Model evaluation and cross-validation

Model evaluation

- Goal: want to evaluate the generalization performance of the model, and possibly use the evaluation to select the model;

- Different types of error
  - Test error / generalization error / prediction error: $Err_{\mathcal{T}} = E_{X^0,Y^0}[L(Y^0, \hat{f}(X^0))|\mathcal{T}]$
  - Expected error: $Err = E_{\mathcal{T}}[Err_{\mathcal{T}}]$
  - Training error: $\overline{err} = \frac{1}{N}\sum_{i=1}^{N} L(y_i, \hat{f}(x_i))$
  - In-sample error: $Err_{in} = \frac{1}{N}\sum_{i=1}^{N} E_{Y^0}[L(Y_i^0, \hat{f}(x_i))|\mathcal{T}]$

- Mallow's C_p, AIC, BIC provide estimates of Err_{in};

- Cross-validation estimates Err.

Necessity of model evaluation

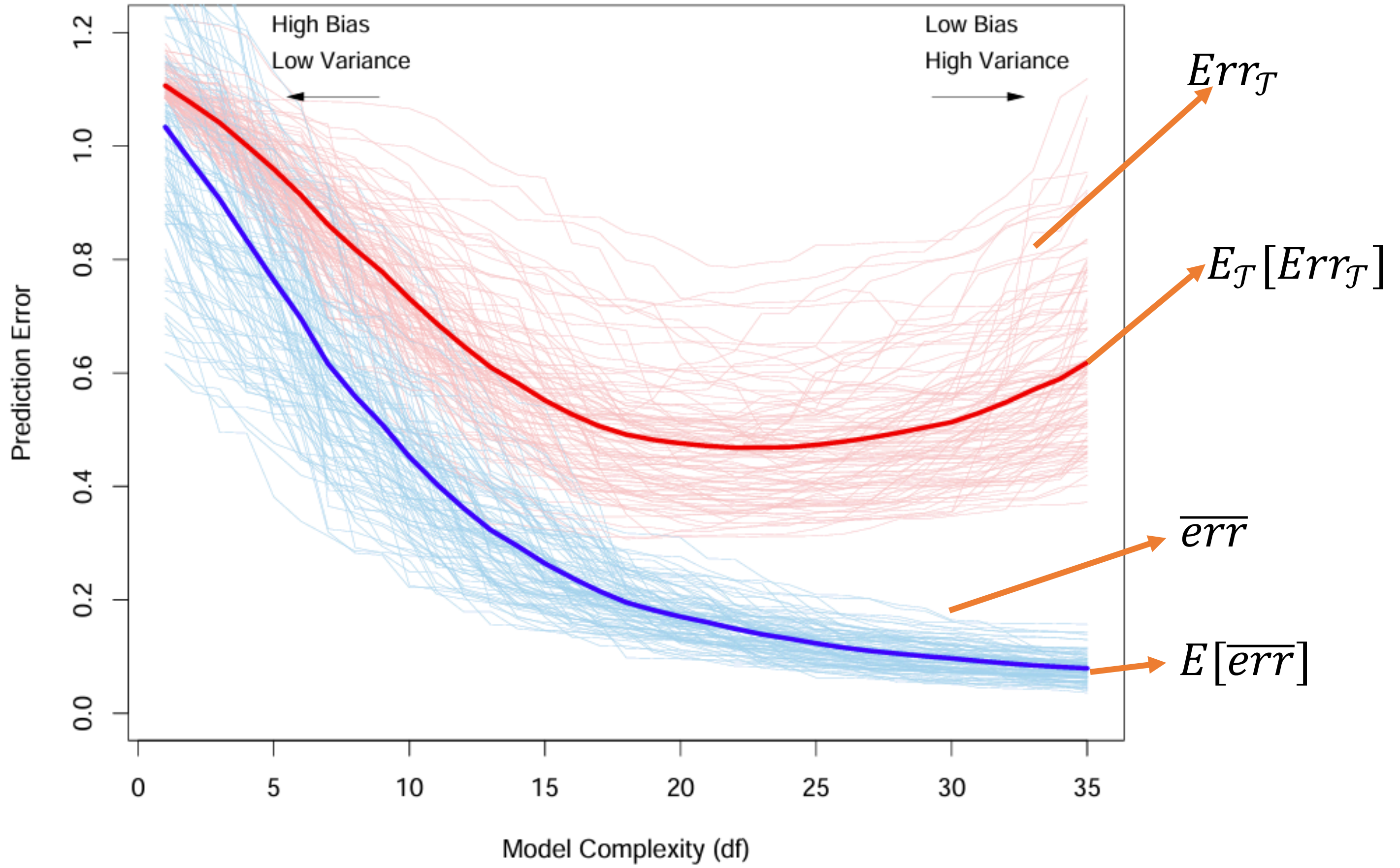
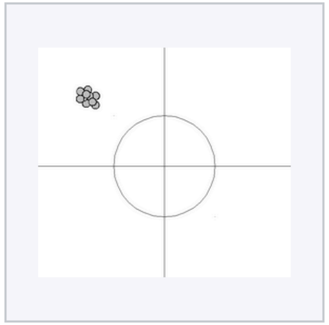
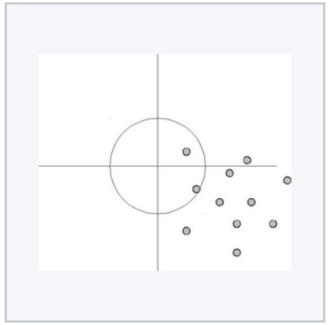


FIGURE 7.1. *Behavior of test sample and training sample error as the model complexity is varied. The light blue curves show the training error* $\overline{\mathrm{err}}$, *while the light red curves show the conditional test error* $\mathrm{Err}_{\mathcal{T}}$ *for* 100 *training sets of size* 50 *each, as the model complexity is increased. The solid curves show the expected test error* Err *and the expected training error* $\mathrm{E}[\overline{\mathrm{err}}]$.

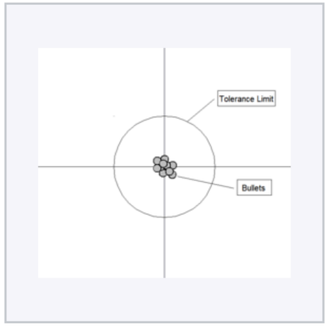
- Model bias: $E\hat{f}(x_0) - f(x_0)$
- Model variance: $E[\hat{f}(x_0) - E\hat{f}(x_0)]^2$



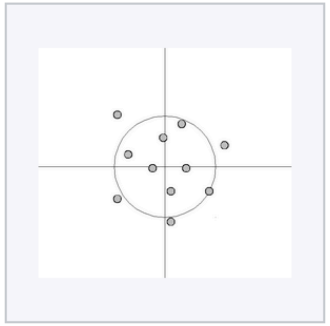
High bias, low variance



High bias, high variance



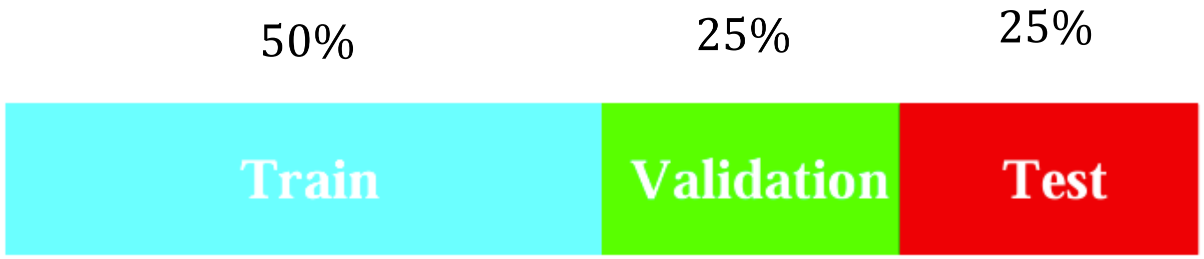
Low bias, low variance



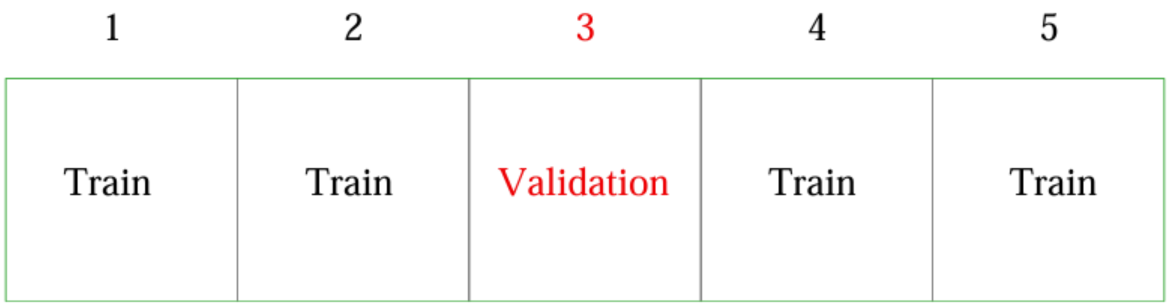
Low bias, high variance

Cross validation

Typical dataset split



5-fold cross validation



Cross-validation

- With a multistep modeling procedure, cross-validation has to be applied to the entire sequence of modeling steps: consider the wrong/right way of cross-validation example in ESL;

- Several replicates can avoid the case of uneven partition in one specific replicate;

- 5-fold/ 10-fold are mostly used.