Regression Analysis HW7

Tuorui Peng, tuoruipeng2028@u.northwestern.edu

(a)

(i)

Use induction: if for $k \geq 0$ we have $(A-B)\sum_{i=0}^{k} A^{-1}(BA^{-1})^i = I - (BA^{-1})^{k+1}$, then at $k+1$:

$$\begin{aligned}(A-B)\sum_{i=0}^{k+1} A^{-1}(BA^{-1})^i =&(A-B)\sum_{0=1}^{k} A^{-1}(BA^{-1})^i + (A-B)A^{-1}(BA^{-1})^{k+1}\\ =&I-(BA^{-1})^{k+1}+(BA^{-1})^{k+1}-(BA^{-1})^{k+2}\\ =&I-(BA^{-1})^{k+2}\end{aligned}$$

And notice that when $k=0$ we have

$$(A-B)\sum_{i=0}^{k=0} A^{-1}(BA^{-1})^i = (A-B)A^{-1} = I - BA^{-1}$$

So by induction, the statement is true for all $k \geq 1$.

(ii)

We have

$$\left\lvert\left\lvert\left\lvert BA^{-1}\right\rvert\right\rvert\right\rvert_{\mathrm{op}} \leq \left\lvert\left\lvert\left\lvert B\right\rvert\right\rvert\right\rvert_{\mathrm{op}} \left\lvert\left\lvert\left\lvert A^{-1}\right\rvert\right\rvert\right\rvert_{\mathrm{op}} < 1 \Rightarrow \left\lvert\left\lvert\left\lvert (BA^{-1})^{k+1}\right\rvert\right\rvert\right\rvert_{\mathrm{op}} \leq \left(\left\lvert\left\lvert\left\lvert BA^{-1}\right\rvert\right\rvert\right\rvert_{\mathrm{op}}\right)^{k+1} \to 0$$

so it's safe to say that $(BA^{-1})^{k+1} \to 0$, thus set $k \to \infty$ in the result of **(i)** we have

$$\begin{cases}(A-B)\sum_{i=0}^{\infty} A^{-1}(BA^{-1})^i = I & \Rightarrow (A-B)^{-1} = \sum_{i=0}^{\infty} A^{-1}(BA^{-1})^i\\ (A+B)\sum_{i=0}^{\infty} A^{-1}(-BA^{-1})^i = I & \Rightarrow (A+B)^{-1} = \sum_{i=0}^{\infty}(-1)^i A^{-1}(BA^{-1})^i\end{cases}$$

(b)

With notation $D(\boldsymbol{\delta}) = \mathrm{diag}(\boldsymbol{\delta})$ we have

$$\hat{\beta}_{\boldsymbol{\delta}} = \arg\min_b \frac{1}{n}(Y-Xb)'(I-D(\boldsymbol{\delta}))(Y-Xb)$$

Solution is given at $\dfrac{\partial}{\partial b} = 0$, which is

$$\begin{aligned} 0 =& \frac{\partial}{\partial b}\frac{1}{n}(Y - Xb)'(I - D(\boldsymbol{\delta}))(Y - Xb) \\ =& -\frac{1}{n}X'(I - D(\boldsymbol{\delta}))Y + \frac{1}{n}X'(I - D(\boldsymbol{\delta}))Xb \\ \Rightarrow \hat{\beta}_{\boldsymbol{\delta}} =& (X'(I - D(\boldsymbol{\delta}))X)^{-1}X'(I - D(\boldsymbol{\delta}))Y \end{aligned}$$

(c)

Using the result from **(a)-(ii)** we have

$$\begin{aligned} (X'(I - D(\boldsymbol{\delta}))X)^{-1} =& (X'X - X'D(\boldsymbol{\delta})X)^{-1} \\ =& \sum_{i=0}^{\infty}(X'X)^{-1}\left(X'D(\boldsymbol{\delta})X(X'X)^{-1}\right)^{i} \\ =& (X'X)^{-1}(I + X'D(\boldsymbol{\delta})X(X'X)^{-1} + O(\|\boldsymbol{\delta}\|^2)) \end{aligned}$$

i.e.

$$\begin{aligned} \hat{\beta}_{\boldsymbol{\delta}} =& (X'(I - D(\boldsymbol{\delta}))X)^{-1}X'(I - D(\boldsymbol{\delta}))Y \\ =& (X'X)^{-1}(I + X'D(\boldsymbol{\delta})X(X'X)^{-1} + O(\|\boldsymbol{\delta}\|^2))X'(I - D(\boldsymbol{\delta}))Y \\ =& (X'X)^{-1}X'Y + ((X'X)^{-1}X'D(\boldsymbol{\delta})X(X'X)^{-1}X'Y - (X'X)^{-1}X'D(\boldsymbol{\delta})Y) + O(\|\boldsymbol{\delta}\|^2) \\ =& \hat{\beta} + \frac{1}{n}\hat{C}^{-1}X'D(\boldsymbol{\delta})(\hat{y} - y) + O(\|\boldsymbol{\delta}\|^2) \end{aligned}$$

in which $\hat{C} = \frac{1}{n}X'X$.

(d)

If $\boldsymbol{\delta} := \delta e_i$, we have $D(\boldsymbol{\delta})_{kl} = \delta_{ki}\delta_{il}$ in which $\delta_{\cdot\,\cdot}$ is the Kronecker delta. So

$$\begin{aligned} \hat{\beta}_{\boldsymbol{\delta}=\delta e_i} =& + \frac{1}{n}\hat{C}^{-1}X'\delta_{\cdot i}\delta_{i\cdot}(\hat{y} - y) + O(\|\boldsymbol{\delta}\|^2) \\ \Rightarrow \lim_{\delta\to 0}\frac{\hat{\beta}_{\delta e_i} - \hat{\beta}_{\boldsymbol{\delta}}}{\delta} =& \lim_{\delta\to 0}\frac{1}{n}\hat{C}^{-1}X'\delta_{\cdot i}\delta_{i\cdot}(\hat{y} - y) + O(\|\boldsymbol{\delta}\|) \\ =& \frac{1}{n}\hat{C}^{-1}x_i(\hat{y}_i - y_i) := \mathsf{iinf}(\hat{\beta}, i) \end{aligned}$$

(e)

From the plot we can see that removing a few points seems do not affect the p-value much, for most variables. Only on `Height` variable, removing one point can change the p-value a lot.

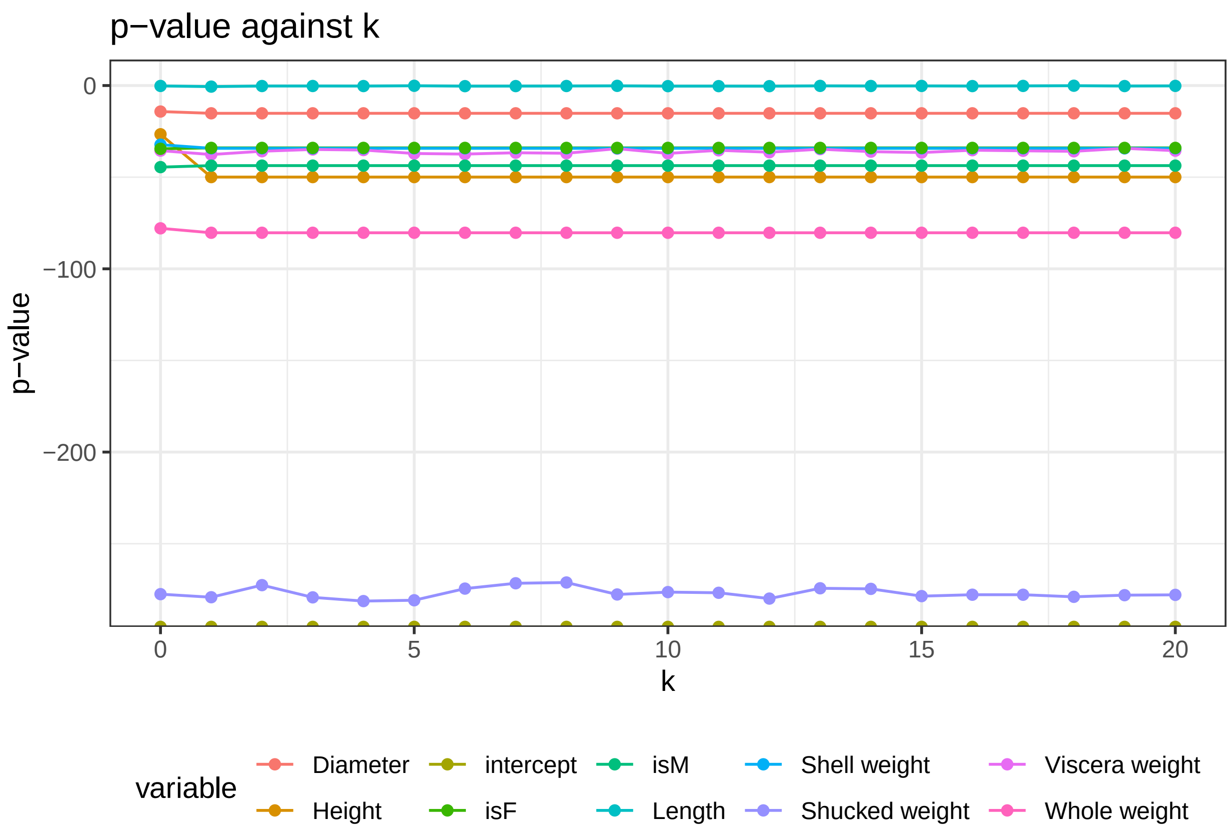
```r
library('tidyverse')
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
# i) data read in and preprocessing
aba <- read.csv('abalone.data', header = FALSE)
df <- data.frame(aba[,2:8])
df$isM <- ifelse(aba[,1] == 'M', 1, 0)
df$isF <- ifelse(aba[,1] == 'F', 1, 0)

# i) normalization to l_2 norm=\sqrt{n} and mean centering, then add intercept and y
df <- df %>% mutate(across(everything(), scale))
df$intercept <- 1
df$y <- aba[,9]

# ii) compute iinf: n\times d matrix
n <- nrow(df)
d <- ncol(df) - 1
X <- df[,1:d] %>% as.matrix()
C.hat <- t(X) %*% X / n
beta.hat <- solve(C.hat, t(X) %*% df$y / n)
y.hat <- X %*% beta.hat
iinf.mat <-  diag(c(y.hat)) %*% X %*% solve(C.hat) / n

# iii) find index set I with |I|=k
find_maximize_idxs <- function(beta, iinf, k){
    order.iinf <- order(iinf, decreasing = TRUE)
    # for simplicity, we just try two ways: all negative and all positive
    upper.idx <- order.iinf[1:k]
    lower.idx <- order.iinf[(n-k+1):n]
    upper.sum <- beta + sum(iinf[upper.idx])
    lower.sum <- beta + sum(iinf[lower.idx])
    ret.idx <- ifelse( abs(upper.sum) > abs(lower.sum), upper.idx, lower.idx)
    return(ret.idx)
}
ks <- 1:20
p.value.mat <- matrix(NA, nrow = d, ncol = length(ks))
```

```
for(j in 1:d){
    iinf.j <- iinf.mat[,j]
    beta.hat.j <- beta.hat[j]

    for(k in ks){
        maximize.idxs <- find_maximize_idxs(beta.hat.j, iinf.j, k)
        # run regression without these indices, and test null: beta_j=0
        X.jk <- X[-maximize.idxs,]
        y.jk <- df$y[-maximize.idxs]
        lm.jk <- lm(y.jk ~ X.jk - 1)
        p.value.jk <- summary(lm.jk)$coefficients[j,4]
        p.value.mat[j,k] <- p.value.jk
    }
}
# k=0 case is the original model
lm.0 <- lm(df$y ~ X - 1)
p.value.mat <- cbind(summary(lm.0)$coefficients[,4], p.value.mat)

# plot p-value against k
p.value.df <- data.frame(t(p.value.mat))
names(p.value.df) <- c('Length', 'Diameter', 'Height', 'Whole weight', 'Shucked weight', 'Viscera wei
p.value.df <- log(p.value.df)
p.value.df$k <- 0:20

p.value.df %>%
    pivot_longer(cols = -k, names_to = 'variable', values_to = 'p.value') %>%
    ggplot(aes(x = k, y = p.value, color = variable)) +
    geom_line() +
    geom_point() +
    theme_bw() +
    theme(legend.position = 'bottom') +
    labs(x = 'k', y = 'p-value', title = 'p-value against k')
```


p−value against k
0
−100
−200
p−value
0
5
10
15
20
k
variable
Diameter
Height
intercept
isF
isM
Length
Shell weight
Shucked weight
Viscera weight
Whole weight