BIS623 Homework 8

Due on 11/21/2019 before the lecture

Joanna Chen

17 November, 2019

Problem 1

Please derive the variance-covariance matrix of $\hat{\beta}_R$ (the estimate of the ridge regression).

We know that

$$\begin{aligned} E\left(\hat{\beta}_R\right) &= E\left(\left(X^T X + \lambda I_p\right)^{-1} X^T Y\right) \\ &= \left(X^T X + \lambda I_p\right)^{-1} X^T X \beta \end{aligned}$$

$$\begin{aligned} \mathrm{Var}\left(\hat{\beta}_R\right) &= \mathrm{Var}\left(\left(X^T X + \lambda I_p\right)^{-1} X^T Y\right) \\ &= \mathrm{Var}\left(\left(X^T X + \lambda I_p\right)^{-1} X^T X \hat{\beta}\right) \quad \text{Let us denote } \left(X^T X + \lambda I_p\right)^{-1} X^T X \text{ ty } A \\ &= A\,\mathrm{Var}(\hat{\beta}) A^\top \\ &= \sigma^2 A \left(X'X\right)^{-1} A^\top \\ &= \sigma^2 \left(X^\top X + \lambda I_p\right)^{-1} X^\top X \left(X'X\right)^{-1} \left[\left(X^\top X + \lambda I_p\right)^{-1} X^\top X\right]^\top \\ &= \sigma^2 \left(X^\top X + \lambda I_p\right)^{-1} X^\top X \left[\left(X^\top X + \lambda I_p\right)^{-1}\right]^\top \quad \text{since } \left(X^\top X\right)^\top = X^\top X \end{aligned}$$

Problem 2

Without a consideration of intercept, assume we have a design matrix (an orthonormal design matrix)

$$X = \frac{1}{2}\begin{pmatrix} -1 & -1 \\ -1 & 1 \\ 1 & -1 \\ 1 & 1 \end{pmatrix}$$

and an outcome Y. Please derive the ordinary least squares estimator and ridge estimator to verify the ridge estimator scales the OLS estimator by a factor in this case. OLS estimate:

$$\hat{\beta} = \left(X^\top X\right)^{-1} X^\top Y$$

$$\begin{aligned} X^\top X &= \frac{1}{2}\begin{pmatrix} -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \end{pmatrix} \frac{1}{2}\begin{pmatrix} -1 & -1 \\ -1 & 1 \\ 1 & -1 \\ 1 & 1 \end{pmatrix} \\ &= \frac{1}{4}\begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= (X^\top X)^{-1} \end{aligned}$$

$$
\begin{aligned}
\left(X^\top X\right)^{-1} X^\top &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \frac{1}{2} \begin{pmatrix} -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \end{pmatrix} \\
&= \frac{1}{2} \begin{pmatrix} -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \end{pmatrix} \\
&= \begin{pmatrix} -0.5 & -0.5 & 0.5 & 0.5 \\ -0.5 & 0.5 & -0.5 & 0.5 \end{pmatrix}
\end{aligned}
$$

Use R to check the answer

```
A = (1/2)*(matrix(c(-1,-1,-1,1,1,-1,1,1),nrow = 4, ncol = 2, byrow = TRUE))
solve(t(A)%*%A)%*%t(A)
```

```
##      [,1] [,2] [,3] [,4]
## [1,] -0.5 -0.5  0.5  0.5
## [2,] -0.5  0.5 -0.5  0.5
```

Therefore,

$$
\hat{\beta} = \begin{pmatrix} -0.5 & -0.5 & 0.5 & 0.5 \\ -0.5 & 0.5 & -0.5 & 0.5 \end{pmatrix} Y
$$

Ridge estimator:

$$
\begin{aligned}
\hat{\beta_R} &= \left(X^\top X + \lambda I_p\right)^{-1} X^\top Y \\
&= \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \right]^{-1} X^\top Y \\
&= \begin{pmatrix} 1+\lambda & 0 \\ 0 & 1+\lambda \end{pmatrix}^{-1} \frac{1}{2} \begin{pmatrix} -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \end{pmatrix} Y \\
&= \begin{pmatrix} \frac{1}{1+\lambda} & 0 \\ 0 & \frac{1}{1+\lambda} \end{pmatrix} \begin{pmatrix} -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{pmatrix} Y \\
&= \begin{pmatrix} -\frac{1}{2(1+\lambda)} & -\frac{1}{2(1+\lambda)} & \frac{1}{2(1+\lambda)} & \frac{1}{2(1+\lambda)} \\ -\frac{1}{2(1+\lambda)} & \frac{1}{2(1+\lambda)} & -\frac{1}{2(1+\lambda)} & \frac{1}{2(1+\lambda)} \end{pmatrix} Y \\
&= \frac{1}{1+\lambda} \begin{pmatrix} -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{pmatrix} Y \\
&= \frac{1}{1+\lambda} \hat{\beta}
\end{aligned}
$$

Therefore, Ridge estimator scales the OLS estimator by the fatocr $\frac{1}{1+\lambda}$ in this case.

Problem 3

Please do a forward stepwise selection for the `state` data used in the class. Is the result consistent with the one from backward stepwise?

```
data(state)
statedata = data.frame(state.x77,row.names=state.abb)
fit <- lm(Life.Exp~.,data = statedata) # Full model
fit.base <- lm(Life.Exp~1,data=statedata) # Intercept-only model
```

I used two ways to do forward stepwise selection. Foe the first way, I used the step() function which automatically compared AIC.

```
fit.forward <- step(fit.base,scope=list(lower=formula(fit.base),upper=formula(fit)), direction = "forwa
```

```
## Start:  AIC=30.44
## Life.Exp ~ 1
##
##              Df Sum of Sq    RSS     AIC
## + Murder      1    53.838 34.461 -14.609
## + Illiteracy  1    30.578 57.721  11.179
## + HS.Grad     1    29.931 58.368  11.737
## + Income      1    10.223 78.076  26.283
## + Frost       1     6.064 82.235  28.878
## <none>                    88.299  30.435
## + Area        1     1.017 87.282  31.856
## + Population  1     0.409 87.890  32.203
##
## Step:  AIC=-14.61
## Life.Exp ~ Murder
##
##              Df Sum of Sq    RSS     AIC
## + HS.Grad     1    4.6910 29.770 -19.925
## + Population  1    4.0161 30.445 -18.805
## + Frost       1    3.1346 31.327 -17.378
## + Income      1    2.4047 32.057 -16.226
## <none>                    34.461 -14.609
## + Area        1    0.4697 33.992 -13.295
## + Illiteracy  1    0.2732 34.188 -13.007
##
## Step:  AIC=-19.93
## Life.Exp ~ Murder + HS.Grad
##
##              Df Sum of Sq    RSS     AIC
## + Frost       1    4.3987 25.372 -25.920
## + Population  1    3.3405 26.430 -23.877
## <none>                    29.770 -19.925
## + Illiteracy  1    0.4419 29.328 -18.673
## + Area        1    0.2775 29.493 -18.394
## + Income      1    0.1022 29.668 -18.097
##
## Step:  AIC=-25.92
## Life.Exp ~ Murder + HS.Grad + Frost
##
##              Df Sum of Sq    RSS     AIC
## + Population  1   2.06358 23.308 -28.161
## <none>                    25.372 -25.920
## + Income      1   0.18232 25.189 -24.280
## + Illiteracy  1   0.17184 25.200 -24.259
## + Area        1   0.02573 25.346 -23.970
##
## Step:  AIC=-28.16
## Life.Exp ~ Murder + HS.Grad + Frost + Population
##
##              Df Sum of Sq    RSS     AIC
## <none>                    23.308 -28.161
## + Income      1 0.0060582 23.302 -26.174
## + Illiteracy  1 0.0039221 23.304 -26.170
## + Area        1 0.0007900 23.307 -26.163
```

```
summary(fit.forward)
```

```
##
## Call:
## lm(formula = Life.Exp ~ Murder + HS.Grad + Frost + Population,
##     data = statedata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.47095 -0.53464 -0.03701  0.57621  1.50683
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.103e+01  9.529e-01  74.542  < 2e-16 ***
## Murder      -3.001e-01  3.661e-02  -8.199 1.77e-10 ***
## HS.Grad      4.658e-02  1.483e-02   3.142  0.00297 **
## Frost       -5.943e-03  2.421e-03  -2.455  0.01802 *
## Population   5.014e-05  2.512e-05   1.996  0.05201 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7197 on 45 degrees of freedom
## Multiple R-squared:  0.736,  Adjusted R-squared:  0.7126
## F-statistic: 31.37 on 4 and 45 DF,  p-value: 1.696e-12
```

```
formula(fit.forward)
```

```
## Life.Exp ~ Murder + HS.Grad + Frost + Population
```

For the second way, I used R function addterm() in the MASS package which applys an F-test criterion or a P-value criterion. I start with intercept-only model and then admit the predictor with the smallest P-value until it's above my Alpha-level

```
library(MASS)
addterm( fit.base, scope = fit, test="F" )
```

```
## Single term additions
##
## Model:
## Life.Exp ~ 1
##            Df Sum of Sq    RSS     AIC F Value     Pr(F)
## <none>                  88.299  30.435
## Population  1     0.409 87.890  32.203   0.223   0.63866
## Income      1    10.223 78.076  26.283   6.285   0.01562 *
## Illiteracy  1    30.578 57.721  11.179  25.429 6.969e-06 ***
## Murder      1    53.838 34.461 -14.609  74.989 2.260e-11 ***
## HS.Grad     1    29.931 58.368  11.737  24.615 9.196e-06 ***
## Frost       1     6.064 82.235  28.878   3.540   0.06599 .
## Area        1     1.017 87.282  31.856   0.559   0.45815
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
NewMod = update( fit.base, .~. + Murder )
addterm( NewMod, scope = fit, test="F" )
```

```
## Single term additions
```

```
##
## Model:
## Life.Exp ~ Murder
##            Df Sum of Sq    RSS     AIC F Value    Pr(F)
## <none>                  34.461 -14.609
## Population  1    4.0161 30.445 -18.805  6.1999 0.016369 *
## Income      1    2.4047 32.057 -16.226  3.5257 0.066636 .
## Illiteracy  1    0.2732 34.188 -13.007  0.3756 0.542910
## HS.Grad     1    4.6910 29.770 -19.925  7.4059 0.009088 **
## Frost       1    3.1346 31.327 -17.378  4.7029 0.035205 *
## Area        1    0.4697 33.992 -13.295  0.6494 0.424375
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
NewMod = update( NewMod, .~. + HS.Grad )
addterm( NewMod, scope = fit, test="F" )
```

```
## Single term additions
##
## Model:
## Life.Exp ~ Murder + HS.Grad
##            Df Sum of Sq    RSS     AIC F Value    Pr(F)
## <none>                  29.770 -19.925
## Population  1    3.3405 26.430 -23.877  5.8141 0.019949 *
## Income      1    0.1022 29.668 -18.097  0.1585 0.692418
## Illiteracy  1    0.4419 29.328 -18.673  0.6931 0.409421
## Frost       1    4.3987 25.372 -25.920  7.9751 0.006988 **
## Area        1    0.2775 29.493 -18.394  0.4329 0.513863
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
NewMod = update( NewMod, .~. + Frost )
addterm( NewMod, scope = fit, test="F" )
```

```
## Single term additions
##
## Model:
## Life.Exp ~ Murder + HS.Grad + Frost
##            Df Sum of Sq    RSS     AIC F Value   Pr(F)
## <none>                  25.372 -25.920
## Population  1   2.06358 23.308 -28.161  3.9841 0.05201 .
## Income      1   0.18232 25.189 -24.280  0.3257 0.57103
## Illiteracy  1   0.17184 25.200 -24.259  0.3069 0.58236
## Area        1   0.02573 25.346 -23.970  0.0457 0.83173
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
AIC(NewMod)
```

```
## [1] 117.9743
```

```
NewMod = update( NewMod, .~. + Population )
addterm( NewMod, scope = fit, test="F" )
```

```
## Single term additions
##
## Model:
```

```
## Life.Exp ~ Murder + HS.Grad + Frost + Population
##            Df Sum of Sq    RSS     AIC   F Value  Pr(F)
## <none>                  23.308 -28.161
## Income      1 0.0060582 23.302 -26.174 0.0114395 0.9153
## Illiteracy  1 0.0039221 23.304 -26.170 0.0074052 0.9318
## Area        1 0.0007900 23.307 -26.163 0.0014914 0.9694
```

```
AIC(NewMod) #AIC reduce, admit the Population Variable.
```

```
## [1] 115.7326
```

```
formula(NewMod)
```

```
## Life.Exp ~ Murder + HS.Grad + Frost + Population
```

```
# Backward Elimination
fit.backward <- step(fit,scope=list(lower=fit.base,upper=fit), direction = "backward")
```

```
## Start:  AIC=-22.18
## Life.Exp ~ Population + Income + Illiteracy + Murder + HS.Grad +
##     Frost + Area
##
##              Df Sum of Sq    RSS     AIC
## - Area        1    0.0011 23.298 -24.182
## - Income      1    0.0044 23.302 -24.175
## - Illiteracy  1    0.0047 23.302 -24.174
## <none>                    23.297 -22.185
## - Population  1    1.7472 25.044 -20.569
## - Frost       1    1.8466 25.144 -20.371
## - HS.Grad     1    2.4413 25.738 -19.202
## - Murder      1   23.1411 46.438  10.305
##
## Step:  AIC=-24.18
## Life.Exp ~ Population + Income + Illiteracy + Murder + HS.Grad +
##     Frost
##
##              Df Sum of Sq    RSS     AIC
## - Illiteracy  1    0.0038 23.302 -26.174
## - Income      1    0.0059 23.304 -26.170
## <none>                    23.298 -24.182
## - Population  1    1.7599 25.058 -22.541
## - Frost       1    2.0488 25.347 -21.968
## - HS.Grad     1    2.9804 26.279 -20.163
## - Murder      1   26.2721 49.570  11.569
##
## Step:  AIC=-26.17
## Life.Exp ~ Population + Income + Murder + HS.Grad + Frost
##
##              Df Sum of Sq    RSS     AIC
## - Income      1     0.006 23.308 -28.161
## <none>                    23.302 -26.174
## - Population  1     1.887 25.189 -24.280
## - Frost       1     3.037 26.339 -22.048
## - HS.Grad     1     3.495 26.797 -21.187
## - Murder      1    34.739 58.041  17.456
##
```

```
## Step:  AIC=-28.16
## Life.Exp ~ Population + Murder + HS.Grad + Frost
##
##              Df Sum of Sq    RSS     AIC
## <none>                    23.308 -28.161
## - Population  1     2.064 25.372 -25.920
## - Frost       1     3.122 26.430 -23.877
## - HS.Grad     1     5.112 28.420 -20.246
## - Murder      1    34.816 58.124  15.528
```

```
fit.backward
```

```
##
## Call:
## lm(formula = Life.Exp ~ Population + Murder + HS.Grad + Frost,
##     data = statedata)
##
## Coefficients:
## (Intercept)   Population       Murder      HS.Grad        Frost
##   7.103e+01    5.014e-05   -3.001e-01    4.658e-02   -5.943e-03
```

```
step(fit,scope=list(lower=fit.base,upper=fit), direction = "backward")
```

```
## Start:  AIC=-22.18
## Life.Exp ~ Population + Income + Illiteracy + Murder + HS.Grad +
##     Frost + Area
##
##              Df Sum of Sq    RSS     AIC
## - Area        1    0.0011 23.298 -24.182
## - Income      1    0.0044 23.302 -24.175
## - Illiteracy  1    0.0047 23.302 -24.174
## <none>                    23.297 -22.185
## - Population  1    1.7472 25.044 -20.569
## - Frost       1    1.8466 25.144 -20.371
## - HS.Grad     1    2.4413 25.738 -19.202
## - Murder      1   23.1411 46.438  10.305
##
## Step:  AIC=-24.18
## Life.Exp ~ Population + Income + Illiteracy + Murder + HS.Grad +
##     Frost
##
##              Df Sum of Sq    RSS     AIC
## - Illiteracy  1    0.0038 23.302 -26.174
## - Income      1    0.0059 23.304 -26.170
## <none>                    23.298 -24.182
## - Population  1    1.7599 25.058 -22.541
## - Frost       1    2.0488 25.347 -21.968
## - HS.Grad     1    2.9804 26.279 -20.163
## - Murder      1   26.2721 49.570  11.569
##
## Step:  AIC=-26.17
## Life.Exp ~ Population + Income + Murder + HS.Grad + Frost
##
##              Df Sum of Sq    RSS     AIC
## - Income      1     0.006 23.308 -28.161
## <none>                    23.302 -26.174
```

```
## - Population  1     1.887 25.189 -24.280
## - Frost       1     3.037 26.339 -22.048
## - HS.Grad     1     3.495 26.797 -21.187
## - Murder      1    34.739 58.041  17.456
##
## Step:  AIC=-28.16
## Life.Exp ~ Population + Murder + HS.Grad + Frost
##
##              Df Sum of Sq    RSS     AIC
## <none>                     23.308 -28.161
## - Population  1     2.064 25.372 -25.920
## - Frost       1     3.122 26.430 -23.877
## - HS.Grad     1     5.112 28.420 -20.246
## - Murder      1    34.816 58.124  15.528
```

```
##
## Call:
## lm(formula = Life.Exp ~ Population + Murder + HS.Grad + Frost,
##     data = statedata)
##
## Coefficients:
## (Intercept)   Population       Murder      HS.Grad        Frost
##   7.103e+01    5.014e-05   -3.001e-01    4.658e-02   -5.943e-03
```

```
summary(fit.backward)
```

```
##
## Call:
## lm(formula = Life.Exp ~ Population + Murder + HS.Grad + Frost,
##     data = statedata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.47095 -0.53464 -0.03701  0.57621  1.50683
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.103e+01  9.529e-01  74.542  < 2e-16 ***
## Population   5.014e-05  2.512e-05   1.996  0.05201 .
## Murder      -3.001e-01  3.661e-02  -8.199 1.77e-10 ***
## HS.Grad      4.658e-02  1.483e-02   3.142  0.00297 **
## Frost       -5.943e-03  2.421e-03  -2.455  0.01802 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7197 on 45 degrees of freedom
## Multiple R-squared:  0.736,  Adjusted R-squared:  0.7126
## F-statistic: 31.37 on 4 and 45 DF,  p-value: 1.696e-12
```

```
formula(fit.backward)
```

```
## Life.Exp ~ Population + Murder + HS.Grad + Frost
```

Comparing with backward selection, the result is consistent.