Theory of Linear Models
Solutions to Exercises 2

27 January 2020

1. For the multiple regression model given by $\boldsymbol{\beta}' = [\beta_0\ \beta_1\ \beta_2]$ and

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}:$$

   (a) find rank($\mathbf{X}$);
   The first two columns of $\mathbf{X}$ are orthogonal, while the third is a linear combination of the other two, so rank($\mathbf{X}$) = 2.

   (b) find a generalized inverse of $\mathbf{X}'\mathbf{X}$;
   We can use $\mathbf{X}^* = [1\ -1\ 1]$, for which

$$\mathbf{X}^{*\prime}\mathbf{X}^* = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix}.$$

   Since

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 4 \\ 0 & 4 & 4 \end{bmatrix},$$

   we have

$$\mathbf{X}'\mathbf{X} + \mathbf{X}^{*\prime}\mathbf{X}^* = \begin{bmatrix} 5 & -1 & 1 \\ -1 & 5 & 3 \\ 1 & 3 & 5 \end{bmatrix},$$

   so that

$$(\mathbf{X}'\mathbf{X})^- = \begin{bmatrix} \frac{1}{4} & \frac{1}{8} & -\frac{1}{8} \\ \frac{1}{8} & \frac{3}{8} & -\frac{1}{4} \\ -\frac{1}{8} & -\frac{1}{4} & \frac{3}{8} \end{bmatrix}.$$

   Note that there are other generalized inverses.

(c) hence find a least squares estimator of $\boldsymbol{\beta}$;

$$\begin{aligned}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y} &= \begin{bmatrix} Y_1+Y_2+Y_3+Y_4 \\ Y_3+Y_4-Y_1-Y_2 \\ Y_3+Y_4-Y_1-Y_2 \end{bmatrix} \\ \Rightarrow \hat{\boldsymbol{\beta}} &= \begin{bmatrix} \bar{Y} \\ \frac{1}{2}(Y_3+Y_4)-\frac{1}{2}(Y_1-Y_2) \\ \frac{1}{2}(Y_3+Y_4)-\frac{1}{2}(Y_1-Y_2) \end{bmatrix}.\end{aligned}$$

(d) check whether or not β_1 is estimable;
$\beta_1 = \mathbf{c}'\boldsymbol{\beta}$, where $\mathbf{c}' = [0\ 1\ 0]$.

$$\begin{aligned}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} &= \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \\ \Rightarrow \mathbf{I}-(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} &= \begin{bmatrix} 0 & 0 & 0 \\ -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix} \\ \Rightarrow \mathbf{c}'\{\mathbf{I}-(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}\} &= \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix},\end{aligned}$$

so β_1 is not estimable.

(e) check whether or not $\beta_1+\beta_2$ is estimable;
$\beta_1+\beta_2 = \mathbf{c}'\boldsymbol{\beta}$, where $\mathbf{c}' = [0\ 1\ 1]$. Hence $\mathbf{c}'\{\mathbf{I}-(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}\} = [0\ 0\ 0]$ and so $\beta_1+\beta_2$ is estimable.

2. Consider the half-replicate fractional factorial design for three factors, each at two levels, in four runs:

X_1	X_2	X_3
-1	-1	-1
-1	1	1
1	-1	1
1	1	-1

Consider fitting the model

$$\mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3,$$

with the usual covariance assumptions.

(a) Write down **X**.

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 \end{bmatrix}.$$

(b) Find rank(**X**).
The first 4 columns are mutually orthogonal and the others are linear combinations of these, hence rank($\mathbf{X}$) = 4.

(c) Find a generalized inverse of $\mathbf{X}'\mathbf{X}$.
We can use (among others)

$$\mathbf{X}^* = \begin{bmatrix} 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 \end{bmatrix},$$

so that

$$\mathbf{X}^{*\prime}\mathbf{X}^* = \begin{bmatrix} 3 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 & -1 & -1 & 3 \\ -1 & -1 & 3 & -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 & 3 & -1 & -1 \\ -1 & -1 & -1 & 3 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 & -1 & 3 & -1 \\ -1 & 3 & -1 & -1 & -1 & -1 & 3 \end{bmatrix}.$$

Since

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 & 0 & -4 \\ 0 & 0 & 4 & 0 & 0 & -4 & 0 \\ 0 & 0 & 0 & 4 & -4 & 0 & 0 \\ 0 & 0 & 0 & 4 & -4 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & -4 & 0 \\ 0 & 4 & 0 & 0 & 0 & 0 & -4 \end{bmatrix},$$

$\mathbf{X}'\mathbf{X} + \mathbf{X}^{*\prime}\mathbf{X}^* = 8\mathbf{I} - \mathbf{J}$, where $\mathbf{J}$ is a matrix of 1s, and $(\mathbf{X}'\mathbf{X})^- = \frac{1}{8}(\mathbf{I} + \mathbf{J})$.

(d) Hence find a least squares estimator of $\boldsymbol{\beta}$.

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} Y_1 + Y_2 + Y_3 + Y_4 \\ Y_3 + Y_4 - Y_1 - Y_2 \\ Y_2 + Y_4 - Y_1 - Y_3 \\ Y_2 + Y_3 - Y_1 - Y_4 \\ Y_1 + Y_4 - Y_2 - Y_3 \\ Y_1 + Y_3 - Y_2 - Y_4 \\ Y_1 + Y_2 - Y_3 - Y_4 \end{bmatrix}$$

and so

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \bar{Y} \\ (Y_3 + Y_4 - Y_1 - Y_2)/4 \\ (Y_2 + Y_4 - Y_1 - Y_3)/4 \\ (Y_2 + Y_3 - Y_1 - Y_4)/4 \\ (Y_1 + Y_4 - Y_2 - Y_3)/4 \\ (Y_1 + Y_3 - Y_2 - Y_4)/4 \\ (Y_1 + Y_2 - Y_3 - Y_4)/4 \end{bmatrix}.$$

(e) Check whether or not β_1 is estimable.

$$\mathbf{I} - (\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X} = -\frac{1}{2}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 1 \\ -1 & 0 & 1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 1 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and $\mathbf{c}' = [0\,1\,0\,0\,0\,0\,0]$. Hence $\mathbf{c}'(\mathbf{I} - (\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X}) = \frac{1}{2}[-1\,1\,0\,0\,0\,0\,1]$ and so β_1 is not estimable.

(f) Check whether or not $\beta_1 - \beta_{23}$ is estimable.

$$\mathbf{c}' = [0\ 1\ 0\ 0\ 0\ 0\ \ -1],$$

so that $\mathbf{c}'(\mathbf{I}-(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}) = \mathbf{0}'$, showing that $\beta_1-\beta_{23}$ is estimable.

(g) Obtain the variance of the least squares estimator of $\beta_1 - \beta_{23}$.

$$V(\mathbf{c}'\hat{\boldsymbol{\beta}}) = \sigma^2\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{c} = \sigma^2/4.$$

3. Derive the maximum likelihood estimator of σ^2 in the general linear model.

   From the likelihood given in lectures, the log-likelihood is

$$\begin{aligned} l(\boldsymbol{\beta}, \sigma^2|\mathbf{Y}) &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}) \\ \Rightarrow \frac{\partial l(\boldsymbol{\beta}, \sigma^2|\mathbf{Y})}{\partial\sigma^2} &= -\frac{n}{2} + \frac{1}{2\sigma^4}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}). \end{aligned}$$

   Equating this to zero, we get

$$\begin{aligned} -n\hat{\sigma}^2 + (\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}) &= 0 \\ \Rightarrow \hat{\sigma}^2 &= \frac{SS_R}{n}. \end{aligned}$$

4. Write down a linear model for which the maximum likelihood estimator of σ^2 is the minimum mean square error estimator.

   Any model whose design matrix has rank 2 will do, e.g. a simple linear regression model, with at least two distinct values of x.

5. Explain what would be done by an M-estimator of $\boldsymbol{\beta}$ with

$$\rho(u) = \begin{cases} u^2 & a \le u \le a; \\ 0 & \text{otherwise.} \end{cases}$$

   (Actually, this is not quite well-defined. We need some side conditions, such as having a such that at least n_0 observations have $-a \le \epsilon_i/s \le a$. You don't need to think about this to answer the question.)

   This is giving zero weight to observations with large residuals and using least squares for the others. Hence it is equivalent to removing outliers and fitting the remaining observations by least squares.

6. Consider fitting a multiple linear regression model to the following data:

Y	X_1	X_2	X_3
10	-1	-1	-1.0001
12	-1	1	-0.9999
20	1	-1	1.0000
22	1	1	1.0000

(a) Calculate $(\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1}$ for $\lambda = 0, 10^{-8}, 10^{-7}, \ldots, 1$. (I suggest using R or another package which inverts matrices.)

Here are my R commands and results:

```
> Y <- c(10,12,20,22)
> X <- matrix(c(1,1,1,1,-1,-1,1,1,-1,1,-1,1,-1.0001,-0.9999,1,1),4,4)
> XtX <- t(X)%*%X
> Var0 <- solve(XtX)
> Var1 <- solve(XtX+0.00000001*diag(4))
> Var2 <- solve(XtX+0.0000001*diag(4))
> Var3 <- solve(XtX+0.000001*diag(4))
> Var4 <- solve(XtX+0.00001*diag(4))
> Var5 <- solve(XtX+0.0001*diag(4))
> Var6 <- solve(XtX+0.001*diag(4))
> Var7 <- solve(XtX+0.01*diag(4))
> Var8 <- solve(XtX+0.1*diag(4))
> Var9 <- solve(XtX+diag(4))
> Var0
     [,1]   [,2]   [,3]   [,4]
[1,] 0.25  0e+00  0e+00  0e+00
[2,] 0.00  1e+08  5e+03 -1e+08
[3,] 0.00  5e+03  5e-01 -5e+03
[4,] 0.00 -1e+08 -5e+03  1e+08
> Var1
     [,1]          [,2]          [,3]          [,4]
[1,] 0.25         0.000     0.0000000         0.000
[2,] 0.00  33333333.659  1666.6666704 -33333333.492
[3,] 0.00      1666.667     0.3333333     -1666.667
[4,] 0.00 -33333333.492 -1666.6666746  33333333.576
> Var2
     [,1]         [,2]        [,3]          [,4]
[1,] 0.25       0.0000   0.0000000        0.0000
[2,] 0.00 4761904.8186 238.0952284 -4761904.6876
```

```
[3,] 0.00      238.0952    0.2619048     -238.0952
[4,] 0.00 -4761904.6876 -238.0952344  4761904.8067
> Var3
          [,1]          [,2]         [,3]          [,4]
[1,] 0.2499999       0.00000   0.0000000       0.00000
[2,] 0.0000000  497512.50028  24.8756125 -497512.37466
[3,] 0.0000000      24.87561   0.2512437     -24.87562
[4,] 0.0000000 -497512.37466 -24.8756187  497512.49904
> Var4
          [,1]          [,2]        [,3]          [,4]
[1,] 0.2499994      0.000000  0.0000000      0.000000
[2,] 0.0000000  49975.074996  2.4987412 -49974.949934
[3,] 0.0000000      2.498741  0.2501243     -2.498747
[4,] 0.0000000 -49974.949934 -2.4987475  49975.074871
> Var5
          [,1]          [,2]        [,3]          [,4]
[1,] 0.2499938     0.0000000  0.0000000     0.0000000
[2,] 0.0000000  4999.8125117  0.2499781 -4999.6875070
[3,] 0.0000000     0.2499781  0.2500062    -0.2499844
[4,] 0.0000000 -4999.6875070 -0.2499844  4999.8124992
> Var6
          [,1]          [,2]         [,3]          [,4]
[1,] 0.2499375    0.00000000   0.00000000    0.00000000
[2,] 0.0000000  500.05999220   0.02499050 -499.93500720
[3,] 0.0000000    0.02499050   0.24993877   -0.02499675
[4,] 0.0000000 -499.93500720  -0.02499675  500.05999095
> Var7
          [,1]          [,2]         [,3]          [,4]
[1,] 0.2493766   0.000000000  0.000000000   0.000000000
[2,] 0.0000000  50.062396973  0.002490651 -49.937552965
[3,] 0.0000000   0.002490651  0.249376683  -0.002496878
[4,] 0.0000000 -49.937552965 -0.002496878  50.062396848
> Var8
          [,1]          [,2]          [,3]          [,4]
[1,] 0.2439024  0.0000000000  0.0000000000  0.0000000000
[2,] 0.0000000  5.0617281452  0.0002408913 -4.9382713489
[3,] 0.0000000  0.0002408913  0.2439024511 -0.0002469136
[4,] 0.0000000 -4.9382713489 -0.0002469136  5.0617281326
> Var9
     [,1]          [,2]          [,3]          [,4]
[1,]  0.2  0.000000e+00  0.000000e+00  0.000000e+00
```

```
[2,]  0.0  5.555556e-01  1.777778e-05 -4.444444e-01
[3,]  0.0  1.777778e-05  2.000000e-01 -2.222222e-05
[4,]  0.0 -4.444444e-01 -2.222222e-05  5.555556e-01
>
```

(b) Obtain the ridge regression estimates of β_1, β_2 and β_3 in each of these cases.

My R commands and output are:

```
> XtY <- t(X)%*%Y
> betahat0 <- Var0%*%XtY
> betahat1 <- Var1%*%XtY
> betahat2 <- Var2%*%XtY
> betahat3 <- Var3%*%XtY
> betahat4 <- Var4%*%XtY
> betahat5 <- Var5%*%XtY
> betahat6 <- Var6%*%XtY
> betahat7 <- Var7%*%XtY
> betahat8 <- Var8%*%XtY
> betahat9 <- Var9%*%XtY
> betahat0[2:4]
[1]  5.000000e+00  1.000000e+00 -7.951166e-08
> betahat1[2:4]
[1] 3.3333166 0.9999167 1.6666834
> betahat2[2:4]
[1] 2.619024 0.999881 2.380976
> betahat3[2:4]
[1] 2.5124126 0.9998754 2.4875867
> betahat4[2:4]
[1] 2.5012213 0.9998726 2.4987725
> betahat5[2:4]
[1] 2.500069 0.999850 2.499869
> betahat6[2:4]
[1] 2.4996750 0.9996251 2.4997000
> betahat7[2:4]
[1] 2.4968552 0.9973817 2.4969026
> betahat8[2:4]
[1] 2.4691118 0.9754893 2.4691604
> betahat9[2:4]
[1] 2.2222045 0.7999111 2.2222444
>
```

(c) Comment on the results.

The results are fairly stable for $\lambda = 10^{-6}$ and higher.