HW4

Jinhong Du - 12243476

2019/10/08

Contents

1. (Faraway 5.4) For the longley data, fit a model with Employed as the response and the other variables as predictors.

(a) Compute and comment on the condition numbers.

There is a wide range in the eigenvalues and several condition numbers are large. This means that the matrix $X^\top X$ is very singular, and problems are being caused by more than just one linear combination (collinearity).

```
library(faraway)
data(longley)

model <- lm(Employed~., longley)
X <- model.matrix(model)[,-1]
eig <- eigen(t(X)%*%X)$values
cat(eig)
```

```
## 66652993 209073 105355 18039.76 24.5573 2.015117
```

```
cat(sqrt(eig[1]/eig))
```

```
## 1 17.85504 25.15256 60.78472 1647.478 5751.216
```

(b) Compute and comment on the correlations between the predictors.

As we can see, `GNP.deflator`, `GNP`, `Population` and `Year` are highly linearly correlated to each other.

```
round(cor(X),2)
```

```
##              GNP.deflator  GNP Unemployed Armed.Forces Population Year
## GNP.deflator         1.00 0.99       0.62         0.46       0.98 0.99
## GNP                  0.99 1.00       0.60         0.45       0.99 1.00
## Unemployed           0.62 0.60       1.00        -0.18       0.69 0.67
## Armed.Forces         0.46 0.45      -0.18         1.00       0.36 0.42
## Population           0.98 0.99       0.69         0.36       1.00 0.99
## Year                 0.99 1.00       0.67         0.42       0.99 1.00
```

(c) Compute the variance inflation factors.

The variance inflation factor(VIF) is given by $(1-R_j^2)^{-1}$ where $R_j^2 = Cor^2(\hat{X}_j, X_j)$ is the coefficient of determination by regressing of X_j on all other predictors. If R_j^2 is close to one, then the variance inflation factor will be and so $Var\hat{\beta}_j = \sigma^2 \frac{1}{1-R_j^2} \frac{1}{\sum_i (X_{ij}-\overline{X}_j)^2}$ will also be large.

There is much variance inflation for covariates `GNP.deflator`, `GNP`, `Population` and `Year`, indicating effect of collinearity.

```
vif(X)
```

```
## GNP.deflator          GNP   Unemployed Armed.Forces   Population 
##    135.53244   1788.51348     33.61889      3.58893    399.15102 
##         Year 
##    758.98060
```

2. (Faraway 5.5) For the prostate data, fit a model with lpsa as the response and the other variables as predictors.

(a) Compute and comment on the condition numbers.

There is a wide range in the eigenvalues and almost all condition numbers are large. This may be the reason that columns of X have highly varying norms instead of collinearity.

```
library(faraway)
data(prostate)

model <- lm(lpsa~., prostate)
X <- model.matrix(model)[,-1]
eig <- eigen(t(X)%*%X)$values
cat(eig)
```

```
## 479082.6 61907.04 210.9042 175.6329 64.79853 44.52379 20.23914 8.093145
```

```
cat(sqrt(eig[1]/eig))
```

```
## 1 2.78186 47.66094 52.22787 85.98499 103.7311 153.8541 243.3025
```

(b) Compute and comment on the correlations between the predictors.

As we can see, there is no correlation very close to 1/-1, so there is no large pairwise collinearities.

```
round(cor(X),2)
```

```
##         lcavol lweight  age  lbph   svi   lcp gleason pgg45
## lcavol    1.00    0.19 0.22  0.03  0.54  0.68    0.43  0.43
## lweight   0.19    1.00 0.31  0.43  0.11  0.10    0.00  0.05
## age       0.22    0.31 1.00  0.35  0.12  0.13    0.27  0.28
## lbph      0.03    0.43 0.35  1.00 -0.09 -0.01    0.08  0.08
## svi       0.54    0.11 0.12 -0.09  1.00  0.67    0.32  0.46
## lcp       0.68    0.10 0.13 -0.01  0.67  1.00    0.51  0.63
## gleason   0.43    0.00 0.27  0.08  0.32  0.51    1.00  0.75
## pgg45     0.43    0.05 0.28  0.08  0.46  0.63    0.75  1.00
```

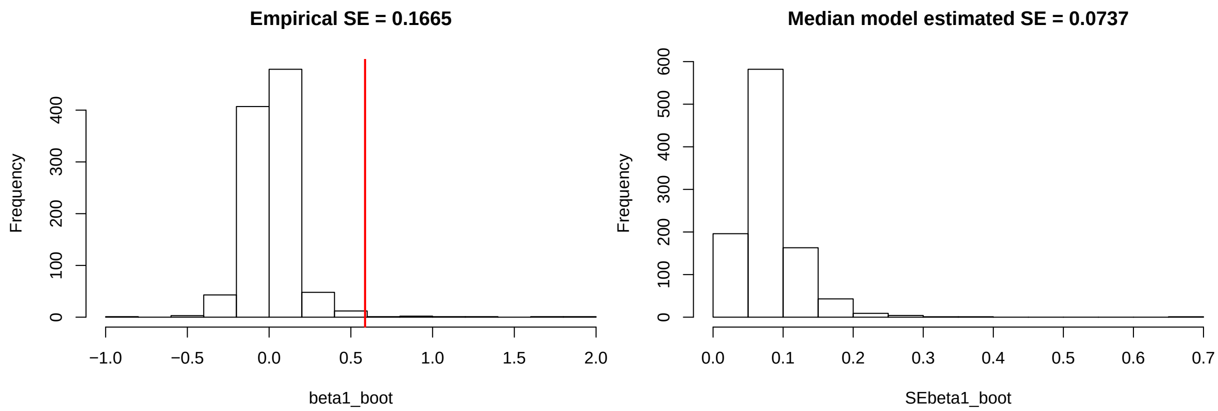
(c) Compute the variance inflation factors.

Since all variance inflation factors are near to 1, there may be only slight effect of collinearity.

```
vif(X)
```

```
##   lcavol  lweight      age     lbph      svi      lcp  gleason    pgg45
## 2.054115 1.363704 1.323599 1.375534 1.956881 3.097954 2.473411 2.974361
```

3. Using the same covariates and response as in Faraway 5.4, use bootstrapping to check whether the usual estimate of the standard error for $\hat{\beta}_{\text{GNP.deflator}}$ seems to estimate the variability appropriately. (Bootstrap the sample, not the residual.) Discuss what you see and any possible explanations.



As we can see from the above plots, the empirical standard error for $\hat{\beta}_{\text{GNP.deflator}}$ is much larger than the theoretical estimated SE. It may be caused by the small size of data set. Every time we bootstrap the samples, usually some of the sampels will not be selected, which is of high ratio in this same data set. Leverage points, outliers and influential points may appear more times and so that may cause much biased in bootstrapped models.

```r
set.seed(0)
nboot = 1000
n <- dim(longley)[1]
beta1_boot = SEbeta1_boot = c(0,nboot)
for(i in 1:nboot){
    boot_sample = sample(n,n,replace=TRUE)
    d <- longley[boot_sample,]
    model_boot = lm(Employed~., d)
    beta1_boot[i] = model_boot$coefficients[2]
    SEbeta1_boot[i] = summary(model_boot)$coefficients[2,2]
}

par(mfrow=c(1,2))
hist(beta1_boot,main=paste0('Empirical SE = ',round(sqrt(var(beta1_boot)),4)))

abline(v = model$coefficients[2],col='red',lwd=2)
hist(SEbeta1_boot,main=paste0('Median model estimated SE = ',round(median(SEbeta1_boot),4)))
```

4. In this problem we'll prove that, no matter the correlation structure of the covariates, variance of estimating a mean at a new $x \in \mathbb{R}^p$ can only increase when you add an additional covariate. You are welcome to collaborate in pairs or groups of three on this problem; if you choose to work in a group, please list your collaborators in your handed in HW.

We will consider two models: with and without X_j. Let $X \in \mathbb{R}^{n\times p}$ be the full matrix of covariates and X_{-j} be the same matrix with the X_j column removed. We will assume that the normal linear model holds in both cases, i.e. the true model for the response is

$$Y_i = \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + N(0, \sigma^2)$$

and we have $\beta_j = 0$ so that this model is true even with X_j removed. We'll write $\hat{\beta}$ for the fitted coefficients using all the covariates, and $\hat{\beta}_{-j}$ for the model using the $p-1$ covariates when X_j is removed. Note that $\hat{\beta}_{-j}$ is not the same as removing the entry j from the vector $\hat{\beta}$—the values may have changed entirely.

If we predict the mean response at a new $x \in \mathbb{R}^p$, we would predict

$$\hat{y} = x_1\hat{\beta}_1 + \cdots + x_p\hat{\beta}_p = x^\top \hat{\beta}.$$

For the reduced model, we would predict

$$\hat{y}_{-j} = x_1(\hat{\beta}_{-j})_1 + \cdots + x_{j-1}(\hat{\beta}_{-j})_{j-1} + x_{j+1}(\hat{\beta}_{-j})_{j+1} + \cdots x_p(\hat{\beta}_{-j})_p = x_{-j}^\top \hat{\beta}_{-j}$$

where x_{-j} is the vector x with entry j removed.

We will use a linear algebra result:

Lemma: For any positive definite matrix $\begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}$, it holds that

$$\begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}^{-1} \succeq \begin{bmatrix} A^{-1} & 0 \\ 0 & 0 \end{bmatrix}.$$

Here $M \succeq N$ is the positive semidefinite ordering, defined on matrices M, N which are themselves positive semidefinite, with $M \succeq N$ equivalent to $M - N \succeq 0$, i.e. $M - N$ is positive semidefinite.

You can use the fact that, for positive semidefinite and invertible M, N, it holds that $M \succeq N$ if and only if $M^{-1} \preceq N^{-1}$.

(a) Write down the variance of $\hat{y}$ and of $\hat{y}_{-j}$, using matrix notation such as $X^\top X$ for short and clean answers.

Since $\hat{\beta} = (X^\top X)^{-1} X^\top Y \sim N(\beta, \sigma^2 (X^\top X)^{-1})$, we have that

$$\begin{aligned} \mathrm{Var}(\hat{y}) &= \mathrm{Var}(x^\top \hat{\beta}) \\ &= x^\top \mathrm{Var}(\hat{\beta}) x \\ &= \sigma^2 x^\top (X^\top X)^{-1} x \end{aligned}$$

Since $\hat{\beta}_{-j} = (X_{-j}^\top X_{-j})^{-1} X_{-j}^\top Y \sim N(\beta, \sigma^2 (X_{-j}^\top X_{-j})^{-1})$, we have that

$$\begin{aligned} \mathrm{Var}(\hat{y}_{-j}) &= \mathrm{Var}(x^\top \hat{\beta}_{-j}) \\ &= x_{-j}^\top \mathrm{Var}(\hat{\beta}_{-j}) x_{-j} \\ &= \sigma^2 x_{-j}^\top (X_{-j}^\top X_{-j})^{-1} x_{-j} \end{aligned}$$

(b) Consider predicting the mean response value at a new $x \in \mathbb{R}^p$. Assuming the lemma is true, prove that $\mathrm{Var}(\hat{y}) \geq \mathrm{Var}(\hat{y}_{-j})$.

Without loss of generality, assume $j = p$, Otherwise, we can rearange the covariates to achieve this.
From the lemma, since $X^\top X$ is positive (by assumption it's invertible and so has no zero eigenvalue) definite ,

$$(X^\top X)^{-1} = \begin{bmatrix} X_{-j}^\top X_{-j} & X_{-j}^\top X_j \\ X_j^\top X_{-j} & X_j^\top X_j \end{bmatrix}^{-1} \succeq \begin{bmatrix} (X_{-j}^\top X_{-j})^{-1} & \mathbf{0}_{(p-1)\times 1} \\ \mathbf{0}_{1\times(p-1)} & 0 \end{bmatrix}$$

Then $(X^\top X)^{-1} - \begin{bmatrix} (X_{-j}^\top X_{-j})^{-1} & 0 \\ 0 & 0 \end{bmatrix}$ is postive semidefinite and

$$x^\top (X^\top X)^{-1} x - x_{-j}^\top (X_{-j}^\top X_{-j})^{-1} x_{-j} = x^\top \left[(X^\top X)^{-1} - \begin{bmatrix} (X_{-j}^\top X_{-j})^{-1} & \mathbf{0}_{(p-1)\times 1} \\ \mathbf{0}_{1\times(p-1)} & 0 \end{bmatrix} \right] x \geq 0$$

Therefore, $\mathrm{Var}(\hat{y}) \geq \mathrm{Var}(\hat{y}_{-j})$.

(c) Prove that if X_j is orthogonal to X_k for every $k \neq j$, and $x_j = 0$, then the variances are in fact equal.

If X_j is orthogonal to X_k $(k \neq j)$, then $X_{-j}^\top X_j = \mathbf{0}_{(p-1)\times 1}$. So

$$(X^\top X)^{-1} = \begin{bmatrix} X_{-j}^\top X_{-j} & \mathbf{0}_{(p-1)\times 1} \\ \mathbf{0}_{1\times(p-1)} & X_j^\top X_j \end{bmatrix}^{-1} = \begin{bmatrix} (X_{-j}^\top X_{-j})^{-1} & \mathbf{0}_{(p-1)\times 1} \\ \mathbf{0}_{1\times(p-1)} & (X_j^\top X_j)^{-1} \end{bmatrix}.$$

Since $x_j = 0$, we have

$$x^\top (X^\top X)^{-1} x = x_{-j}^\top (X_{-j}^\top X_{-j})^{-1} x_{-j}$$

and so $\mathrm{Var}(\hat{y}) = \mathrm{Var}(\hat{y}_{-j})$.

(d) Now we'll prove the lemma. One simple way to do this is with a limiting argument—we'll prove that

$$\begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}^{-1} \succeq \begin{bmatrix} \frac{1}{1+\epsilon} A^{-1} & 0 \\ 0 & 0 \end{bmatrix} \tag{$*$}$$

If this is true for any $\epsilon > 0$ then taking a limit, the lemma will be true. We'll break the proof into steps:

i. Prove that if $\begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}$ is positive semidefinite, then so is $\begin{bmatrix} \epsilon A & -B \\ -B^\top & \epsilon^{-1} C \end{bmatrix}$.

Suppose that $A \in \mathbb{R}^{m\times m}, B \in \mathbb{R}^{m\times n}$ and $C \in \mathbb{R}^{n\times n}$. $\forall\ \epsilon > 0$, $\forall\ x = \begin{bmatrix} y \\ z \end{bmatrix} \in \mathbb{R}^{m+n}$ where $y \in \mathbb{R}^m$ and $z \in \mathbb{R}^n$,

$$\begin{aligned} x^\top \begin{bmatrix} \epsilon A & -B \\ -B^\top & \epsilon^{-1} C \end{bmatrix} x &= \epsilon y^\top A y - y^\top B z - z^\top B^\top y + \epsilon^{-1} z^\top C z \\ &= \begin{bmatrix} \epsilon^{\frac{1}{2}} y \\ \epsilon^{-\frac{1}{2}} z \end{bmatrix}^\top \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} \begin{bmatrix} \epsilon^{\frac{1}{2}} y \\ \epsilon^{-\frac{1}{2}} z \end{bmatrix} \geq 0 \end{aligned}$$

since $\begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}$ is positive semidefinite. So $\begin{bmatrix} \epsilon A & -B \\ -B^\top & \epsilon^{-1} C \end{bmatrix}$ is also positive semidefinite.

ii. Using the previous step, prove that $\begin{bmatrix} \epsilon A & -B \\ -B^\top & c\mathbf{I} - C \end{bmatrix} \succeq 0$ for a sufficiently large constant c (you should specify c in terms of the other quantities in the problem).

If $\begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}$ is positive semidefinite, $\forall\ x = \begin{bmatrix} 0_{m\times 1} \\ z \end{bmatrix} \in \mathbb{R}^{m+n}$ where $z \in \mathbb{R}^n$,

$$x^\top \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} x = z^\top C z \geq 0,$$

i.e., C is postive semidefinite. Suppose that C has eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$. For $c \geq (1+\epsilon^{-1})\|C\|_2 = (1+\epsilon^{-1})\lambda_1$, $\lambda(c\mathbf{I} - (1+\epsilon^{-1})C) = c - (1+\epsilon^{-1})\lambda(C) \geq 0$, i.e., the eigenvalues of $c\mathbf{I} - (1+\epsilon^{-1})C$ are all non-negative. So $c\mathbf{I} - (1+\epsilon^{-1})C$ is positive semidefinite.

So $\forall\ x = \begin{bmatrix} y \\ z \end{bmatrix} \in \mathbb{R}^{m+n}$ where $y \in \mathbb{R}^m$ and $z \in \mathbb{R}^n$,

$$\begin{aligned} z^\top (c\mathbf{I} - C) z - \epsilon^{-1} z^\top C z &= c z^\top z - z^\top C z - \epsilon^{-1} z^\top C z \\ &= z^\top (c\mathbf{I} - (1+\epsilon^{-1})C) z \geq 0, \end{aligned}$$

and

$$\begin{aligned} x^\top \begin{bmatrix} \epsilon A & -B \\ -B^\top & c\mathbf{I} - C \end{bmatrix} x &= \epsilon y^\top A y - y^\top B z - z^\top B^\top y + \epsilon^{-1} z^\top (c\mathbf{I} - C) z \\ &\geq \epsilon y^\top A y - y^\top B z - z^\top B^\top y + \epsilon^{-1} z^\top C z \\ &= x^\top \begin{bmatrix} \epsilon A & -B \\ -B^\top & \epsilon^{-1} C \end{bmatrix} x \geq 0, \end{aligned}$$

i.e., $\begin{bmatrix} \epsilon A & -B \\ -B^\top & c\mathbf{I} - C \end{bmatrix}$ is positive semidefinite when $c \geq (1+\epsilon^{-1})\|C\|_2$.

iii. Using the previous step, prove that this implies $\begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}^{-1} \succeq \begin{bmatrix} \frac{1}{1+\epsilon} A^{-1} & 0 \\ 0 & c^{-1}\mathbf{I} \end{bmatrix}$ and that implies the equation marked with a $(*)$ above.

Since

$$\begin{bmatrix} \epsilon A & -B \\ -B^\top & c\mathbf{I} - C \end{bmatrix} = \begin{bmatrix} (1+\epsilon) A & 0 \\ 0 & c\mathbf{I} \end{bmatrix} - \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} \succeq 0.$$

So

$$\begin{bmatrix} (1+\epsilon) A & 0 \\ 0 & c\mathbf{I} \end{bmatrix} \succeq \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}.$$

Taking inverse of both sides, we have $\begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}^{-1} \succeq \begin{bmatrix} \frac{1}{1+\epsilon} A^{-1} & 0 \\ 0 & c^{-1}\mathbf{I} \end{bmatrix}$. Let $c \to \infty$,

$$\begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}^{-1} \succeq \begin{bmatrix} \frac{1}{1+\epsilon} A^{-1} & 0 \\ 0 & 0 \end{bmatrix}.$$

Let $\epsilon \to 0^+$, we have

$$\begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}^{-1} \succeq \begin{bmatrix} A^{-1} & 0 \\ 0 & 0 \end{bmatrix}.$$