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
* 2.4 #4
* 2.4 #8