Stat 760
Jakob Lovato & Will Bliss
Homework 6

Exercise 1

The log-likelihood function for N observations in the multinomial case is

$$l(\theta) = \sum_{i=1}^{n} \log P_{g_i}(x_i;\theta) \qquad (4.19)$$

Where $p_k = (x_i;\theta) = P(G=k \mid X = x_i;\theta)$

Using 2-class coding, we can say the response $y_i = 1$ when $g_i = 1$ and $y_i = 0$ when $g_i = 2$. Say $p_1(x;\theta) = p(x;\theta)$ and $p_2(x;\theta) = 1 - p(x;\theta)$. Then,

$$l(\beta) = \sum_{i=1}^{N} \{ y_i \beta^T x_i - \log(1 + e^{\beta^T x_i}) \} \qquad (4.20)$$

To maximize log-likelihood, we set

$$\frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^{N} x_i (y_i - p(x;\beta)) = 0 \qquad (4.21)$$

Which is a set of $p+1$ nonlinear equations in β. This requires an expanded input vector X to be of size $N \times (p+1)$.

Following from Newton-Raphson in the binomial case, the algorithm updates the vector of βs by:

$$\beta^{new} = \beta^{old} - \left(\frac{\partial^2 l(\beta)}{\partial\beta\,\partial\beta^T}\right)^{-1} \frac{\partial l(\beta)}{\partial \beta}$$

$$= \beta^{old} + (X^T W X)^{-1} X^T (y - p)$$

$$= (X^T W X)^{-1} X^T W z$$

Where $\dfrac{\partial l(\beta)}{\partial \beta} = X^T (y - p)$

$$\frac{\partial^2 l(\beta)}{\partial \beta} = -X^T W X$$

And z is the weight matrix $X\beta^{old} + W^{-1}(y - p)$. This can be repeated iteratively until β converges.

Exercise 2.

We will have two classes of data:

$$\begin{cases} y_i = 1 & \text{if} \quad x_i \ge x_0 \\ y_i = 0 & \text{if} \quad x_i < x_0 \end{cases}$$

Recall

$$\ell(\beta) = \sum_{i=1}^{N} \{ y_i \beta^T x_i - \log(1 + e^{\beta^T x_i}) \}$$

From exercise 1. Since $x \in \mathbb{R}$, this can be rewritten as

$$\ell(\beta) = \sum_{i=1}^{N} \{ y_i (\beta_0 + \beta_1 x_i) - \log(1 + e^{\beta_0 + \beta_1 x_i}) \}$$

Maximizing by setting the derivative of $\ell(\beta)$ equal to zero yields

$$\frac{\partial \ell(\beta)}{\partial \beta} = \sum_{i=1}^{N} x_i \left(y_i - \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} \right)$$

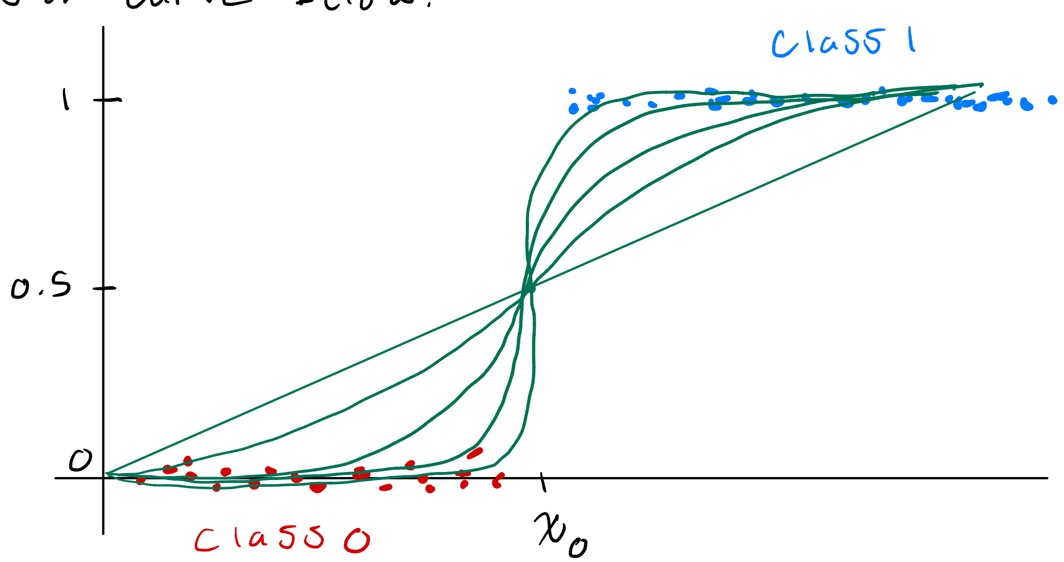
$$= \sum_{x_i \ge x_0} x_i \left(1 - \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} \right) - \sum_{x_i < x_0} x_i \left(\frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} \right)$$

$$= \sum_{x_i \ge x_0} x_i - \sum_{x_i \ge x_0} x_i \left(1 - \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} \right) - \sum_{x_i < x_0} x_i \left(\frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} \right)$$

$$\Rightarrow \sum_{x_i > 0} x_i = \sum_{i=1}^{N} x_i \left(1 - \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} \right) - \sum_{x_i < x_0} x_i \left(\frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} \right)$$

We can see that for any set of x_i's, $\beta = (\beta_0, \beta_1) \to \infty$ and hence cannot have a maximized likelihood.

We can observe this trend visually with the logistic regression curve below.



Stat 760 Homework 6 Question 3

Will Bliss and Jakob Lovato

3/8/2022

```r
#read data
data <- read.delim("/Users/jakoblovato/Desktop/Stat 760/HW 6/SAHeart.txt", header = TRUE, sep = ',')
data <- data[,-1]
#convert categroical data
data$famhist <- as.numeric(data$famhist == "Present")
```

```r
X <- data[,-ncol(data)]
beta0 <- rep(1, nrow(data))
X <- cbind(beta0, X)
X <- as.matrix(X)

Y <- data[,ncol(data)]

beta <- as.matrix(rep(0, ncol(X)), ncol = 1)
p <- rep(1/2, nrow(data))
W <- diag(nrow(data))

#iterate
while(TRUE){
  z <- X %*% beta + solve(W) %*% (Y - p)
  temp <- beta
  beta <- beta + solve(t(X) %*% W %*% X) %*% t(X) %*% (Y - p)
  for(i in 1:nrow(data)){
    p[i] <- exp(t(beta) %*% X[i, ]) / (1 + exp(t(beta) %*% X[i, ]))
  }
  W <- diag(p * (1 - p))
  if(abs(temp - beta) < 0.01){
    break
  }
}

#output coefficients
beta
```

```
##                    [,1]
## beta0     -6.1507208649
## sbp        0.0065040171
## tobacco    0.0793764457
## ldl        0.1739238981
## adiposity  0.0185865682
## famhist    0.9253704194
## typea      0.0395950250
## obesity   -0.0629098693
```

```
## alcohol    0.0001216624
## age        0.0452253496
```

```r
#store the names for later
betanames <- rownames(beta)
```

```r
#bootstrap
M <- 100
N <- nrow(data)
coefs <- list()

for(k in 1:M){
  Xboot <- matrix(rep(0), nrow = N, ncol = ncol(X))
  index <- sample(1:N, N, replace = TRUE)
  Y <- data[index ,ncol(data)]

  for(j in 1:N){
    Xboot[j, ] <- X[index[j],]
  }

  beta <- as.matrix(rep(0, ncol(Xboot)), ncol = 1)
  p <- rep(1/2, nrow(data))
  W <- diag(nrow(data))

  #iterate
  while(TRUE){
    z <- Xboot %*% beta + solve(W) %*% (Y - p)
    temp <- beta
    beta <- beta + solve(t(Xboot) %*% W %*% Xboot) %*% t(Xboot) %*% (Y - p)
    for(i in 1:nrow(data)){
      p[i] <- exp(t(beta) %*% Xboot[i, ]) / (1 + exp(t(beta) %*% Xboot[i, ]))
    }
    W <- diag(p * (1 - p))
    if(abs(temp - beta) < 0.01){
      break
    }
  }

  coefs[[k]] <- beta
}

means <- c()
vars <- c()
for(i in 1:10){
  means <- c(means, mean(unlist(lapply(coefs, '[[', i))))
  vars <- c(vars, var(unlist(lapply(coefs, '[[', i))))
}
means <- data.frame(means)
vars <- data.frame(vars)
rownames(means) <- betanames
rownames(vars) <- betanames

#round to avoid scientific notation
round(means, 7)
```

```
##                means
## beta0     -6.2862752
## sbp        0.0068853
## tobacco    0.0841379
## ldl        0.1754985
## adiposity  0.0173020
## famhist    0.9421337
## typea      0.0423831
## obesity   -0.0659686
## alcohol    0.0002619
## age        0.0455974
```

```r
round(vars, 7)
```

```
##                vars
## beta0     1.6599122
## sbp       0.0000336
## tobacco   0.0006416
## ldl       0.0038941
## adiposity 0.0009445
## famhist   0.0540465
## typea     0.0001559
## obesity   0.0023934
## alcohol   0.0000232
## age       0.0001643
```