1. Simulated data
   1. Generated predictor/error
   2. Generated response
   3. Best Subset
   4. Forward/Backward
   5. Lasso
   6. When altering the response vector from a 3rd degree polynomial of x, with non-zero coefficients for every single term, to a 7th degree polynomial with only an intercept and 7th degree coefficient, the results were considerably different. All 3 of BIC, Cp, and Adjusted R^2 in the best subset selection selected an 8 variable model, but with non-zero coefficients for every single term included. Lasso, on the other hand, predicted an optimal model with coefficients that only involved 0 coefficients other than the intercept.
2. College
   1. Split
   2. LM test error = 1135758
   3. Ridge Regression test error = 976261.5
   4. Lasso test error = 1115901 (All 18 coefficient estimates were non-zero)
   5. PCR test error = 1135758 (M=17 chosen by CV)
   6. PLS test error = 1135758 (M=17 chosen by CV)
   7. The error rates are all quite high in the number of college applications received. Apart from ridge regression, there isn’t too much difference regarding the test errors from each of the 5 different methods used. Since PCR and PLS returned the exact same test error as ordinary least squares, this means that all of the predictors were needed to obtain the least amount of unexplained data in the relationship between the response and its predictors, thereby affirming that all of the predictors are at least somewhat related to the response. This is also why lasso is only a slight improvement, as it focuses more on variable selection rather than impact reduction, which ridge regression does, allowing it to in turn perform significantly better than standard least squares. It is highly likely that many of the variables were not strongly correlated with the response, and so dampening many of their impacts allowed for a much smaller test error and a more accurate model of the relationship between the number of applications and the other variables.
3. Simulation: Training error vs test error
   1. Generated training set with p=20 and n=1000 according to Y=XB+e, where some of the B=0.
   2. Split.
   3. Best subset selection on training set
   4. Plotted test MSE with best model of each size
   5. The test MSE takes its minimum value when the model includes 10 variables. This matches exactly with the true coefficient vector in this case, which contained 10 non-zero coefficients and 10 zero coefficients. The test MSE plotted against number of variables included decreased steeply until the 10-variable model, reached a minimum at 10, and proceeded to monotonically increase very slightly afterward. This makes sense, as there is a penalty incurred for higher variable models that, for those sizes, is not offset by a sufficient reduction in the test error.
   6. As stated above, the 10-variable model which minimizes the test MSE contained the same number of non-zero variables. In addition, all of the non-zero coefficient estimates in the best subset selection model had an error margin that was at most a multiple of 10^-3 in size. This shows that best subset selection, after performing an exhaustive search, can approximate the true relationship between the predictors and response very well.
   7. The overall shape of the plot of the square root of the sum of squared difference between the coefficients in every sized model from best subset selection and the actual coefficient vector matches the overall trend of the test MSE plot from d). The minimum occurs at the model of size 10, with significant decrease occurring before and minimal monotonic increase occurring after that point. The only real difference is that the test MSE plot was concave up in its decreasing section, while the decreasing section of this plot was concave down in its decreasing section, likely due to the square root involved here in lieu of the linear constant 1/n from the test MSE plot.
4. Save for later, use as review if necessary(end of page 264)