title: “Data Science 1 : A2”

author: “Abigail Chen

output: pdf_document”

Exercise Part 1

Consider the simplest possible regression model

$$Y_i = \beta_0 + \epsilon_i$$

where ϵ_i *i = 1,…,n* are independent and identically distributed random variables with $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2$. The ridge estimator of β_0 solves.

$$\min_b [\sum_{i=1}^{n} (Y_i - b)^2 + \lambda b^2]$$

For some $\lambda \geq 0$. In the special case $\lambda = 0$, the solution is of course the OLS estimator. (a) Show that the solution to this problem is given by $\hat{beta}_0^{ridge} = \sum_{i=1}^{n} Y_i/(n+\lambda)$. Compare this to the OLS estimator $\hat{beta}_0^{OLS} = \bar{Y}$.

$$\begin{aligned}
\min_b [\sum_{i=1}^{n} (Y_i - b)^2 + \lambda b^2] = \frac{\partial}{\partial b}[\sum_{i=1}^{n} Y_i^2 - 2b\sum_{i=1}^{n} Y_i + b^2 n + \lambda b^2] &= 0 \\
-2\sum_{i=1}^{n} Y_i + 2bn + 2\lambda b &= 0 \\
-2\sum_{i=1}^{n} Y_i + 2b(n+\lambda) &= 0 \\
\beta_0^{ridge} &= \frac{\sum_{i=1}^{n} Y_i}{n+\lambda}
\end{aligned}$$

b. Suppose that $\beta_0 = 1$ and $\epsilon \sim N(0, \sigma^2)$ with $\sigma^2 = 4$. Generate a sample of size *n = 10* from a model and compute $\hat{beta}_0^{ridge}$ for a grid λ values over the interval $[0, 20]$.

```
sampleBeta <- function(len_beta) {
    4 / seq(len_beta)^2
}
functionyx <- function(xvalue) {
    beta <- sampleBeta(dim(xvalue)[2])
    xvalue %*% beta
}
```

```
sim <- function(xgen, x0 = 0.2, lambdas = seq(0, 20, 1)) {
    # sample generator
    xvalue <- xgen()
    y_e <- functionyx(xvalue)
    yvalue <- y_e + rnorm(length(y_e)) * 4


    # x0 value generator
    x_check <- matrix(x0, ncol = dim(xvalue)[2])

    map_df(lambdas, ~{
        model <- glmnet(xvalue,  yvalue, alpha = 0, lambda = .x)
        tibble(
            lambda = .x,
            fhat = as.numeric(predict(model, newx =   x_check )),
            error = as.numeric(functionyx(  x_check )) - fhat
        )
    })
}
```

```
sim_visualizer <- function(results) {
    group_by(results, lambda) %>%
    summarise(bias2 = mean(error)^2, var = var(fhat)) %>%
    mutate(MSE = bias2 + var) %>%
    pivot_longer(bias2:MSE, names_to = "metric") %>%
    mutate(metric = factor(metric, levels = c("bias2", "var", "MSE"))) %>%
    ggplot(aes(lambda, value, color = metric)) + geom_line(size = 2)
}
```

c. Repeat part *b)*, say 1000 times so that you end up with 1000 estiamtes of β_0 for all the λ values that you have picked. For each value of λ, compute $bias^2[\hat{beta}_0^{ridge}]$, $Var[\hat{beta}_0^{ridge}]$ and $MSE[\hat{beta}_0^{ridge}] = bias^2[\hat{beta}_0^{ridge}] + Var[\hat{beta}_0^{ridge}]$.

```
indie <- function(n = 10, p = 1) {
    matrix(rnorm(n * p), nrow = n, ncol = 2)
}
```

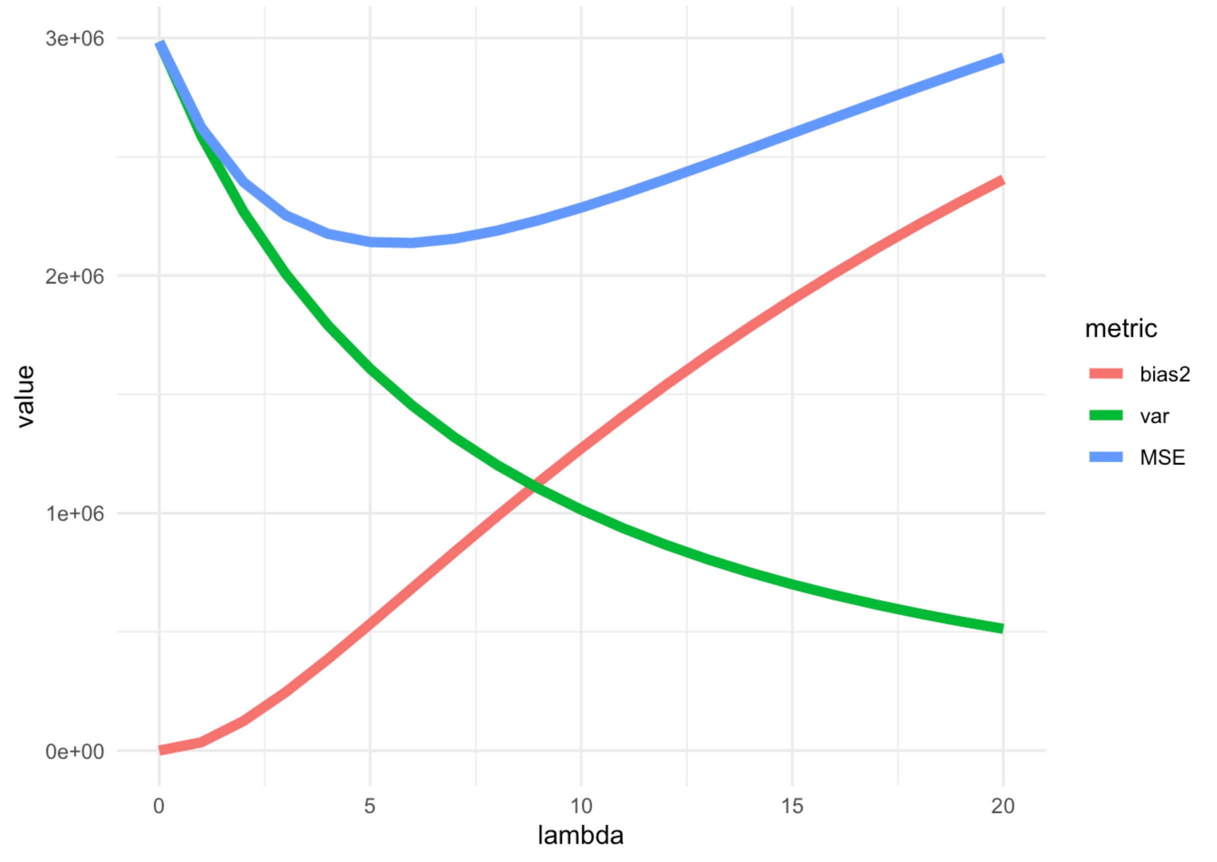
```
# repeat for 1000 times
sim(xgen = indie)
```

```
## # A tibble: 21 × 3
##     lambda  fhat   error
##      <dbl> <dbl>   <dbl>
##  1       0 1.95  -0.954
##  2       1 1.67  -0.669
##  3       2 1.42  -0.423
##  4       3 1.21  -0.210
##  5       4 1.02  -0.0232
##  6       5 0.858  0.142
##  7       6 0.711  0.289
##  8       7 0.579  0.421
##  9       8 0.460  0.540
## 10       9 0.353  0.647
## # … with 11 more rows
```

```
sim_count <- 1000
sim_2<- map_df(
    seq(sim_count),
    sim,
    xgen = indie
)
```

d. Plot $bias^2[\hat{beta}_0^{ridge}]$, $Var[\hat{beta}_0^{ridge}]$ and $MSE[\hat{beta}_0^{ridge}]$ as a function of λ and interpret the result.

```
sim_visualizer(sim_2)
```



This figure shows us the increase for bias 2 and the decrease for the variance.

Exercise Part 2

Let X and Y be two random variables with zero mean. The population version of the optimization problem that defines the first principal component of the two variables is

$$\max_{u1,u2} Var(u_1 X + u_2 Y)$$

Subject to

$$u_1^2 + u_2^2 = 1$$

The following questions ask you to examine some insightful special cases.

a. Suppose that *Var(X) > Var(Y)* and *cov(X, Y) = E(XY) = 0*. Derive the first principle component vector. Draw an illustrative picture and explain the result intuitively. (Hint: expand the variance formula and substitute the constraint.Then carry out the minimization.)

$$\begin{aligned}
Var(u_1 X + u_2 Y) = u_1^2 Var(X) + 2u_1 u_2 Cov(X, Y) + u_2^2 Var(Y) \\
Cov(X, Y) = 0 \\
u_1^2 Var(X) + 2u_1 u_2 0 + u_2^2 Var(Y) \\
\text{This is what we'll get} \\
u_1^2 Var(X) + u_2^2 Var(Y) \\
\\
u_1^2 + u_2^2 = 1 \\
u_1^2 = 1 - u_2^2 \\
\\
(1 - u_2^2)Var(X) + u_2^2 Var(Y) = 0 \\
Var(X) - u_2^2 Var(X) + u_2^2 Var(Y) = 0 \\
\frac{\partial}{\partial u_2}(Var(X) - u_2^2 Var(X) + u_2^2 Var(Y)) = 0 \\
-2u_2 Var(X) + 2u_2 Var(Y) = 0 \\
2u_2(-Var(X) + Var(Y)) = 0 \\
u_2 = 0 \\
u_1^2 + u_2^2 = 1 \\
u_1^2 + 0 = 1 \\
u_1^2 = 1 \\
\sqrt{u_1^2} = \sqrt{1} \\
u_1 = \pm 1 \\
\text{Here is the result} \\
(u_1, u_2) \\
(1, 0), (-1, 0) \\
\text{This shows the result after substitution } u_2^2 \\
(u_1, u_2) \\
(0, 1), (0, -1)
\end{aligned}$$

After expanding the variance formula and substituting the constraint. The minimization is carried out and this is what we get $(1, 0), (-1, 0), (0, 1)$ and $(0, -1)$.

b. Suppose that *Var(X) = Var(Y) = 1* (principle component analysis is often performed after standardization) and *cov(X, Y) = E(XY) = 0*. Show that in this case any vector (u1, u2) with length 1 is a principal component vector (i.e., it solves the problem above). Explain intuitively this puzzling result. (A picture can help.)

$$\begin{aligned}
Var(X) = Var(Y) = 1 \\
cov(X, Y) = E(XY) = 0 \\
Var(u_1 X + u_2 Y) = u_1^2 Var(X) + 2u_1 u_2 Cov(X, Y) + u_2^2 Var(Y) \\
u_1^2 Var(X) + u_2^2 Var(Y) \\
\\
u_1^2 + u_2^2 = 1 \\
u_2^2 = 1 - u_1^2 \\
\\
u_1^2 Var(X) + (1 - u_1^2)Var(Y) = 0 \\
u_1^2 Var(X) + Var(Y) - u_1^2 Var(Y) = 0 \\
\frac{\partial}{\partial u_1}(u_1^2 Var(X) + Var(Y) - u_1^2 Var(Y)) = 0 \\
2u_1 Var(X) - 2u_1 Var(Y) = 0 \\
2u_1(Var(X) - Var(X)) = 0 \\
2u_1 * 0 = 0 \\
0 = 0
\end{aligned}$$

Here we’ll see that *Var(X) = Var(Y) = 1*, after the principle component analysis(PCA) is performed after standardization. We want to minimize the ellipse size from OLS and circle simultaneously in the given ridge regression.

Exercise Part 3

ISLR Exercise 3 in Section 6.8: Suppose we estimate the regression coefficients in a linear regression model by minimizing for a particular value of s. For parts (a) through (e), indicate which of i. through v. is correct. Justify your answer.

$$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p} \beta_i x_{ij})^2$$

Subject to

$$\sum_{j=1}^{p} |\beta_j| \leq s$$

- i. Increase initially, and then eventually start decreasing in an inverted U shape.
- ii. Decrease initially, and then eventually start increasing in a U shape.
- iii. Steadily increase.
- iv. Steadily decrease.
- v. Remain constant.

a. As we increase s from 0, the training RSS will: When we increase *s* from 0, the training RSS will *(iv) steadily decrease* since the RSS is subject to the given constraint. Our model becomes more flexible as the *s* gets larger, and the restriction of the beta is reducing thus minimizing our RSS.

b. Repeat (a) for test RSS.

As we increase *s* from 0, the test RSS will *(ii) initially decrease* and then increase, making a U shape. If the constraint loosens, the model flexibility and the *s* will both increase.

c. Repeat (a) for variance.

As we increase *s* from 0 the variance *(iii) steadily increase*, because the increase in *s* from 0 means a shrinkage reduction, where lambda is decreasing, resulting to an increase in model flexibility.

d. Repeat (a) for (squared) bias. As we increase *s* from 0, squared bias will *(iv) steadily decrease/* because as the model flexibility increases the bias decreases.

e. Repeat (a) for the irreducible error. As we increase *s* from 0, the irreducible error will, *(v) remain constant* because it is independent of the model parameters making it constant all throughout.