

[STAT409] Homework 3

1. Assuming floating number representation with $B = 10$ and $d = 4$, show that $\mathbf{X}^T \mathbf{X}$ is not invertible (Hint: Compute its determinant.)

$$\mathbf{X} = \begin{bmatrix} 1 & 1.000 \\ 1 & 1.000 \\ 1 & 1.001 \\ 1 & 1.001 \end{bmatrix}$$

2. Hitters's data that record salary of Major League Baseball (MLB) players is available in R. You can download the data by running the following code.

```
library(ISLR)
data("Hitters")
Hitters <- na.omit(Hitters) # remove missing values

# response: Salary of Baseball Player
y <- Hitters$Salary # y
# predictors: Players Stats
X <- cbind(Hitters$AtBat, # x1: number of times at batting in the game
           Hitters$Hits, # x2: number of hits
           Hitters$HmRun, # x3: number of homeruns
           Hitters$Runs, # x4: number of runs
           Hitters$Walks, # x5: number of walks
           Hitters$Years) # x6: number of years played in the league
```

For regression computation, you first center the data as follows:

```
n <- nrow(X); p <- ncol(X)
my <- mean(y); mX <- apply(X, 2, mean)
y.c <- y - my; X.c <- t(t(X) - mX)
```

- (a) Please write your own code to compute the OLS estimator of β , fitted values, and residuals for the linear regression for the Hitters data problem using “qr()” function in “R”. (Hint: You may simply apply the code given in the Lecture note)
- (b) Please compare your results obtained from (a) to the result from “lm()” given below:

```
obj <- lm(Salary ~ AtBat + Hits + HmRun + Runs + Walks + Years, data = Hitters)
est <- coef(obj)
y.hat <- fitted(obj)
resid <- resid(obj)
```

4. (Hitters's data) Apply OLS regression, LASSO, Ridge, and Elastic Net Regression to Hitters's data (with appropriately selected λ for the latter regularized methods), and provide their coefficient estimates.