homework-3

```r
library(bis557)
```

#1.CASL 5.8 Exercises problem number 2

The Hessian matrix can be written as $H(l) = X^T DX$, where D is a diagonal matrix with elements

$$D_{i,i} = p_i(1 - p_i)$$

The textbook CASL shows that in order to make the logistic Hessian ill-conditioned, the probablity p should close to 0 or 1.

```r
set.seed(100)
n <- 100; m <- 25
X <- cbind(1, matrix(rnorm(n * (m-1)), ncol = m-1))
H_linear <- t(X) %*% X
svals <- svd(H_linear)$d
max(svals) / min(svals) #the condition number can be computed from values
#> [1] 6.371581

beta <- rep(10, m)
p <- 1 / (1 + exp(-X %*% beta)) #p is either close to 0 or close to 1
D <- diag(as.vector(p) * (1 - as.vector(p)))
H_logistic <- t(X) %*% D %*% X
svals1 <- svd(H_logistic)$d
max(svals1) / min(svals1)
#> [1] 541745.6
```

The condition number of the linear Hessian is small, so it is well-conditioned. The condition number of the logistic Hessian is large, so it is ill-conditioned.

#2 I create a function that provides first-order solution for the GLM maximum likelihood problem using gradient information,and I can choose to use either a constant step size or a momentum method to optimaize the parameter.

```r
set.seed(999)
n <- 1000; p <- 5;
X <- cbind(1, matrix(rnorm(n * (p-1)), ncol = p-1))
beta <- c(-1, 0.3, 2, 0.1, 0.5)
Y <- rpois(n, exp(X %*% beta))
data <- as.data.frame(cbind(Y, X))
fit_glm <- glm(Y ~ . -1, data, family = poisson(link = "log"))
fit_cos <-glm_gradient(X,Y,family=poisson(link = "log"),step = "constant")
fit_mom <-glm_gradient(X,Y,family=poisson(link = "log"),step = "momentum")
fit_glm$coefficients
#>          V2          V3          V4          V5          V6
#> -0.99522326  0.32790708  2.00796080  0.07941207  0.49548509
fit_cos
#> $coefficients
#>             [,1]
```

```
#> [1,] -0.97142587
#> [2,]  0.32368153
#> [3,]  1.99735203
#> [4,]  0.07910379
#> [5,]  0.49240947
fit_mom
#> $coefficients
#>             [,1]
#> [1,] -0.40957879
#> [2,]  0.23161214
#> [3,]  1.73891071
#> [4,]  0.07170525
#> [5,]  0.42120510
```

Compared to the perfomance with momentum method, the performance with a constant step size is better since the estimated coefficients are more similar to the glm method and the true beta.

#3
I create a classification model generalizing logistic regression to accommodate more than two classes. I use a second-order solution (Hessian matrix) to optimaize the parameter. Then I use softmax function to get the probablity of each observation belongs to three classes and assign the obersvation to the class with the highest probability among its three probabilities.

```
data(iris)
iris1 <- iris[sample(150,replace = F),]
iris.X <- as.matrix(iris1[,-5])
iris.y <- iris1$Species
classification <- class_logistic(iris.X, iris.y, maxiter=50)
length(iris.y[(iris.y != classification$prediction)==TRUE])/length(iris.y)
#> [1] 0.04
```

The misclassification rate is 0.04. Our classification is somehow accuracy.