midterm

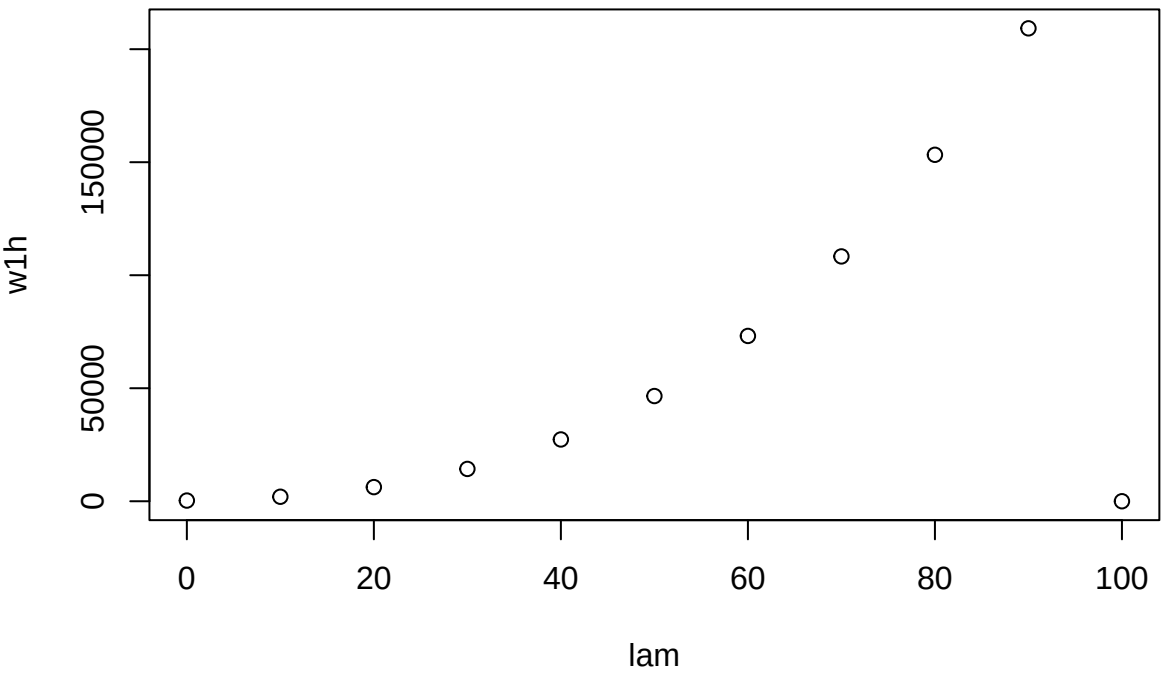
Enbo Tian

2022/3/4

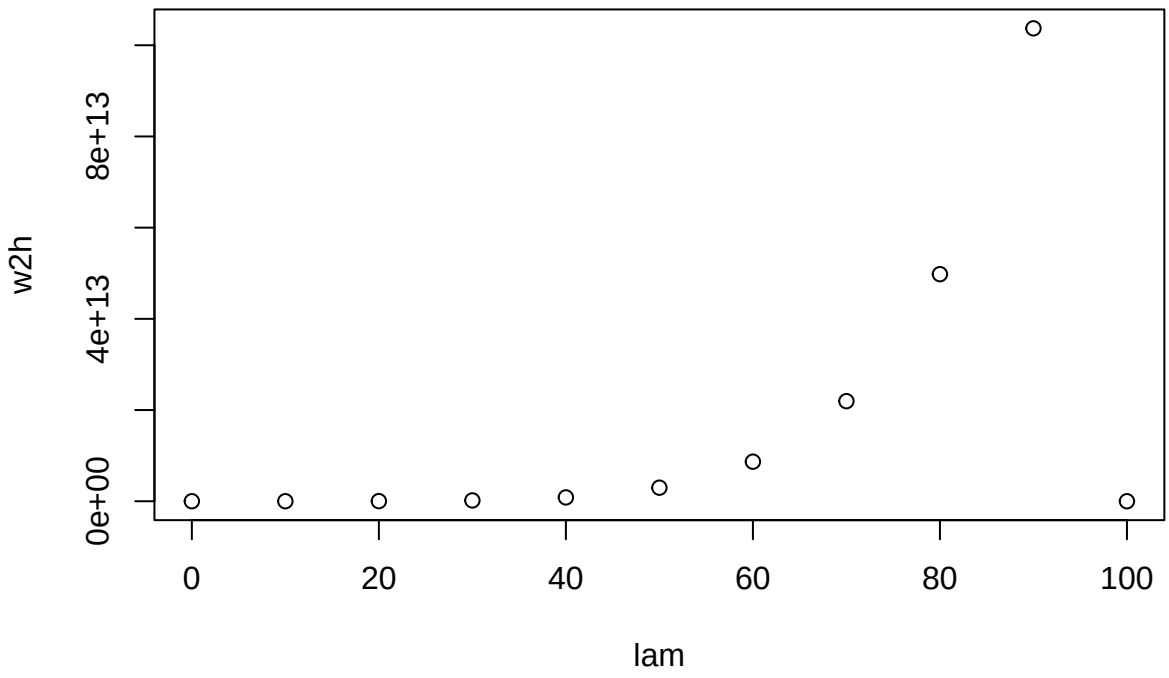
problem 3

c

```
x <- rnorm(100, 0,1)
e <- rnorm(100,0,0.1)

ti <- 1 + 0.2*x -1*x^2 +e


wl1 <- function(lam1,l){
    w1l <- e%*%e+lam1*sqrt(sum(w1*w1))
    w1l
}

wl2 <- function(lam2,l){
   w2l <-  e%*%e+lam2*(sum(w2*w2))
   w2l
}

w1h <- rep(0,11)
w2h <- rep(0,11)
h=0
lam = c(0,10,20,30,40,50,60,70,80,90,100)
for( i in lam){
    w1 = 0.2
    w2 = 1
    for(k in 1: 3){
      w1 <- wl1(lam1 = i,l=k)
      w2 <- wl2(lam2 = i,l=k)
    }
    w1h[h] = w1
    w2h[h] = w2
    h=h+1
}

plot(lam,w1h)
```




```
plot(lam,w2h)
```



d)

w1,2 is increasing when lambda1,2 increasing

problem 6

d)

```
library("readxl")
train<-read_excel("training.xlsx",col_names = c("x1","x2","group"))
fit1 <- glm(group~ x1+x2,data=train)
summary(fit1)
```

```
##
## Call:
## glm(formula = group ~ x1 + x2, data = train)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.76184  -0.15292   0.00143   0.17241   0.63974
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.194053   0.043108  50.897   <2e-16 ***
## x1          -0.214846   0.007725 -27.813   <2e-16 ***
## x2           0.017735   0.014143   1.254    0.211
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.06226525)
##
##     Null deviance: 66.667  on 299  degrees of freedom
## Residual deviance: 18.493  on 297  degrees of freedom
## AIC: 23.442
##
## Number of Fisher Scoring iterations: 2
```

```
fit2 <- glm(group~ x1+x2-1,data = train)
summary(fit2)
```

```
##
## Call:
## glm(formula = group ~ x1 + x2 - 1, data = train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6618  -0.2639   0.1928   0.7601   1.9616
##
## Coefficients:
##    Estimate Std. Error t value Pr(>|t|)
## x1 -0.03808    0.02148  -1.773   0.0772 .
## x2  0.60288    0.02564  23.514   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.6033199)
##
##     Null deviance: 900.00  on 300  degrees of freedom
## Residual deviance: 179.79  on 298  degrees of freedom
## AIC: 703.76
##
## Number of Fisher Scoring iterations: 2
```

e)

from the coding in d), AIC for model 1 is 23.442, AIC for model 2 is 703.76

f)

```
trdata1 <- predict(fit1, newdata = train, type = "response")
trdata2 <- predict(fit2, newdata = train, type = "response")
test<-read_excel("test.xlsx",col_names = c("x1","x2","group"))
```

```
testdata1 <- predict(fit1, newdata = test, type = "response" )
testdata2 <- predict(fit2, newdata = test, type = "response")
## training model 1
glm.pred1=rep(1,300)
glm.pred1[trdata1 >1.5]=2
table(glm.pred1,factor(train$group))
```

```
##
## glm.pred1   1   2
##         1  90   5
##         2  10 195
```

```
## training model 2
glm.pred2=rep(1,300)
glm.pred2[trdata2 >1.5]=2
table(glm.pred2,factor(train$group))
```

```
##
## glm.pred2   1   2
##         1  62 103
##         2  38  97
```

```
## test model 1
glm.test1 = rep(1,40)
glm.test1[testdata1 >1.5]=2
table(glm.test1,factor(test$group))
```

```
##
## glm.test1  1  2
##         1 19  1
##         2  1 19
```

```
## test model 2
glm.test2=rep(1,40)
glm.test2[testdata2 >1.5]=2
table(glm.test2,factor(test$group))
```

```
##
## glm.test2  1  2
##         1 12 12
##         2  8  8
```