Bias Variance Trade-off

- Intuition:
  - If the model is too simple, the solution is biased and does not fit the data
  - If the model is too complex then it is very sensitive to small changes in the data

Bias

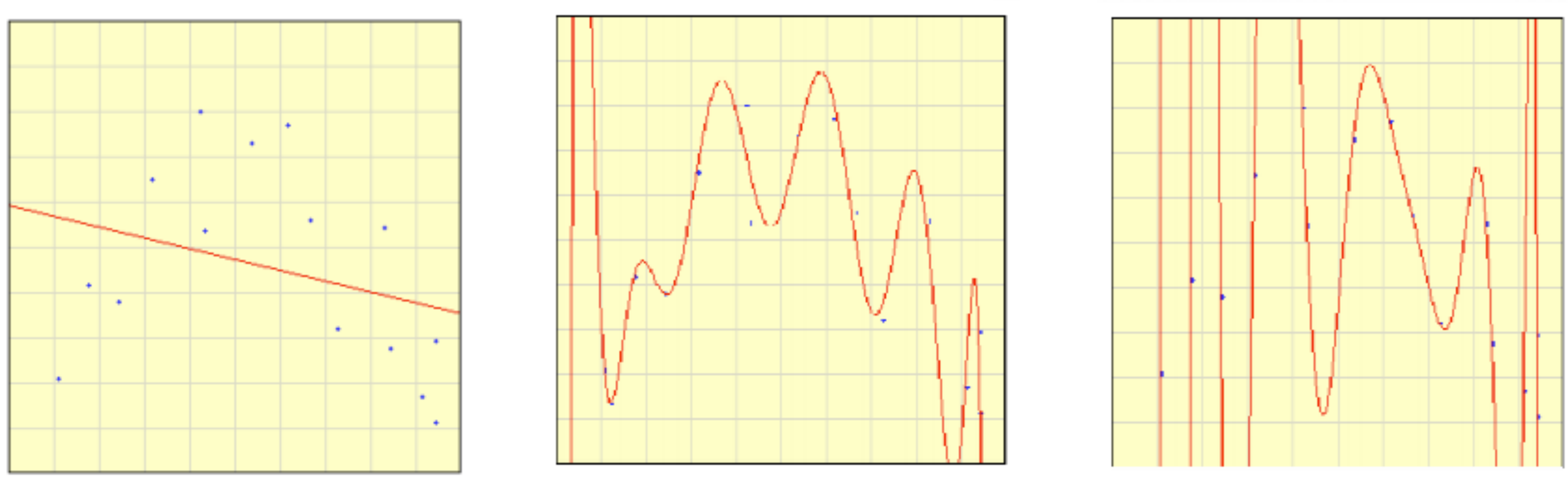
- If you sample a dataset D multiple times you expect to learn a different h(x)
- Expected hypothesis is $E_D[h(x)]$
- Bias: difference between the truth and what you expect to learn
  - $$bias^2 = \int_x \{E_D[h(x)] - t(x)^2\}^2 p(x)dx$$
  - Decreases with more complex models

Variance

- Variance: difference between what you learn from a particular dataset and what you expect to learn
  - $$variance = \int_x \{E_D[(h(x) - \bar{h}(x))^2]\}p(x)dx$$
    $$\bar{h}(x) = E_D[h(x)]$$
  - Decreases with simpler models

Bias-Variance Tradeoff

- The choice of hypothesis class introduces a learning bias
  - More complex class: less bias and more variance.



Training error

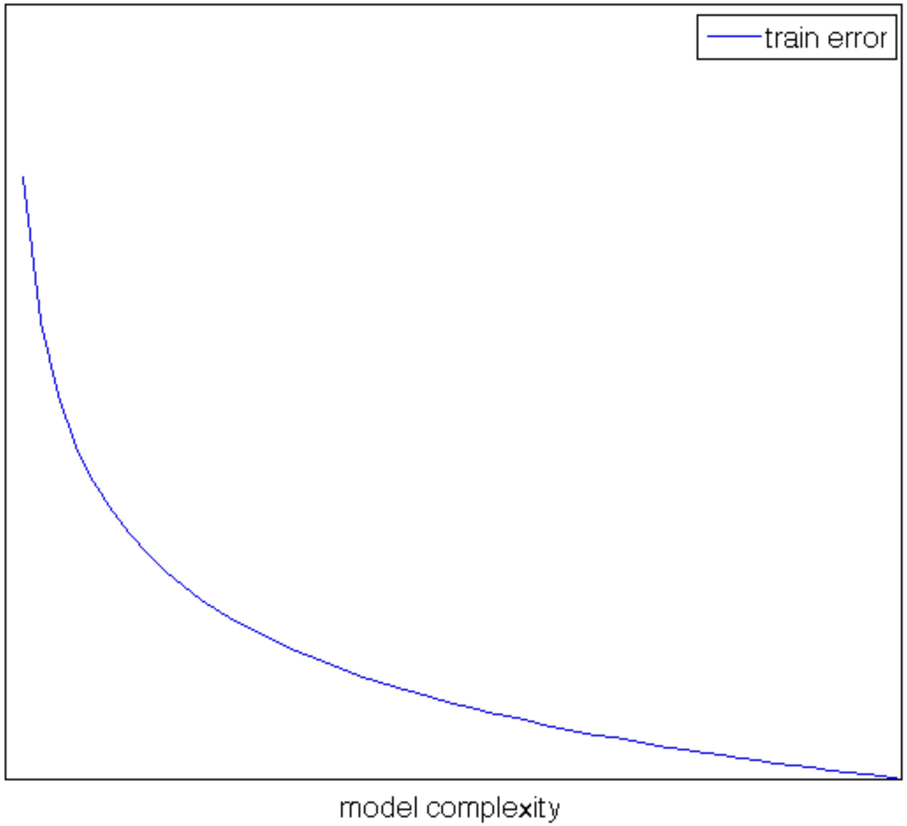
- Given a dataset
- Chose a loss function (L_2 for regression for example)

  - Training set error:

$$error_{train} = \frac{1}{N_{train}} \sum_{j=1}^{N_{train}} \Big(I(y_i \neq h(x)) \Big)$$

$$error_{train} = \frac{1}{N_{train}} \sum_{j=1}^{N_{train}} \Big(y_i - w.\mathbf{x_i} \Big)^2$$

Training error as a function of complexity

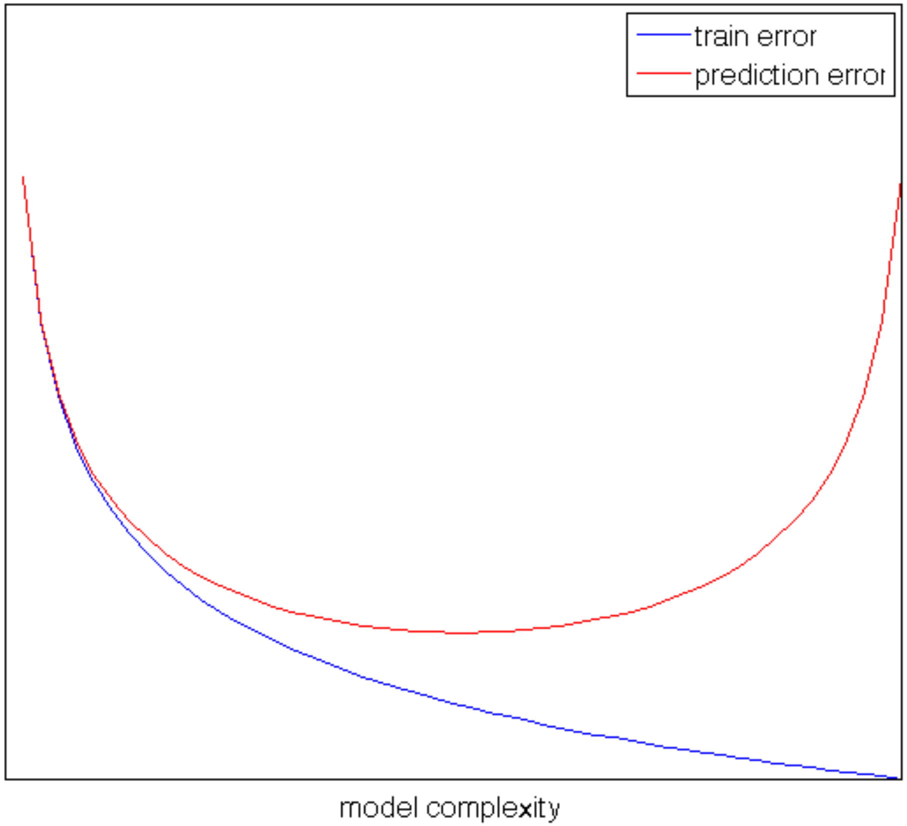


Prediction error

- Training error is not necessary a good measure
- We care about the error over all inputs points:

$$error_{true} = E_x\Big(I(y \neq h(x))\Big)$$

Prediction error as a function of complexity



Prediction error

- Training error is not necessary a good measure
- We care about the error over all inputs points:

$$error_{true} = E_x\Big(I(y \neq h(x))\Big)$$

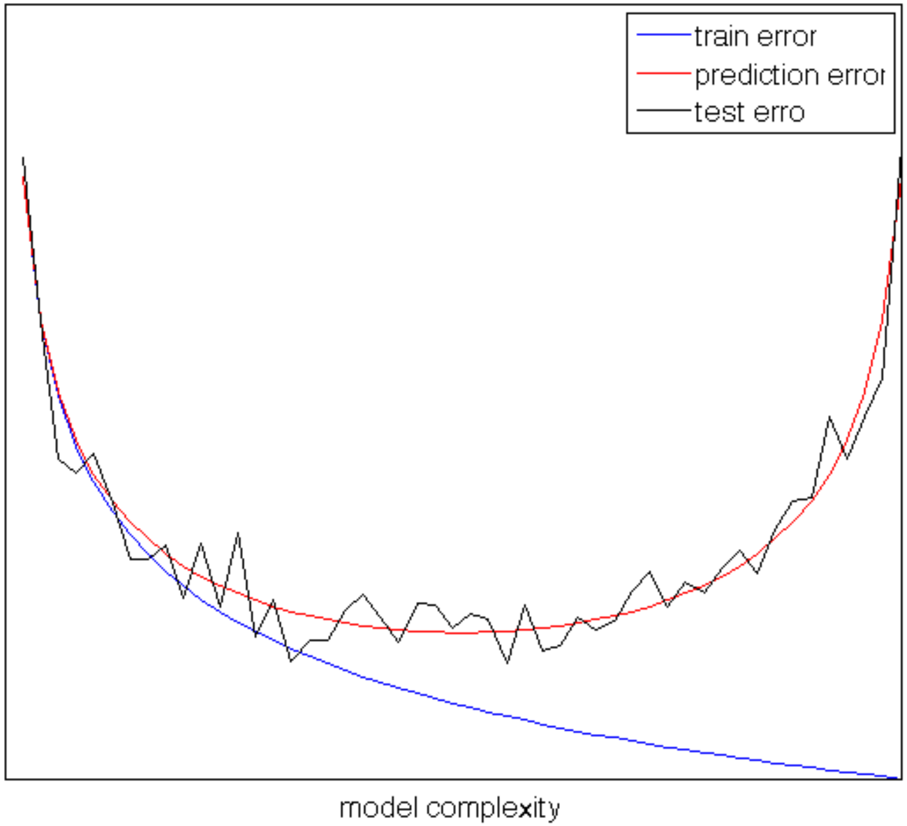
- Training error is an optimistically biased estimate of prediction error. You optimized with respect to training set.

Train-test

- In practice:
  - Randomly divide the dataset into test and train.
  - Use training data to optimize parameters.
  - Test error:

$$error_{test} = \frac{1}{N_{test}} \sum_{i=1}^{N_{test}} \Big(I(y_i \neq h(x_i)) \Big)$$

Test error as a function of complexity

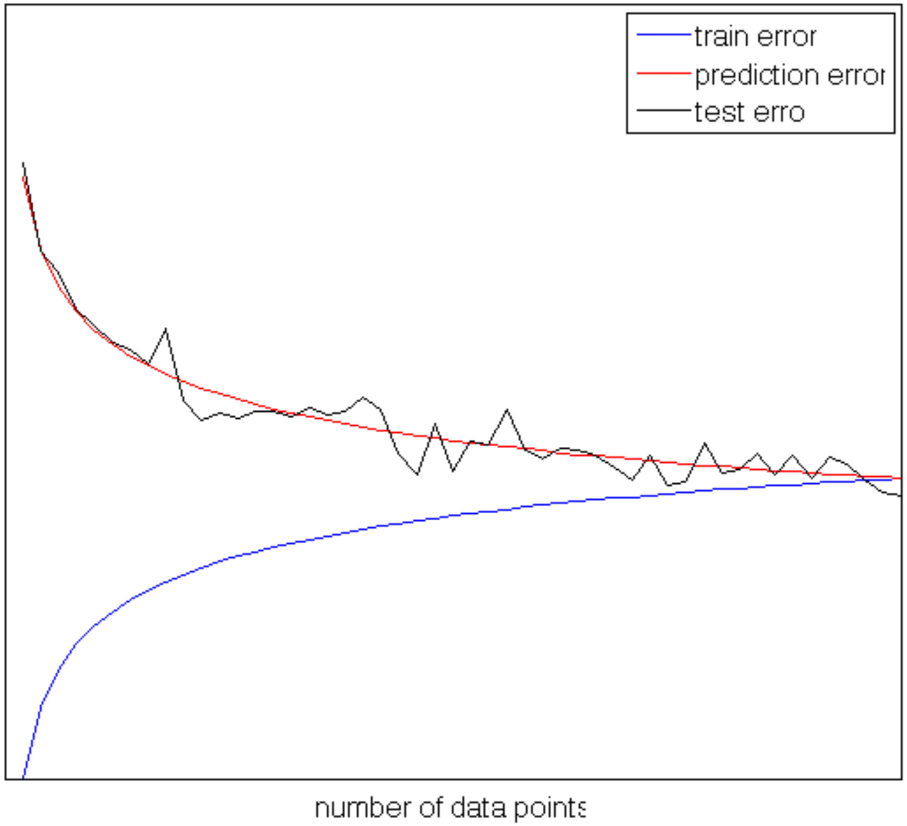


Overfitting

- Overfitting happens when we obtain a model h when there exist another solution h' such that:

$$[error_{train}(h) < error_{train}(h')] \wedge [error_{true}(h) > error_{true}(h')]$$

Error as a function of data size for fixed complexity



Careful

- Test set only unbiased if never ever do any learning on it (including parameter selection!).