Kaicheng Luo

Assignment 2

September 16, 2019

Problem 1

Proof : Consider the vector case,

$$Y = X\beta + \mu \tag{1}$$

where X is a matrix of independent variables with a column of constant 1. To fit the model using OLS, we minimize sum of squares

$$\begin{aligned}\hat{\beta} &= \arg\min_{\beta} \{u^T u\} \\ &= \arg\min_{\beta}[(Y - X\beta)^T (Y - X\beta)] \\ &= \arg\min_{\beta} (Y^T Y - Y^T X\beta - \beta^T X^T Y + \beta^T X^T X\beta)\end{aligned} \tag{2}$$

F.O.C.

$$\frac{\partial u^T u}{\partial \beta^T} = -X^T Y + (X^T X)\hat{\beta} = 0 \tag{3}$$

Consider the first row of X^T, which should be a vector of "1"s.

$$\begin{vmatrix}1 & 1 & ... & 1\end{vmatrix} (Y - X\hat{\beta}) = 0 \tag{4}$$

Hence we got

$$\sum_{i=1}^{n} \mu_i = 0 \tag{5}$$

Problem 2

Proof for symmetry :

$$\begin{aligned}H^T &= [X(X^T X)^{-1} X^T]^T \\ &= X[(X^T X)^{-1}]^T X^T \\ &= X[(X^T X)^T]^{-1} X^T \\ &= X(X^T X)^{-1} X^T \\ &= H\end{aligned} \tag{6}$$

Proof for idempotency :

$$\begin{aligned}H^2 &= X(X^T X)^{-1} X^T X(X^T X)^{-1} X^T \\ &= X(X^T X)^{-1} X^T \\ &= H\end{aligned} \tag{7}$$

Problem 3

Proof for symmetry :

$$\begin{aligned} Q^T = (I-H)^T &= I^T - [X(X^TX)^{-1}X^T]^T \\ &= I - X[(X^TX)^{-1}]^TX^T \\ &= I - X[(X^TX)^T]^{-1}X^T \\ &= I - X(X^TX)^{-1}X^T \\ &= I - H = Q \end{aligned} \tag{8}$$

Proof for idempotency :

$$\begin{aligned} (I-H)^2 &= I^2 + X(X^TX)^{-1}X^TX(X^TX)^{-1}X^T - 2IH \\ &= I + X(X^TX)^{-1}X^T - 2H \\ &= I - H \end{aligned} \tag{9}$$

Problem 4

Proof :
For any idempotent matrix A, its eigenvector α, and its eigenvalue λ, we have

$$A^2 = A \qquad A\alpha = \lambda\alpha \tag{10}$$

Using the properties of eigenvalues,

$$\begin{aligned} \lambda\alpha &= A\alpha \\ &= A^2\alpha = A\lambda\alpha \\ &= \lambda A\alpha = \lambda^2\alpha \end{aligned} \tag{11}$$

Hence $\lambda = \lambda^2$, $\lambda = 0$ or $\lambda = 1$

Problem 5

y_0 is, in most cases, more meaningful. Note that the intercept b_0 is the mean of dependent variable when all predictors = 0. But when 0 is out of the range of data, that value becomes meaningless. When you center X so that a value within the dataset becomes 0, the intercept becomes the mean of Y at the value you centered on. So in reality, mean-centering the data will help us interpret the intercept of our result.

Problem 6

Implementing Regression by Successive Orthogonalization:
Note that in our expectation, Gamma[] shall be an upper triangular matrix
z[] shall be the residual matrix, and Beta[] shall be a vector showing coefficents

```r
data <- read.csv("Advertising.csv", row.names = "X")
z <- matrix(nrow = nrow(data), ncol = ncol(data)+1)
Gamma <- matrix(nrow = nrow(data), ncol = ncol(data)+1)
Beta <- vector(length = ncol(data)-2)
```

```r
# Step 1 initialize X[,1] and Z[,1]
temp <- rep(1,nrow(data))
data <- data %>%
  add_column(temp, .before = T)
z[,1] = temp
```

```r
# Step 2 & 3
for (p in 2:ncol(data)){
  for (j in 2:p){
    for (l in 1:(j-1)){
      Gamma[l,j] = t(z[,l]) %*% data[,j] / t(z[,l] %*% z[,l])
    }
    z[,j] = data[,j]
    for (k in 1:(j-1)){
      z[,j] = z[,j] - Gamma[k,j] * z[,k]
    }
  }
  Beta[p-1] = t(z[,p]) %*% data[,'sales'] / (t(z[,p]) %*% z[,p])
}
# We can compare our results with those implemented by the built-in library of R
result <- lm(sales~. ,data = data)
result$coefficients
```

```
##  (Intercept)         temp           TV        radio    newspaper
##  2.938889369           NA  0.045764645  0.188530017 -0.001037493
```

```r
Beta[-4]
```

```
## [1]  0.047536640  0.187994227 -0.001037493
```

Table 1

	Dependent variable:
	sales
temp	
TV	0.047^{***}
	(0.001)
radio	0.189^{***}
	(0.009)
newspaper	-0.001
	(0.006)
Constant	2.939^{***}
	(0.312)
Observations	200
R^2	0.897
Adjusted R^2	0.896
Residual Std. Error	1.686 (df = 196)
F Statistic	570.271^{***} (df = 3; 196)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$