HW3-GLM-LogitClassification

```r
library(bis557)
#> Loading required package: doParallel
#> Loading required package: foreach
#> Loading required package: iterators
#> Loading required package: parallel
#> Loading required package: ggplot2
#> Loading required package: purrr
#>
#> Attaching package: 'purrr'
#> The following objects are masked from 'package:foreach':
#>
#>     accumulate, when
#> Loading required package: rsample
#> Loading required package: tibble
#> Warning: replacing previous import 'foreach::when' by 'purrr::when' when loading
#> 'bis557'
#> Warning: replacing previous import 'foreach::accumulate' by 'purrr::accumulate'
#> when loading 'bis557'
library(palmerpenguins)
```

1. CASL 5.8 Exercise number 2. Generate a matrix X and probabilities p such that the linear Hessian $(X^t X)$ is well-conditioned but the logistic variation is not.

As we learned from the lecture and text book, when the probabilities become very close to zero or one, the logistic variation would be ill-conditioned. When p or (1-p) is too small, $X^t DX$ becomes very small and makes the inverse matrix too big and is bad for convergence. Given:

$$H(l) = X^t \cdot D \cdot X$$

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

```r
# example

beta = matrix(rep(0,5),5)

beta_old <- beta
p <- runif(10000,0,0.0000001)
Y <- rbinom(10000,1,p)
X <- matrix(rnorm(50000),10000)

D <- matrix(rep(0,10000^2),10000)
diag(D) <- p*(1-p)

#X
XtDX <- t(X)%*%D%*%X # small
head(XtDX)
```

```
#>               [,1]          [,2]          [,3]         [,4]          [,5]
#> [1,]  4.976206e-04 -3.114382e-06  9.822194e-06 7.545930e-06  4.739811e-07
#> [2,] -3.114382e-06  4.854818e-04  7.844031e-06 1.466690e-06 -5.878301e-06
#> [3,]  9.822194e-06  7.844031e-06  4.888099e-04 1.350984e-05 -4.091128e-06
#> [4,]  7.545930e-06  1.466690e-06  1.350984e-05 4.944152e-04  1.619733e-05
#> [5,]  4.739811e-07 -5.878301e-06 -4.091128e-06 1.619733e-05  4.997250e-04
XtX <- t(X)%*%X
head(XtX)
#>             [,1]       [,2]        [,3]      [,4]        [,5]
#> [1,] 10102.56842  -36.70522   187.84845  178.9823   107.81256
#> [2,]   -36.70522 9941.13873   123.96224  109.7524   -96.11722
#> [3,]   187.84845  123.96224 10069.38069  159.6297   -47.73439
#> [4,]   178.98226  109.75240   159.62972 9971.1266   241.54256
#> [5,]   107.81256  -96.11722   -47.73439  241.5426 10018.12520

inv_H_v <- -solve(XtDX) # absolute value is too big and is bad for convergence
inv_H_v
#>              [,1]         [,2]        [,3]         [,4]         [,5]
#> [1,] -2010.884202   -13.619177    39.81649    29.606731     1.113428
#> [2,]   -13.619177 -2060.740670    32.96890     6.211674   -24.159150
#> [3,]    39.816493    32.968902 -2048.81107    55.875412   -18.234134
#> [4,]    29.606731     6.211674    55.87541 -2026.757201    66.194683
#> [5,]     1.113428   -24.159150   -18.23413    66.194683 -2003.680609
inv_H <- -solve(XtX)
inv_H #Small and is good for convergence
#>               [,1]          [,2]          [,3]          [,4]          [,5]
#> [1,] -9.906181e-05 -3.977149e-07  1.830416e-06  1.728305e-06  1.029314e-06
#> [2,] -3.977149e-07 -1.006307e-04  1.223869e-06  1.118986e-06 -9.823519e-07
#> [3,]  1.830416e-06  1.223869e-06 -9.938734e-05  1.557358e-06 -5.190661e-07
#> [4,]  1.728305e-06  1.118986e-06  1.557358e-06 -1.004165e-04  2.420654e-06
#> [5,]  1.029314e-06 -9.823519e-07 -5.190661e-07  2.420654e-06 -9.990041e-05
```

2. Describe and implement a first-order solution for the GLM maximum likelihood problem using only gradient information, avoiding the Hessian matrix. Include both a constant step size along with an adaptive one. You may use a standard adaptive update Momentum, Nesterov, AdaGrad, Adam, or your own. Make sure to explain your approach and compare it's performance with a constant step size.

I used frist-order gradient descent with momentum adaptive update. Compare to the constant step size, the adaptive update is faster in converging because it helps accelerate gradient vectors in the right rirection. When chosing a bad step size, the constant step size may not work but the adaptive update may work (accep a greater range of step sizes).

```
set.seed(10)
# Momentum Algorithm
X <- cbind(rep(1,100),matrix(rnorm(1000),200))
##poisson simulation
beta <- c(1, 0.1, 0.2, 0.1, 0.3, -1)
y <- rpois(nrow(X), exp(X%*%beta))
data <- as.data.frame(cbind(y,X[,-1]))

glm(y~.,data,family = poisson)
#>
```

```
#> Call:  glm(formula = y ~ ., family = poisson, data = data)
#> 
#> Coefficients:
#> (Intercept)           V2           V3           V4           V5           V6  
#>     0.91536      0.08518      0.26944      0.14617      0.30211     -0.99768  
#> 
#> Degrees of Freedom: 199 Total (i.e. Null);  194 Residual
#> Null Deviance:       986.8 
#> Residual Deviance: 223.9     AIC: 717

# with adaptive step size update
gradient_descent_mmt(y,X,family = poisson(link = "log"),update = T)
#> $coefficients
#> [1]  0.92292367  0.08504586  0.26621883  0.14487870  0.30355635 -1.00574020
# constant step size
gradient_descent_mmt(y,X,family = poisson(link = "log"),update = F)
#> $coefficients
#> [1]  0.91535819  0.08517696  0.26944273  0.14616556  0.30210794 -0.99767596
```

3. Describe and implement a classification model generalizing logistic regression to accommodate more than two classes.

Change y into rows of binary (3 classes = 3 rows) classes and then apply a regular logistic model to each line of data. This method gives coefficients as a matrix, make predictions, and calculates am error rate.

```
data("penguins")
X = penguins[-which(is.na(penguins[,c(3,4,5,6)])),c(3,4,5,6)]
#> Warning: The `i` argument of ``[.tbl_df`()` must lie in [-rows, 0] if negative, as of tibble
  3.0.0.
#> Use `NA_integer_` as row index to obtain a row full of `NA` values.
#> This warning is displayed once every 8 hours.
#> Call `lifecycle::last_warnings()` to see where this warning was generated.
X = cbind(1,scale(X))
y = (unlist(penguins[-which(is.na(penguins[,c(3,4,5,6)])),1]))
# y = (penguins$species=="Adelie")
data = cbind(X,y)
logit_multiclass(X,y)
#> $coefficients
#>                        bill_length_mm bill_depth_mm flipper_length_mm
#> is Adelie    -6.605742   -27.88142422    17.6819179         0.7378325
#> is Gentoo    -8.897498     0.06448071   -11.0255843         8.1020486
#> is Chinstrap -9.613699    15.22723862     0.2107683        -2.8183472
#>              body_mass_g
#> is Adelie       5.037175
#> is Gentoo       6.883006
#> is Chinstrap  -12.276150
#> 
#> $predict
#>   [1] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie   
#>   [8] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie   
#>  [15] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie   
#>  [22] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie   
#>  [29] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie   
#>  [36] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie   
```

```
#>  [43] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
#>  [50] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
#>  [57] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
#>  [64] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
#>  [71] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
#>  [78] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
#>  [85] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
#>  [92] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
#>  [99] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
#> [106] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
#> [113] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
#> [120] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
#> [127] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
#> [134] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
#> [141] Adelie    Adelie    Adelie    Adelie    Adelie    Adelie    Adelie
#> [148] Adelie    Adelie    Adelie    Adelie    Gentoo    Gentoo    Gentoo
#> [155] Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo
#> [162] Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo
#> [169] Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo
#> [176] Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo
#> [183] Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo
#> [190] Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo
#> [197] Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo
#> [204] Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo
#> [211] Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo
#> [218] Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo
#> [225] Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo
#> [232] Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo
#> [239] Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo
#> [246] Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo
#> [253] Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo
#> [260] Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo
#> [267] Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo    Gentoo
#> [274] Gentoo    Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
#> [281] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
#> [288] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
#> [295] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
#> [302] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
#> [309] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
#> [316] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
#> [323] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
#> [330] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
#> [337] Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap Chinstrap
#> Levels: Adelie Chinstrap Gentoo
#>
#> $error
#> [1] 0
```