HW4 - Least Squares Review

William Morgan, Jared Scolaro, Mitchell O'Brien

February 19, 2018

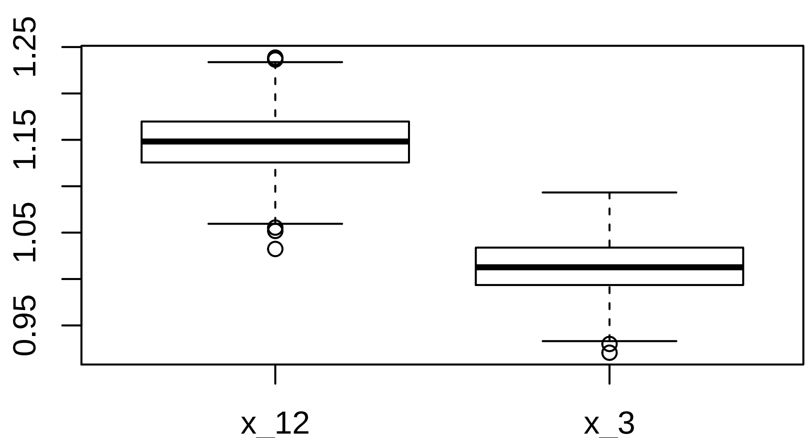
Problems

---

1. Out-of-sample predictive performance of variable subsets

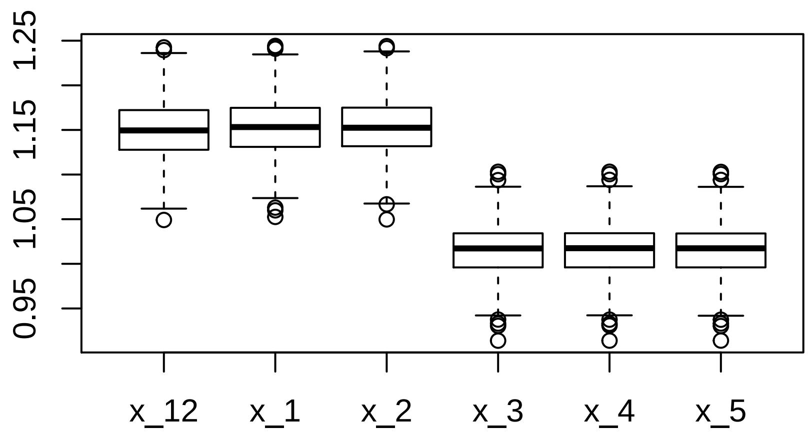
1a. Is X_3 really better than X_1, X_2 in terms of test error?

Based on the class example, it is pretty obvious that X_3 outperformed X_1, X_2 in nearly every model-estimation iteration. To be certain, we increase the number of iterations in the test to see if our conclusion changes.



This is pretty solid evidence that having X_3 alone results in a better performing model (in terms of test error) than X_1 and X_2 together.

1b. Modify the code to compare the model containing X_1 and X_2 as predictors against all subsets containing only one variable



Based on the box plot, it is clear that the three models that include X_1 and/or X_2 underperform compared to the models that exclude them.

---

2. Properties of Least Squares

2.1 $\hat{\beta}$ by Matrix Operations

- Compute $\hat{\beta}$ and check the first order conditions for the data from problem 1

```
# Define model matrix
X <- as.matrix(dta[, 2:6], ncol = 5)
X <-  cbind(1, X)

Y <- as.matrix(dta[, 1], ncol = 1)

# Find (X'X)^-1
X_tX_i <- t(X) %*% X %>%
  solve()

# Find Beta
bhat <-  X_tX_i %*% t(X) %*% Y

# Check FOCs
t(X) %*% (Y - X %*% bhat)
```

```
##                y
##    -1.583336e-08
## x1 -7.795033e-08
## x2 -7.794550e-08
```

```
## x3 -8.653670e-09
## x4 -8.737823e-09
## x5 -8.736231e-09
```

2.2 $\hat{\sigma}$ and Standard Errors

Get $\hat{\sigma}$ and $se(\hat{\beta}_i)$ for the data from problem 1 directly from the formulas using R matrix operations and vector calculations

```
# Grab predictions and find sample variance of Y
yhat <-  X %*% bhat
shat <-  (1/1994) * sum((Y - yhat)^2)

# Find Std. Error of beta_hat
inv = diag(solve(t(X) %*% X))
std_err <- sqrt(inv * shat)

# Grab results from lm() function
fit <- lm(y ~ ., data = dta)
lm_shat <-  summary(fit)$sigma
lm_sterr <-  coef(summary(fit))[,2]

cat('The direct calculation of sigma hat is ', shat, '\n')
```

```
## The direct calculation of sigma hat is  1.013872
```

```
cat('The lm() calculation of sigmat is ', lm_shat, '\n')
```

```
## The lm() calculation of sigmat is  1.006912
```

```
cat('The direct calculation of the standard errors is ', std_err, '\n')
```

```
## The direct calculation of the standard errors is  0.08915702 0.007685841 0.007741829 10.8887 7.77313
```

```
cat('The lm() calculation of the standard errors is ', lm_sterr, '\n')
```

```
## The lm() calculation of the standard errors is  0.08915702 0.007685841 0.007741829 10.8887 7.773133
```

2.3 Correlations

- 2.3a: How do the outputs of the regression between the demeaned and raw data compare?

Only the intercept β_0 has been affected by the demeaning; Its estimate has increased to 1.211 and its standard error has decreased by a solid margin (relative to the standard error of the non-demeaned regression)

- 2.3b: Why are the residuals uncorrelated with the fitted values?

The residuals are uncorrelated with the fitted values by construction. Specifically, the first order condition that the gradient of the loss function equals 0 implies that the residuals $y - X\hat{\beta}$ are orthogonal (i.e. uncorrelated) to each column of X

- 2.3b: Square the correlation between y and yhat; How does it compare with R^2?

```
## [1] 0.2433844
```

The squared correlation between y and yhat is equal to the R^2 from the regression (save for some rounding error)

3. Orthogonalized Regression

3a: How do the coefficients from the last regression compare to the previous?

The estimates of the coefficients do not change, but the standard error of the estimates increase slightly

3b: How does the e_5 coefficient compare to the x_5 coefficient in the previous problem?

The coefficients and their standard errors are equivalent

3c: What is this number? Confirm that this is the standard error for the coefficient of x_5

The number outputted from the code is the standard error of the estimate for the coefficient on e_5

```r
# Refit the data with x5 in the model
fit <- lm(y~., dta)

# Extract the standard error of the coefficient for x5
coef(summary(fit))[6,2]
```

```
## [1] 7.835586
```

3d: What is the R^2 from the regression of x_5 on X_1, X_2, X_3, X_4?

Looking to the output that's already written on the assignment, we can see that the R^2 from that model is .2433

3e: Run the regression of y on just x5. How does the SE for the coefficient for x5 compare to the SE of the one in the full model? Explain why they are so different

```
## [1] 0.07743788
```

The large difference between the two standard errors is a result of multicollinearity between x_3, x_4, and x_5. This makes it difficult to explain how responsible each variable is for variation in Y and leads to inflated standard errors.

---

4. Predictive Variance

Suppose we have training data (X, y) and x_f at which we wish to predict the future Y_f. Then our usual prediction is $\hat{Y_f} = x_f\hat{\beta}$. Obtain a nice matrix formula for:

$$Var[E_f] = Var[Y_f - \hat{Y_f}]$$

In this answer we use the fact that Y are iid observations and thus have $cov(Y, Y') = 0$

$$\begin{aligned} Var[E_f] &= Var[Y_f] + Var[\hat{Y_f}] + 2cov[Y_f, \hat{Y_f}] \\ &= Var[Y_f] + Var[X_f\hat{\beta}] \\ &= \sigma^2 + X_f(\sigma^2(X^TX)^{-1})X_f^T \end{aligned}$$

5. R-squared

Show that the square of the correlation between y and the fitted values is indeed the same as the usual formula for R-squared

$$cor(\hat{y}, y)^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

To simplify the algebra of this a little bit, we assume that Y is demeaned. We begin with the formula for correlation between y and $\hat{y}$:

$$cor(y, \hat{y}) = \frac{\sum y_i \hat{y}_i}{\sqrt{\sum y_i^2 * \sum \hat{y_i}^2}}$$

Rewrite using inner product notation:

$$cor(y, \hat{y}) = \frac{\langle y_i, \hat{y}_i \rangle}{\sqrt{\langle y_i, y_i \rangle}\sqrt{\langle \hat{y}_i, \hat{y}_i \rangle}}$$

Note that since $\hat{y}$ and ϵ are orthogonal, the following statement holds:

$$\langle \hat{y}_i, y \rangle = \langle \hat{y}, \hat{y} + \epsilon \rangle = \langle \hat{y}, \hat{y} \rangle + \langle \hat{y}, \epsilon \rangle = \langle \hat{y}, \hat{y} \rangle + 0 = \langle \hat{y}, \hat{y} \rangle$$

Square the expression and reduce:

$$cor(y, \hat{y}) = \frac{\langle \hat{y}_i, \hat{y}_i \rangle}{\langle y_i, y_i \rangle}$$

Square the previous expression to get the statement we sought to show.