Homework 2 Solution

Bryan Liu

September 18, 2018

Problem 1

Let x_i^T be the ith row of X. Since $EY_i = x_i^T \beta$, we have

$$E\Big[\sum_{i=1}^{n} a_i Y_i\Big] = \sum_{i=1}^{n} a_i x_i^T \beta \tag{1}$$

$$= a^T X \beta \tag{2}$$

where $a = (a_1, ..., a_n)^T$. By assumption $a^T X \beta = \beta_1$. Let $\lambda^T = a^T X$, so $\lambda = X^T a \in C(X^T)$. Hence, β_1 is identifiable.

Problem 2

a)

The design matrix for this problem is the $n \times 2$ matrix of all 1s.

$$X = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ ... & ... \\ 1 & 1 \end{pmatrix} \tag{3}$$

b)

The normal equations are given by

$$X^T X \beta = X^T y \tag{4}$$

Writing this out explicitly with our defintion of X from part (a), we find that β_0 and β_1 must satisfy

$$n(\beta_0 + \beta_1) = \sum_{i=1}^{n} y_i \tag{5}$$

or more concisely, $\beta_0 + \beta_1 = \bar{y}$. We see that there are infinitely many possible β_0 and β_1 which solve this equation. Two for example, are $\beta_0 = 0, \beta_1 = \bar{y}$, or $\beta_0 = \bar{y}, \beta_1 = 0$

c)

Define $\gamma = \beta_1 + \beta_0$. We seek to find

$$\text{argmin}_{\gamma \in \mathbb{R}} \left\{ \sum_{i=1}^{n} (y_i - \gamma)^2 \right\} \tag{6}$$

We show that $\gamma = \bar{y}$ minimizes the objective above. We compute

$$\begin{aligned}
\sum_{i=1}^{n} (y_i - \gamma)^2 &= \sum_{i=1}^{n} (y_i - \gamma + \bar{y} - \bar{y})^2 && (7)\\
&= \sum_{i=1}^{n} (y_i - \bar{y})^2 + \sum_{i=1}^{n} (\bar{y} - \gamma)^2 + 2(\bar{y} - \gamma) \sum_{i=1}^{n} (\bar{y} - y_i) && (8)\\
&= \sum_{i=1}^{n} (y_i - \bar{y})^2 + \sum_{i=1}^{n} (\bar{y} - \gamma)^2 && (9)
\end{aligned}$$

where we used the fact that $\sum_{i=1}^{n} (\bar{y} - y_i)$. The first term does not depend on z, so it suffices ot minimize the second term. The second term can be set to zero and minimized, by setting $\gamma = \bar{y}$.

An easier way to show this is to use estimability: here $\beta_0 + \beta_1 = \lambda^T \beta$, where $\lambda^T = (1, 1)$. Pick any least square solution from part (b), e.g $\hat{\beta} = (\bar{y}, 0)^T$. Then $\lambda^T \hat{\beta} = \bar{y}$ is the least squares estimate for $\lambda^T \beta$.

d)

β_1 is not estimable. Let $\lambda^T = (0, 1)$, so $\beta_1 = \lambda^T \beta$. λ is not in the column-space of X^T.

e)

Here,

$$X = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ ... & ... \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \tag{10}$$

And $X^T X = \begin{pmatrix} n+1 & n+2 \\ n+2 & n+4 \end{pmatrix}$. We must solve the normal equations $X^T X \hat{\beta} = X^T y$, and in this case $X^T X$ is invertible.

We use formula for inverting a 2×2 matrix:

$$(X^T X)^{-1} = \frac{1}{(n+1)(n+4) - (n+2)^2} \begin{pmatrix} n+4 & -n-2 \\ -n-2 & n+1 \end{pmatrix} \tag{11}$$

We next compute $X^T y$ as

$$X^T y = \begin{pmatrix} \sum_{i=1}^n y_i + y_{n+1} \\ \sum_{i=1}^n y_i + 2y_{n+1} \end{pmatrix} \tag{12}$$

The regression coefficients are then $(X^T X)^{-1} X^T y$.

Problem 3

The design matrix for this problem is

$$X = \left(\mathbf{1_n}, \mathbf{I_{n\times n}}\right) \tag{13}$$

where $\mathbf{1_n}$ is a n dimensional column vector of ones, and $\mathbf{I_{n\times n}}$ is the $n \times n$ identity matrix. Hence, X is a matrix with dimensions $n \times (n+1)$.

We find a solution to the normal equations, $X^T X\beta = X^T y$, that we will use for the remainder of the problem. Firstly,

$$X^T X = \begin{pmatrix} n & \mathbf{1_n^T} \\ \mathbf{1_n} & \mathbf{I_{n\times n}} \end{pmatrix} \tag{14}$$

Secondly,

$$X^T y = \begin{pmatrix} \sum_{i=1}^n y_i \\ y_1 \\ y_2 \\ \dots \\ \dots \\ y_n \end{pmatrix} \tag{15}$$

By inspection, we see that $\beta_{ls} = (0, y_1, ..., y_n)^T$ is a solution to the normal equations.

a)

Let $\lambda = (1, 0, 1, 0, ..., 0)$, that is, the vector of length $n+1$, of all zeros except for ones in the first and third entry. Note that $\lambda^T \beta = \beta_0 + \beta_2$. We check that $\lambda \in \mathcal{C}(X^T)$. Let $a = (1, -1, 1, -1, -1, ...,)$, the vector of all negative ones, except for a positive one in the first and third entries. We note that $X^T a = \lambda$, so $\beta_0 + \beta_2$ is estimable.

The least squares estimate for $\beta_0 + \beta_2$ is then $\lambda^T \beta_{ls} = y_2$.

b)

Let $\lambda = (0, 1, 0, 0..., 0)^T$ so that $\beta_1 = \lambda^T \beta$. We show that λ is not in $\mathcal{C}(X^T)$. Let x_i be the be the ith column of X. Then by inspection, if there exist $a_0, ..., a_n$ such that $\lambda = \sum_{i=0}^n a_i x_i$, then $a_1, ..., a_n$ must be zero. However, λ is not in the span of x_0, so no such a_0 exists.

c)

Let $\lambda = (0, 1, -1, 0..., 0)^T$ so that $\beta_1 - \beta_2 = \lambda^T\beta$. Then let $a = (1, -1, 0, 0, ..., 0)^T$, so $X^Ta = \lambda$. Hence $\beta_1 - \beta_2$ is estimable. The least squares estimate for $\beta_1 - \beta_2$ is then $\lambda^T\beta_{ls} = y_1 - y_2$.

d)

Let $\lambda = (0, 1, 1, 1, -3, 0, 0, ..., 0)^T$, so $\beta_1 + \beta_2 + \beta_3 - 3\beta_4 = \lambda^T\beta$.

Note that $X^TX\lambda = \lambda$, so $\lambda \in \mathcal{C}(X^TX)$. Hence, $\lambda^T\beta$ is estimable.

Problem 4

Our model is

$$
\begin{aligned}
BODYFAT &= \beta_0 + \beta_1 AGE + \beta_2 WEIGHT + \beta_3 HEIGHT + \beta_4(AGE + 10WEIGHT + 3HEIGHT) + e && (16)\\
&= \beta_0 + (\beta_1 + \beta_4)AGE + (\beta_2 + 10\beta_4)WEIGHT + (\beta_3 + 3\beta_4)HEIGHT + e && (17)\\
&= \alpha_0 + \alpha_1 AGE + \alpha_2 WEIGHT + \alpha_3 HEIGHT + e && (18)
\end{aligned}
$$

where

$$
\begin{aligned}
\alpha_0 &= \beta_0 && (19)\\
\alpha_1 &= \beta_1 + \beta_4 && (20)\\
\alpha_2 &= \beta_2 + 10\beta_4 && (21)\\
\alpha_3 &= \beta_3 + 3\beta_4 && (22)
\end{aligned}
$$

We can solve for the least squares estimates of α in R:

```r
bodyfat_ds <- read.csv('./Bodyfat.csv')
# head(bodyfat_ds)

bodyfat_lm <- lm(bodyfat ~ Age + Weight + Height, data = bodyfat_ds)
bodyfat_lm
```

```
##
## Call:
## lm(formula = bodyfat ~ Age + Weight + Height, data = bodyfat_ds)
##
## Coefficients:
## (Intercept)          Age       Weight       Height
##     17.7674       0.1698       0.1982      -0.5943
```

```r
alpha <- bodyfat_lm$coefficients # regression coefficients in estimable model
```

Hence we find α_0 = 17.77, α_1 = 0.17, α_2 = 0.2, and α_3 = -0.59

a)

Any choice of $\beta_0, ..., \beta_4$ that satisfies equations 19 - 22 is a least squares solution. We use R to print three different solutions:

```
# note that we have one free parameter in the betas
# let us choose three different \beta_4s.
# the remaining betas will be functions of \beta_4 and \alpha

get_beta_estimates <- function(alpha, beta_4){
  alpha <- as.vector(alpha)
  multiplier <- c(0, -1, -10, -3)
  return(as.vector(c(alpha + beta_4 * multiplier, beta_4)))
}

# These are three different estimates for beta
beta_4 <- 0
get_beta_estimates(alpha, beta_4)
```

```
## [1] 17.7673848  0.1697902  0.1981519 -0.5943339  0.0000000
```

```
beta_4 <- 1
get_beta_estimates(alpha, beta_4)
```

```
## [1] 17.7673848 -0.8302098 -9.8018481 -3.5943339  1.0000000
```

```
beta_4 <- 2
get_beta_estimates(alpha, beta_4)
```

```
## [1]  17.767385  -1.830210 -19.801848  -6.594334   2.000000
```

```
# we assert that they actually solve the least squares problem

# This the matrix with four columns:
# age, weight, height, and age + 10weight + 3height
X <- matrix(1, ncol = 5, nrow = dim(bodyfat_ds)[1])
X[, 2] <- bodyfat_ds[, 'Age']
X[, 3] <- bodyfat_ds[, 'Weight']
X[, 4] <- bodyfat_ds[, 'Height']

X[, 5] <- X[, 2] + 10 * X[, 3] + 3 * X[, 4]

X <- as.matrix(X)

# X^T X
XtX <- t(X) %*% X

all.equal(XtX %*% as.matrix(get_beta_estimates(alpha, 0)),
          t(X) %*% bodyfat_ds[, 'bodyfat'])
```

```
## [1] TRUE
```

```
all.equal(XtX %*% as.matrix(get_beta_estimates(alpha, 1)),
          t(X) %*% bodyfat_ds[, 'bodyfat'])
```

```
## [1] TRUE
```

```r
all.equal(XtX %*% as.matrix(get_beta_estimates(alpha, 2)),
          t(X) %*% bodyfat_ds[, 'bodyfat'])
```

```
## [1] TRUE
```

b)

β_1 is not estimable because we showed in part (a) that there is not a unique solution for β_1.

c)

The least squares estimates are what we called α at the beginning, see 19 - **??**. Reprinted here, they are

```r
alpha
```

```
## (Intercept)         Age      Weight      Height
##  17.7673848   0.1697902   0.1981519  -0.5943339
```

d)

```r
head(bodyfat_ds)
```

```
##   Density bodyfat Age Weight Height Neck Chest Abdomen   Hip Thigh Knee
## 1  1.0708    12.3  23 154.25  67.75 36.2  93.1    85.2  94.5  59.0 37.3
## 2  1.0853     6.1  22 173.25  72.25 38.5  93.6    83.0  98.7  58.7 37.3
## 3  1.0414    25.3  22 154.00  66.25 34.0  95.8    87.9  99.2  59.6 38.9
## 4  1.0751    10.4  26 184.75  72.25 37.4 101.8    86.4 101.2  60.1 37.3
## 5  1.0340    28.7  24 184.25  71.25 34.4  97.3   100.0 101.9  63.2 42.2
## 6  1.0502    20.9  24 210.25  74.75 39.0 104.5    94.4 107.8  66.0 42.0
##   Ankle Biceps Forearm Wrist
## 1  21.9   32.0    27.4  17.1
## 2  23.4   30.5    28.9  18.2
## 3  24.0   28.8    25.2  16.6
## 4  22.8   32.4    29.4  18.2
## 5  24.0   32.2    27.7  17.7
## 6  25.6   35.7    30.6  18.8
```

```r
summary(lm(bodyfat ~ Age + Weight + Height + I(Age + 10*Weight +3*Height), data = bodyfat_ds))
```

```
##
## Call:
## lm(formula = bodyfat ~ Age + Weight + Height + I(Age + 10 * Weight +
##     3 * Height), data = bodyfat_ds)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.3960  -4.5038  -0.0326   3.8324  15.7154
##
## Coefficients: (1 not defined because of singularities)
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         17.76738    7.47935   2.376   0.0183 *
## Age                                  0.16979    0.02956   5.744 2.70e-08 ***
```

```
## Weight                               0.19815    0.01313  15.095  < 2e-16 ***
## Height                              -0.59433    0.10690  -5.560 6.97e-08 ***
## I(Age + 10 * Weight + 3 * Height)        NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.809 on 248 degrees of freedom
## Multiple R-squared:  0.524,  Adjusted R-squared:  0.5182
## F-statistic: 90.99 on 3 and 248 DF,  p-value: < 2.2e-16
```

Running the lm function above, we see that they computed what we defined α – the least squares estimate when we ignore the degenerate column.

Problem 5

a)

We compute X^TX to be equal to $8 \cdot I_{4\times 4}$. To find the least squares estimates for $\beta_1, \beta_2, \beta_3$, and β_4, we solve $X^TX\beta = X^Ty$, where $y \in \mathbb{R}^d$ are the 8 observations in our weighing.

Hence, $\beta = (X^TX)^{-1}X^Ty = \frac{1}{8}X^Ty$.

Writing it out, we have

$$\beta_1 = \frac{1}{8}\sum_{i=1}^{8} y_i \tag{23}$$

$$\beta_2 = \frac{1}{8}\sum_{i=1}^{8} y_i(-1)^i \tag{24}$$

$$\beta_3 = \frac{1}{8}\sum_{i=1,2,5,6} y_i - \frac{1}{8}\sum_{i=3,4,7,8} y_i \tag{25}$$

$$\beta_4 = \frac{1}{8}\sum_{i=1,4,5,8} y_i - \frac{1}{8}\sum_{i=2,3,6,7} y_i \tag{26}$$

```r
# The X matrix
col1 <- rep(1, 8)
col2 <- rep(-1, 8) ** seq(0, 7)

X <- matrix(c(col1, col2), nrow = 8)
X
```

```
##      [,1] [,2]
## [1,]    1    1
## [2,]    1   -1
## [3,]    1    1
## [4,]    1   -1
## [5,]    1    1
## [6,]    1   -1
## [7,]    1    1
## [8,]    1   -1
```

```r
# col3 <- c(1, 1, -1, -1, 1, 1, -1, -1)
# col4 <- col1 * col2 * col3
# X <- matrix(c(col1, col2, col3, col4), nrow = 8)

# XTX
XtX <- t(X) %*% X
print(XtX)
```

```
##      [,1] [,2]
## [1,]    8    0
## [2,]    0    8
```

b)

The covariance of $\hat{\beta}$ is given by $\sigma^2(X^T X)^{-1} = (\sigma^2/8)I_{4\times 4}$. Hence, the estimates for $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$, and $\hat{\beta}_4$ are independent each with variance $\sigma^2/8$.