HW7

Hanao Li

October 25, 2019

Problem 1

a)

```
if(!require("pacman")) install.packages("pacman")
```

```
## Loading required package: pacman
```

```
p_load(glmnet, penalized)

data <- ChickWeight
data$weight_group[ChickWeight$weight > 240] <- 1
data$weight_group[ChickWeight$weight <= 240] <- 0

day21 <- subset(data[which(data$Time == 21) ,])
day21 <- rbind(subset(day21[which(day21$Diet == 1),]), subset(day21[which(day21$Diet == 4),]))
day21$diet_group[day21$Diet == 1] <- 0
day21$diet_group[day21$Diet == 4] <- 1
fit <- glm(weight_group ~ diet_group, family = "binomial", data=day21)
summary(fit)
```

```
##
## Call:
## glm(formula = weight_group ~ diet_group, family = "binomial",
##     data = day21)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9005  -0.5168  -0.5168  -0.5168   2.0393
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.9459     0.7559  -2.574    0.010 *
## diet_group    1.2528     1.0351   1.210    0.226
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 25.020  on 24  degrees of freedom
## Residual deviance: 23.514  on 23  degrees of freedom
## AIC: 27.514
##
## Number of Fisher Scoring iterations: 4
```

```
# From the result, we could see that we have our logistic model as logit = -1.9459 + 1.2528 * Di
et_group.  beta_0 denotes the log odds ratio of Weight > 240 for Diet Group 1, the odds ratio of
Weight > 240 for group 1 is exp(-1.9459). beta_1 denotes the log odds ratio of weight > 240 for
 diet group 4 relative to the group 1. The odds ratio of weight > 240 for group 4 is exp(1.252
8). The odds ratio of weighh > 240 for group 1 is exp(-1.9459 + 1.2528).
# Since p values are both larger than 0.05, we do not reject the null and we conclud there is no
significance association between group 1 and group 4.
```

b)

```
BirthWeight<-subset(data,select=c(weight,Chick),Time==0)
names(BirthWeight)[names(BirthWeight)=="weight"]<-"BirthWeight"
dayadjust<-merge(day21,BirthWeight,by.y = "Chick")

fitadjust <- glm(weight_group ~ diet_group + BirthWeight, "binomial", dayadjust)
summary(dayadjust)
```

```
##      Chick        weight           Time    Diet    weight_group
##  13     : 1   Min.   : 96.0   Min.   :21   1:16   Min.   :0.0
##  9      : 1   1st Qu.:157.0   1st Qu.:21   2: 0   1st Qu.:0.0
##  20     : 1   Median :204.0   Median :21   3: 0   Median :0.0
##  10     : 1   Mean   :199.6   Mean   :21   4: 9   Mean   :0.2
##  17     : 1   3rd Qu.:237.0   3rd Qu.:21          3rd Qu.:0.0
##  19     : 1   Max.   :322.0   Max.   :21          Max.   :1.0
##  (Other):19
##    diet_group    BirthWeight
##  Min.   :0.00   Min.   :39.00
##  1st Qu.:0.00   1st Qu.:41.00
##  Median :0.00   Median :41.00
##  Mean   :0.36   Mean   :41.32
##  3rd Qu.:1.00   3rd Qu.:42.00
##  Max.   :1.00   Max.   :43.00
##
```

```
# From the result, we could see that we have our logistic model as y = 21.836 + 0.9053 * Diet_gr
oup - 0.5743 * BirthWeight. beta_0 denotes the log odds ratio of Weight > 240 for Diet Group 1,
 the odds ratio of Weight > 240 for group 1 is exp(21.836-0.5743 * BirthWeight). beta_1 denotes
 the log odds ratio of weight > 240 for diet group 4 relative to the group 1. The odds ratio of
 weight > 240 for group 4 is exp(21.836+0.9053 - 0.5743*BirthWeight).
# Since p values are both larger than 0.05, we do not reject the null and we conclud there is no
significance association between group 1 and group 4 with adjusting for BirthWeight.
```

Problem 2

a)

```
dayall <- subset(data[which(data$Time == 21) ,])
dayall$Group1 <- rep(0,dim(dayall)[1])
dayall$Group1[dayall$Diet==1] <- 1
dayall$Group2 <- rep(0,dim(dayall)[1])
dayall$Group2[dayall$Diet==2] <- 1
dayall$Group3 <- rep(0,dim(dayall)[1])
dayall$Group3[dayall$Diet==3] <- 1
fitall <- glm(weight_group ~ Group1 + Group2 + Group3, "binomial", dayall)
summary(fitall)
```

```
##
## Call:
## glm(formula = weight_group ~ Group1 + Group2 + Group3, family = "binomial",
##     data = dayall)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5518  -0.9005  -0.5168   0.8446   2.0393
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.6931     0.7071  -0.980    0.327
## Group1       -1.2528     1.0351  -1.210    0.226
## Group2        0.2877     0.9574   0.300    0.764
## Group3        1.5404     0.9880   1.559    0.119
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 58.574  on 44  degrees of freedom
## Residual deviance: 49.191  on 41  degrees of freedom
## AIC: 57.191
##
## Number of Fisher Scoring iterations: 4
```

```
# From the result, we could see that we have our logistic model as y = -0.6931 -1.2528 * Group1
 + 0.2877 * Group2 + 1.5404 * Group3. beta_0 denotes the log odds ratio of Weight > 240 for Diet
Group 4, the odds ratio of Weight > 240 for group 4 is exp(-0.6931). beta_1 denotes the log odds
ratio of weight > 240 for diet group 1 relative to the group 4. The odds ratio of weight > 240 f
or group 1 is exp(-0.6931-1.2528).beta_2 denotes the log odds ratio of weight > 240 for diet gro
up 2 relative to the group 4. The odds ratio of weight > 240 for group 2 is exp(-0.6931+0.2877).
beta_3 denotes the log odds ratio of weight > 240 for diet group 3 relative to the group 4. The
 odds ratio of weight > 240 for group 3 is exp(-0.6931+1.5404).
# Since p values are both larger than 0.05, we do not reject the null and we conclud there is no
significance association between group 1, group 2, group 3 and group 4 without adjusting for Bir
thWeight.
```

b)

```
dayalladj <- merge(dayall,BirthWeight,by.y = "Chick")
fitadjall <- glm(weight_group ~ Group1+Group2+Group3+BirthWeight,family="binomial",data=dayallad
j)
summary(fitadjall)
```

```
##
## Call:
## glm(formula = weight_group ~ Group1 + Group2 + Group3 + BirthWeight,
##     family = "binomial", data = dayalladj)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5672  -0.8997  -0.5148   0.8459   2.0348
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.08183   12.88220   0.006    0.995
## Group1      -1.24004    1.05632  -1.174    0.240
## Group2       0.28409    0.95936   0.296    0.767
## Group3       1.53889    0.98834   1.557    0.119
## BirthWeight -0.01895    0.31463  -0.060    0.952
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 58.574  on 44  degrees of freedom
## Residual deviance: 49.188  on 40  degrees of freedom
## AIC: 59.188
##
## Number of Fisher Scoring iterations: 4
```

```
# From the result, we could see that we have our logistic model as y = 0.08183 -1.2404 * Group1
 + 0.28409 * Group2 + 1.53889 * Group3 - 0.01895 * BirthWeight. beta_0 denotes the log odds rati
o of Weight > 240 for Diet Group 4, the odds ratio of Weight > 240 for group 4 is exp(0.08183 -
 0.01895 * BirthWeight). beta_1 denotes the log odds ratio of weight > 240 for diet group 1 rela
tive to the group 4. The odds ratio of weight > 240 for group 1 is exp(0.08183-1.2404-0.01895*Bi
rthWeight).beta_2 denotes the log odds ratio of weight > 240 for diet group 2 relative to the gr
oup 4. The odds ratio of weight > 240 for group 2 is exp(0.08183+0.28409-0.01895*BirthWeight). b
eta_3 denotes the log odds ratio of weight > 240 for diet group 3 relative to the group 4. The o
dds ratio of weight > 240 for group 3 is exp(0.08183+1.53889-0.01895*BirthWeight).
# Since p values are both larger than 0.05, we do not reject the null and we conclud there is no
significance association between group 1, group 2, group 3 and group 4 with adjusting for BirthW
eight.
```

Problem 3

```
fitl1 <- optL1(dayadjust$weight_group, penalized = -as.matrix(dayadjust$diet_group), fold = 5, m
odel = "logistic")
```

```
## lambda= 0.6684405    1□2□3□4□5□cvl= -16.22268
## lambda= 1.081559     1□2□3□4□5□cvl= -15.72875
## lambda= 1.336881     1□2□3□4□5□cvl= -15.43013
## lambda= 1.494678     1□2□3□4□5□cvl= -15.25894
## lambda= 1.592203     1□2□3□4□5□cvl= -15.15735
## lambda= 1.652476     1□2□3□4□5□cvl= -15.09609
## lambda= 1.689727     1□2□3□4□5□cvl= -15.05879
## lambda= 1.712749     1□2□3□4□5□cvl= -15.03595
## lambda= 1.726978     1□2□3□4□5□cvl= -15.02191
## lambda= 1.735771     1□2□3□4□5□cvl= -15.01326
## lambda= 1.741206     1□2□3□4□5□cvl= -15.00793
## lambda= 1.744565     1□2□3□4□5□cvl= -15.00464
## lambda= 1.746641     1□2□3□4□5□cvl= -15.00261
## lambda= 1.747924     1□2□3□4□5□cvl= -15.00135
## lambda= 1.748717     1□2□3□4□5□cvl= -15.00058
## lambda= 1.749207     1□2□3□4□5□cvl= -15.0001
## lambda= 1.74951  1□2□3□4□5□cvl= -14.9998
## lambda= 1.749697     1□2□3□4□5□cvl= -14.99962
```

```
fitl1$fullfit
```

```
## Penalized logistic regression object
## 2 regression coefficients of which 1 are non-zero
##
## Loglikelihood =   -12.51006
## L1 penalty =  0  at lambda1 =  1.749697
```

```
fitl1_adj <- optL1(dayadjust$weight_group, penalized = -dayadjust$BirthWeight + as.matrix(dayadj
ust$diet_group), fold = 5, model = "logistic")
```

```
## lambda= 1.432373     1□2□3□4□5□cvl= -13.91647
## lambda= 2.317627     1□2□3□4□5□cvl= -13.81029
## lambda= 2.864745     1□2□3□4□5□cvl= -13.85336
## lambda= 2.30739  1□2□3□4□5□cvl= -13.81029
## lambda= 2.309537     1□2□3□4□5□cvl= -13.81029
## lambda= 2.309484     1□2□3□4□5□cvl= -13.81029
## lambda= 2.308463     1□2□3□4□5□cvl= -13.81029
## lambda= 2.309094     1□2□3□4□5□cvl= -13.81029
```

```
fitl1_adj$fullfit
```

```
## Penalized logistic regression object
## 2 regression coefficients
##
## Loglikelihood =   -11.74999
## L1 penalty =  0.5743996  at lambda1 =  2.309484
```