Overfitting, Regularization
Bias-Variance Decomposition

Lyle Ungar

Computer and information Science

Learning Objectives
Understand training and testing error vs. complexity
Use cross validation
Bias-variance decomposition and trade-off
Overfitting and model complexity

- Test error = Training error + model complexity
Estimating test error

- Training set ("in sample") vs. Testing set ("out of sample")
- Leave one out cross validation (LOOCV)
  - With and without replacement
- 10-fold CV
- Two uses of CV
  - Picking hyperparameters
  - Estimating true error
Bias and Variance

**Bias**
- Are the estimates in expectation over training sets high or low?
- Can be of model parameters or of predictions

**Variance**
- How much do the estimates change if you change the training set.

\[
\begin{align*}
Bias(\hat{\theta}) &= E[\hat{\theta} - \theta] = E[\hat{\theta}] - E[\theta] \\
Var(\hat{\theta}) &= E[(\hat{\theta} - E[\hat{\theta}])^2]
\end{align*}
\]

Bias and Variance

- Bias and Variance are in expectation over training sets
  - What does that look like?
  - How does that relate to model complexity?

\[
\text{Bias}(\hat{\theta}) = E[\hat{\theta} - \theta] = E[\hat{\theta}] - E[\theta]
\]

\[
\text{Var}(\hat{\theta}) = E[(\hat{\theta} - E[\hat{\theta}])^2]
\]
Bias Variance Tradeoff

- Test Error = Variance + Bias^2 + Noise

\[
\mathbb{E}_{x,y,D}[ (h(x; D) - y)^2 ] = \mathbb{E}_{x,D}[ (h(x; D) - \overline{h}(x))^2 ] + \mathbb{E}_x[ (\overline{h}(x) - \overline{y}(x))^2 ] + \mathbb{E}_{x,y}[ (\overline{y}(x) - y)^2 ]
\]

- Variance
- Bias^2
- Noise

\(D=\text{training data, } x,y = (\text{infinite) test data}\)

\(y(x) \text{ is the label; } h(x) \text{ is the model prediction}\)

See the [wiki page](#) for the derivation
Bias Variance Tradeoff - OLS

- What is the bias of the estimate of \( w \)?
- What is the bias of the estimate of \( y \)?
- What is the variance of \( y \)?
- What is the variance of the estimate of \( y \)?
Bias Variance Tradeoff - OLS

- Test Error = Variance + Bias² + Noise

\[
E_{x,y,D}[(h(x;D) - y)^2] = E_{x,D}[(h(x;D) - \bar{h}(x))^2] + E_x[(\bar{h}(x) - \bar{y}(x))^2] + E_{x,y}[(\bar{y}(x) - y)^2]
\]

\[
\text{Error} = E[(y - \hat{y})^2] = \text{Bias}(\hat{y})^2 + \text{Var}(\hat{y}) + \sigma^2
\]
What you should know

◆ Overfitting and model complexity
◆ Cross validation
  ● LOOCV, 10-fold CV
  ● sampling with and without replacement
  ● uses of CV: setting hyper-parameters, estimating test error
◆ Learning curves
◆ Bias-variance trade-off
◆ Unbiased estimator