS. TEST ERROR

- Boston Housing data
- Polynomial regression (without interactions)

The training error...

- decreases with smaller training set size as it becomes easier for the model to learn all observed patterns perfectly.



BIAS AND VARIANCE

- Test error is a good estimator of GE, given a) we have enough data b) test data is representative i.i.d.
- Estimates for smaller test sets can fluctuate considerably – this is why we use resampling in such situations.

Repeated $\frac{2}{3} / \frac{1}{3}$ holdout splits:

iris ($n = 150$) and sonar ($n = 208$).

