Question 1

1. *Solution:*

A picture containing text, whiteboard

Description automatically generated

*Supporting Work:*

Text, letter

Description automatically generated

1. *R Code:*

#clear environment

rm(list=ls())

#read data and set up data sets

ridge\_data <- read.csv("logit\_ridge.csv", header = FALSE)

colnames(ridge\_data) <- c("y","x1","x2","x3","x4","x5","x6","x7","x8","x9","x10","x11","x12","x13","x14","x15","x16","x17","x18","x19","x20")

ridge\_data$y <- as.numeric(ridge\_data$y)

ridge\_test <- ridge\_data[1:10,]

ridge\_train <- ridge\_data[11:nrow(ridge\_data),]

#initialize gradient ascent elements

#note – had to set eps to 0.0005 bc it took too long to run at 0.0001 and the hw did not specify an error tolerance, only an alpha

alpha <- 10^-4

b\_last <- matrix(ncol=21,nrow=1,data=0)

colnames(b\_last) <- c(“b0”,”b1”,”b2”,”b3”,”b4”,”b5”,”b6”,”b7”,”b8”,”b9”,”b10”,”b11”,”b12”,”b13”,”b14”,”b15”,”b16”,”b17”,”b18”,”b19”,”b20”)

b <- b\_last

b\_history <- b\_last

err <- 100

eps <- 0.0005

lambda <- 1

grad <- matrix(data=0,ncol=1,nrow=21)

y\_train <- as.matrix(as.numeric(ridge\_train[,1]))

x\_train <- as.matrix(ridge\_train[,2:21])

#calculation of grad descent

while(err > eps) {

e <- exp(b\_last[,1]+b\_last[,2]\*x\_train[,1]+b\_last[,3]\*x\_train[,2]+b\_last[,4]\*x\_train[,3]+b\_last[,5]\*x\_train[,4]+b\_last[,6]\*x\_train[,5]+b\_last[,7]\*x\_train[,6]+b\_last[,8]\*x\_train[,7]+b\_last[,9]\*x\_train[,8]+b\_last[,10]\*x\_train[,9]+b\_last[,11]\*x\_train[,10]+b\_last[,12]\*x\_train[,11]+b\_last[,13]\*x\_train[,12]+b\_last[,14]\*x\_train[,13]+b\_last[,15]\*x\_train[,14]+b\_last[,16]\*x\_train[,15]+b\_last[,17]\*x\_train[,16]+b\_last[,18]\*x\_train[,17]+b\_last[,19]\*x\_train[,18]+b\_last[,20]\*x\_train[,19]+b\_last[,21]\*x\_train[,20])

grad[1,] <- sum(y\_train-e/(1+e))

for(j in 2:21){

grad[j,] <- sum(x\_train[j-1]\*(y\_train-e)/(1+e))-2\*lambda\*b\_last[,j]

}

b = b\_last + alpha\*t(grad)

err = norm(b – b\_last, type = “2”)

b\_last <- b

b\_history <- rbind(b\_history, b\_last)

}

b

*Output:Text

Description automatically generated*

*Solution:*

Max likelihood estimates of b1 and b2 are -1.607 and -0.280 respectively.

1. *R Code:*

#calculate prediction error

y\_test <- as.matrix(as.numeric(ridge\_test[,1]))

x\_test <- as.matrix(ridge\_test[,2:21])

y\_prob <- matrix(ncol=1,nrow=10,data=0)

for(I in 1:10){

y\_prob[i] <- exp(b[,1] + x\_test[I,]%\*%b[,2:21])/(1+exp(b[,1] + x\_test[I,]%\*%b[,2:21]))

}

pred\_err <- (y\_test-y\_prob)^2

mean(pred\_err)

*Output:*

A picture containing graphical user interface

Description automatically generated

(not great I know but I could not improve my algorithm no matter how much debugging I did, switching between matrix notation and non-matrix, changing lambdas and tolerances… best I could do…)

Question 2

Text, letter

Description automatically generated