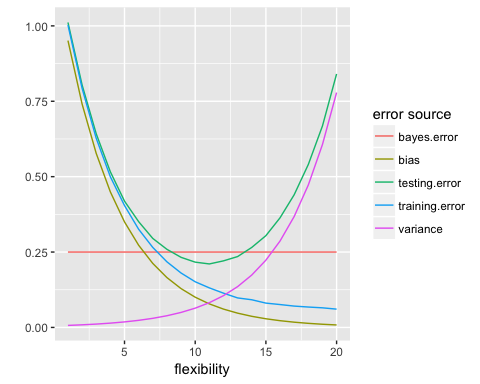
Week 1 Assignment

David Russo

1/15/2017

* 2.4 #1
  1. A flexible statistical model would perform **better** than an inflexible statistical model when *n* is extremely large and *p* is small. With a large sample size and small number of predictors, we are less likely to overfit the data. Furthermore, flexible models will have lower bias than inflexible models which will contribute to a lower overall test error.
  2. A flexible statistical model would perform **worse** than an inflexible statistical model when *n* is small and *p* is extremely large. Using a flexible model with a small sample size would introduce a large amount of variance due to overfitting.
  3. A flexible statistical model would perform **better** than an inflexible statistical model when the relationship between the predictors and the response is highly non-linear. Flexibile statistical models are better equipped to handle non-linearities than inflexible statistical models. Inflexible statistical models will often only detect linear relationships between predictors and the response.
  4. A flexible statistical model would perform **worse** than an inflexible statistical model when the variance of the error terms is high. If the variance of the error terms is high, a flexible model will be likely to overfit the data, resulting in higher variance and higher test error.
* 2.4 #2
  1. Regression setting; inference; n = 500, p = profit, number of employees, industry
  2. Classification setting; prediction; n = 20; p = price charged for the product, marketing budget, competition price
  3. Regression setting; prediction; n = 52; p = % change in US market, % change in British market, % change in German market
* 2.4 #3
  1. 
     1. The bias decreases with model flexibility because models that are more flexible are able to better fit the training data. As the fit improves, the bias decreases.
     2. The variance increases with model flexibility because more flexible models tend to overfit the training data. When the model is applied to new data that wasn't used to train the model, the model will have a high degree of variation. This is because the highly flexible model can only model the training data, and not new data.
     3. The training error decreases with model flexibility because more flexible models are able to model the training data well. As flexibility increases, data overfitting increases which decreases bias because the model is able to predict the training data well.
     4. The testing error has a U-shape because it demonstrates the bias-variance trade off. As model flexibility increases, bias decreases but variance increases. The minimum point of the U-shape represents the degree of flexibility that yields the best trade off between bias and variance in terms of predictive ability on new data. Models that are not flexible enough will have high bias but low variance when tested on new data. Models that are too flexible will overfit the data and have low bias but high variance when tested on new data.
     5. The Bayes error is a horizontal line because it is constant. It represents the true noise in the data that cannot be captured by the model.
* 2.4 #4
  1. Classification
* 2.4 #8