Logistic Regression with caret

Max Kuhn
Michael Weylandt
Kenny Bozarth
Jason Hughes

Table of Contents

1 Introduction
2 Preliminaries
2.1 Data
2.2 Loading the data into memory
2.3 Data Splitting
2.4 Preprocessing
3 Model Building
3.1 Model Fitting
3.2 Model Prediction
3.3 Model Performance and Feature Selection
3.4 Tuning Model Parameters
3.5 Model Comparison
3.6 Fitting Other Models with Caret
4 Model Performance Differences Due to Data Splitting
4.1 Model Fitting
4.2 Model Prediction
4.3 Model Performance
4.4 Model Comparison
5 Cross-Validation Resampling Profiles
5.1 Model Fitting
5.2 Model Prediction
5.3 Model Performance
5.4 Model Comparison
6 Bootstrap Resampling Profiles
6.1 Model Fitting
6.2 Model Prediction
6.3 Model Performance
6.4 Model Comparison
7 Conclusions
7.1 Session Info
7.2 Bibliography

1 Introduction

2 Preliminaries

2.1 Data

```
data(BinomialDat)
str(Fatality, vec.len = 2, strict.width = "cut")
## 'data.frac':  337 obs. of  7 variables:
##  $ Age     : int  21 19
##  $ Weight  : int  117 148
##  $ Drinks  : int  3 8
##  $ Beltuse : int  0 0
##  $ Female  : int  0 0
##  $ Restraint: int  1 1
##  $ Fatality : int  0 0
```

2.2 Loading the data into memory

```
set.seed(1)
trainX <- Fatality[,c("Age","Weight","Drinks","Beltuse","Female","Restraint")]
trainY <- Fatality$Fatality
table(trainY)
## trainY
##   0   1
## 153 184
```

2.3 Data Splitting

```
trainY <- factor(paste("Class", trainY, sep = "_"))
table(trainY)
## trainY
## Class_0 Class_1
## 153 184
```

2.4 Preprocessing

```
nzv <- nearZero(trainX)
if(length(nzv) > 0) trainX <- trainX[,-nzv]
dim(trainX)
## [1] 337   6
```

3 Model Building

3.1 Model Fitting

```
ctrl <- trainControl(method = "cv",
                     classProbs = TRUE,
                     summaryFunction = twoClassSummary)
set.seed(2)
glmFit <- train(trainX,
                trainY,
                method = "glm",
                family = binomial(),
                metric = "ROC",
                trControl = ctrl)
glmFit
## Generalized Linear Model
##
## 337 samples
##   6 predictors
##   2 classes: 'Class_0', 'Class_1'
##
## No pre-processing
## Resampling: Cross-Validation (10 fold)
## Summary of sample sizes: 304, 304, 304, 304, 304, 304, ...
## Resampling results:
##
##   ROC  Sensitivity  Specificity
##   0.9  0.8          0.8
```

3.2 Model Prediction

```
predictTest <- predict(glmFit, newdata = trainX)
table(predictTest, trainY)
##         trainY
## predictTest Class_0 Class_1
##   Class_0     123      36
##   Class_1      30     148
```

3.3 Model Performance and Feature Selection

```
summary(glmFit, cor = TRUE)
##
## Call:
## NULL
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3000  -0.5000   0.0000   0.5000   2.3000
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.0000     0.2000  -5.00    0.00
## Age          -1.0000     0.2000  -5.00    0.00
## Weight       -1.0000     0.2000  -5.00    0.00
## Drinks       -1.0000     0.2000  -5.00    0.00
## Beltuse      -1.0000     0.2000  -5.00    0.00
## Female       -1.0000     0.2000  -5.00    0.00
## Restraint    -1.0000     0.2000  -5.00    0.00
##
## Correlation of Coefficients:
##             (Intercept) Age  Weight Drinks Beltuse Female
## Age          0.          1.  0.     0.     0.     0.
## Weight       0.          0.  1.     0.     0.     0.
## Drinks       0.          0.  0.     1.     0.     0.
## Beltuse      0.          0.  0.     0.     1.     0.
## Female       0.          0.  0.     0.     0.     1.
## Restraint    0.          0.  0.     0.     0.     0.
##             Restraint
## Age          0.
## Weight       0.
## Drinks       0.
## Beltuse      0.
## Female       0.
## Restraint    1.
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 0.  on 336  degrees of freedom
## Residual deviance: 0.  on 330  degrees of freedom
## AIC: 14.
##
## Number of Fisher Scoring iterations: 4
```

3.4 Tuning Model Parameters

```
set.seed(3)
glmFit2 <- train(trainX,
                 trainY,
                 method = "glm",
                 family = binomial(),
                 preProcess = c("pca"),
                 metric = "ROC",
                 trControl = ctrl)
glmFit2
## Generalized Linear Model
##
## 337 samples
##   6 predictors
##   2 classes: 'Class_0', 'Class_1'
##
## Pre-processing: principal component signal extraction (6), scaled (6), centered (6)
## Resampling: Cross-Validation (10 fold)
## Summary of sample sizes: 304, 304, 304, 304, 304, 304, ...
## Resampling results:
##
##   ROC  Sensitivity  Specificity
##   0.9  0.8          0.8
```

3.5 Model Comparison

```
resamp <- resamples(list(glmFit = glmFit, glmFit2 = glmFit2))
summary(resamp)
##
## Call:
## summary.resamples(object = resamp)
##
## Models: glmFit, glmFit2
## Number of resamples: 10
##
## ROC
##          Min
## glmFit   0.9
## glmFit2  0.9
##          1st Qu.
## glmFit   0.9
## glmFit2  0.9
##          Median
## glmFit   0.9
## glmFit2  0.9
##          Mean
## glmFit   0.9
## glmFit2  0.9
##          3rd Qu.
## glmFit   0.9
## glmFit2  0.9
##          Max
## glmFit   0.9
## glmFit2  0.9
## NA's
## glmFit   0
## glmFit2  0
##
## Sensitivity
##          Min
## glmFit   0.8
## glmFit2  0.8
##          1st Qu.
## glmFit   0.8
## glmFit2  0.8
##          Median
## glmFit   0.8
## glmFit2  0.8
##          Mean
## glmFit   0.8
## glmFit2  0.8
##          3rd Qu.
## glmFit   0.8
## glmFit2  0.8
##          Max
## glmFit   0.8
## glmFit2  0.8
## NA's
## glmFit   0
## glmFit2  0
##
## Specificity
##          Min
## glmFit   0.8
## glmFit2  0.8
##          1st Qu.
## glmFit   0.8
## glmFit2  0.8
##          Median
## glmFit   0.8
## glmFit2  0.8
##          Mean
## glmFit   0.8
## glmFit2  0.8
##          3rd Qu.
## glmFit   0.8
## glmFit2  0.8
##          Max
## glmFit   0.8
## glmFit2  0.8
## NA's
## glmFit   0
## glmFit2  0
```

3.6 Fitting Other Models with Caret

```
set.seed(4)
plsFit <- train(trainX,
                trainY,
                method = "pls",
                tuneLength = 6,
                preProcess = c("pca"),
                metric = "ROC",
                trControl = ctrl)
plsFit
## Partial Least Squares
##
## 337 samples
##   6 predictors
##   2 classes: 'Class_0', 'Class_1'
##
## Pre-processing: principal component signal extraction (6), scaled (6), centered (6)
## Resampling: Cross-Validation (10 fold)
## Summary of sample sizes: 304, 304, 304, 304, 304, 304, ...
## Resampling results across tuning parameters:
##
##   ncomp  ROC  Sensitivity  Specificity
##   1      0.9  0.8          0.8
##   2      0.9  0.8          0.8
##   3      0.9  0.8          0.8
##   4      0.9  0.8          0.8
##   5      0.9  0.8          0.8
##   6      0.9  0.8          0.8
##
## ROC was used to select the optimal model using  the largest value.
## The final value used for the model was ncomp = 1.
```

4 Model Performance Differences Due to Data Splitting

4.1 Model Fitting

```
set.seed(5)
glmFit3 <- train(trainX,
                 trainY,
                 method = "glm",
                 family = binomial(),
                 metric = "ROC",
                 trControl = ctrl)
glmFit3
## Generalized Linear Model
##
## 337 samples
##   6 predictors
##   2 classes: 'Class_0', 'Class_1'
##
## No pre-processing
## Resampling: Cross-Validation (10 fold)
## Summary of sample sizes: 304, 304, 304, 304, 304, 304, ...
## Resampling results:
##
##   ROC  Sensitivity  Specificity
##   0.9  0.8          0.8
```

4.2 Model Prediction

```
predictTest <- predict(glmFit3, newdata = trainX)
table(predictTest, trainY)
##         trainY
## predictTest Class_0 Class_1
##   Class_0     123      36
##   Class_1      30     148
```

4.3 Model Performance

```
summary(glmFit3, cor = TRUE)
##
## Call:
## NULL
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3000  -0.5000   0.0000   0.5000   2.3000
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.0000     0.2000  -5.00    0.00
## Age          -1.0000     0.2000  -5.00    0.00
## Weight       -1.0000     0.2000  -5.00    0.00
## Drinks       -1.0000     0.2000  -5.00    0.00
## Beltuse      -1.0000     0.2000  -5.00    0.00
## Female       -1.0000     0.2000  -5.00    0.00
## Restraint    -1.0000     0.2000  -5.00    0.00
##
## Correlation of Coefficients:
##             (Intercept) Age  Weight Drinks Beltuse Female
## Age          0.          1.  0.     0.     0.     0.
## Weight       0.          0.  1.     0.     0.     0.
## Drinks       0.          0.  0.     1.     0.     0.
## Beltuse      0.          0.  0.     0.     1.     0.
## Female       0.          0.  0.     0.     0.     1.
## Restraint    0.          0.  0.     0.     0.     0.
##             Restraint
## Age          0.
## Weight       0.
## Drinks       0.
## Beltuse      0.
## Female       0.
## Restraint    1.
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 0.  on 336  degrees of freedom
## Residual deviance: 0.  on 330  degrees of freedom
## AIC: 14.
##
## Number of Fisher Scoring iterations: 4
```

4.4 Model Comparison

```
resamp <- resamples(list(glmFit = glmFit, glmFit3 = glmFit3))
summary(resamp)
##
## Call:
## summary.resamples(object = resamp)
##
## Models: glmFit, glmFit3
## Number of resamples: 10
##
## ROC
##          Min
## glmFit   0.9
## glmFit3  0.9
##          1st Qu.
## glmFit   0.9
## glmFit3  0.9
##          Median
## glmFit   0.9
## glmFit3  0.9
##          Mean
## glmFit   0.9
## glmFit3  0.9
##          3rd Qu.
## glmFit   0.9
## glmFit3  0.9
##          Max
## glmFit   0.9
## glmFit3  0.9
## NA's
## glmFit   0
## glmFit3  0
##
## Sensitivity
##          Min
## glmFit   0.8
## glmFit3  0.8
##          1st Qu.
## glmFit   0.8
## glmFit3  0.8
##          Median
## glmFit   0.8
## glmFit3  0.8
##          Mean
## glmFit   0.8
## glmFit3  0.8
##          3rd Qu.
## glmFit   0.8
## glmFit3  0.8
##          Max
## glmFit   0.8
## glmFit3  0.8
## NA's
## glmFit   0
## glmFit3  0
##
## Specificity
##          Min
## glmFit   0.8
## glmFit3  0.8
##          1st Qu.
## glmFit   0.8
## glmFit3  0.8
##          Median
## glmFit   0.8
## glmFit3  0.8
##          Mean
## glmFit   0.8
## glmFit3  0.8
##          3rd Qu.
## glmFit   0.8
## glmFit3  0.8
##          Max
## glmFit   0.8
## glmFit3  0.8
## NA's
## glmFit   0
## glmFit3  0
```

5 Cross-Validation Resampling Profiles

5.1 Model Fitting

```
set.seed(6)
glmFit4 <- train(trainX,
                 trainY,
                 method = "glm",
                 family = binomial(),
                 metric = "ROC",
                 trControl = ctrl)
glmFit4
## Generalized Linear Model
##
## 337 samples
##   6 predictors
##   2 classes: 'Class_0', 'Class_1'
##
## No pre-processing
## Resampling: Cross-Validation (10 fold)
## Summary of sample sizes: 304, 304, 304, 304, 304, 304, ...
## Resampling results:
##
##   ROC  Sensitivity  Specificity
##   0.9  0.8          0.8
```

5.2 Model Prediction

```
predictTest <- predict(glmFit4, newdata = trainX)
table(predictTest, trainY)
##         trainY
## predictTest Class_0 Class_1
##   Class_0     123      36
##   Class_1      30     148
```

5.3 Model Performance

```
summary(glmFit4, cor = TRUE)
##
## Call:
## NULL
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3000  -0.5000   0.0000   0.5000   2.3000
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.0000     0.2000  -5.00    0.00
## Age          -1.0000     0.2000  -5.00    0.00
## Weight       -1.0000     0.2000  -5.00    0.00
## Drinks       -1.0000     0.2000  -5.00    0.00
## Beltuse      -1.0000     0.2000  -5.00    0.00
## Female       -1.0000     0.2000  -5.00    0.00
## Restraint    -1.0000     0.2000  -5.00    0.00
##
## Correlation of Coefficients:
##             (Intercept) Age  Weight Drinks Beltuse Female
## Age          0.          1.  0.     0.     0.     0.
## Weight       0.          0.  1.     0.     0.     0.
## Drinks       0.          0.  0.     1.     0.     0.
## Beltuse      0.          0.  0.     0.     1.     0.
## Female       0.          0.  0.     0.     0.     1.
## Restraint    0.          0.  0.     0.     0.     0.
##             Restraint
## Age          0.
## Weight       0.
## Drinks       0.
## Beltuse      0.
## Female       0.
## Restraint    1.
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 0.  on 336  degrees of freedom
## Residual deviance: 0.  on 330  degrees of freedom
## AIC: 14.
##
## Number of Fisher Scoring iterations: 4
```

5.4 Model Comparison

```
resamp <- resamples(list(glmFit = glmFit, glmFit4 = glmFit4))
summary(resamp)
##
## Call:
## summary.resamples(object = resamp)
##
## Models: glmFit, glmFit4
## Number of resamples: 10
##
## ROC
##          Min
## glmFit   0.9
## glmFit4  0.9
##          1st Qu.
## glmFit   0.9
## glmFit4  0.9
##          Median
## glmFit   0.9
## glmFit4  0.9
##          Mean
## glmFit   0.9
## glmFit4  0.9
##          3rd Qu.
## glmFit   0.9
## glmFit4  0.9
##          Max
## glmFit   0.9
## glmFit4  0.9
## NA's
## glmFit   0
## glmFit4  0
##
## Sensitivity
##          Min
## glmFit   0.8
## glmFit4  0.8
##          1st Qu.
## glmFit   0.8
## glmFit4  0.8
##          Median
## glmFit   0.8
## glmFit4  0.8
##          Mean
## glmFit   0.8
## glmFit4  0.8
##          3rd Qu.
## glmFit   0.8
## glmFit4  0.8
##          Max
## glmFit   0.8
## glmFit4  0.8
## NA's
## glmFit   0
## glmFit4  0
##
## Specificity
##          Min
## glmFit   0.8
## glmFit4  0.8
##          1st Qu.
## glmFit   0.8
## glmFit4  0.8
##          Median
## glmFit   0.8
## glmFit4  0.8
##          Mean
## glmFit   0.8
## glmFit4  0.8
##          3rd Qu.
## glmFit   0.8
## glmFit4  0.8
##          Max
## glmFit   0.8
## glmFit4  0.8
## NA's
## glmFit   0
## glmFit4  0
```

6 Bootstrap Resampling Profiles

6.1 Model Fitting

```
set.seed(7)
glmFit5 <- train(trainX,
                 trainY,
                 method = "glm",
                 family = binomial(),
                 metric = "ROC",
                 trControl = ctrl)
glmFit5
## Generalized Linear Model
##
## 337 samples
##   6 predictors
##   2 classes: 'Class_0', 'Class_1'
##
## No pre-processing
## Resampling: Cross-Validation (10 fold)
## Summary of sample sizes: 304, 304, 304, 304, 304, 304, ...
## Resampling results:
##
##   ROC  Sensitivity  Specificity
##   0.9  0.8          0.8
```

6.2 Model Prediction

```
predictTest <- predict(glmFit5, newdata = trainX)
table(predictTest, trainY)
##         trainY
## predictTest Class_0 Class_1
##   Class_0     123      36
##   Class_1      30     148
```

6.3 Model Performance

```
summary(glmFit5, cor = TRUE)
##
## Call:
## NULL
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3000  -0.5000   0.0000   0.5000   2.3000
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.0000     0.2000  -5.00    0.00
## Age          -1.0000     0.2000  -5.00    0.00
## Weight       -1.0000     0.2000  -5.00    0.00
## Drinks       -1.0000     0.2000  -5.00    0.00
## Beltuse      -1.0000     0.2000  -5.00    0.00
## Female       -1.0000     0.2000  -5.00    0.00
## Restraint    -1.0000     0.2000  -5.00    0.00
##
## Correlation of Coefficients:
##             (Intercept) Age  Weight Drinks Beltuse Female
## Age          0.          1.  0.     0.     0.     0.
## Weight       0.          0.  1.     0.     0.     0.
## Drinks       0.          0.  0.     1.     0.     0.
## Beltuse      0.          0.  0.     0.     1.     0.
## Female       0.          0.  0.     0.     0.     1.
## Restraint    0.          0.  0.     0.     0.     0.
##             Restraint
## Age          0.
## Weight       0.
## Drinks       0.
## Beltuse      0.
## Female       0.
## Restraint    1.
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 0.  on 336  degrees of freedom
## Residual deviance: 0.  on 330  degrees of freedom
## AIC: 14.
##
## Number of Fisher Scoring iterations: 4
```

6.4 Model Comparison

```
resamp <- resamples(list(glmFit = glmFit, glmFit5 = glmFit5))
summary(resamp)
##
## Call:
## summary.resamples(object = resamp)
##
## Models: glmFit, glmFit5
## Number of resamples: 10
##
## ROC
##          Min
## glmFit   0.9
## glmFit5  0.9
##          1st Qu.
## glmFit   0.9
## glmFit5  0.9
##          Median
## glmFit   0.9
## glmFit5  0.9
##          Mean
## glmFit   0.9
## glmFit5  0.9
##          3rd Qu.
## glmFit   0.9
## glmFit5  0.9
##          Max
## glmFit   0.9
## glmFit5  0.9
## NA's
## glmFit   0
## glmFit5  0
##
## Sensitivity
##          Min
## glmFit   0.8
## glmFit5  0.8
##          1st Qu.
## glmFit   0.8
## glmFit5  0.8
##          Median
## glmFit   0.8
## glmFit5  0.8
##          Mean
## glmFit   0.8
## glmFit5  0.8
##          3rd Qu.
## glmFit   0.8
## glmFit5  0.8
##          Max
## glmFit   0.8
## glmFit5  0.8
## NA's
## glmFit   0
## glmFit5  0
##
## Specificity
##          Min
## glmFit   0.8
## glmFit5  0.8
##          1st Qu.
## glmFit   0.8
## glmFit5  0.8
##          Median
## glmFit   0.8
## glmFit5  0.8
##          Mean
## glmFit   0.8
## glmFit5  0.8
##          3rd Qu.
## glmFit   0.8
## glmFit5  0.8
##          Max
## glmFit   0.8
## glmFit5  0.8
## NA's
## glmFit   0
## glmFit5  0
```

7 Conclusions

7.1 Session Info

```
sessionInfo()
## R version 3.2.2 (2015-08-14)
## Platform: x86_64-apple-darwin13.4.0 (64-bit)
## Running under: OS X 10.9.5 (Mavericks)
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] caret_6.0-41 ggplot2_1.0.1 lattice_0.20-33
##
## loaded via a namespace (and not attached):
## [1] Rcpp_0.12.0      plyr_1.8.3       iterators_1.0.8  tools_3.2.2
## [5] digest_0.6.8     lubridate_1.3.3  jsonlite_0.9.17  evaluate_0.7
## [9] gtable_0.1.2     nlme_3.1-122     mgtools_1.0.0    lars_1.2
## [13] psych_1.5.8     parallel_3.2.2   foreach_1.4.2    DEoptimR_1.0-2
## [17] proto_0.3-10    e1071_1.6-7      class_7.3-14     grid_3.2.2
## [21] reshape2_1.4.1  MASS_7.3-45      Matrix_1.2-3     formatR_1.2.1
## [25] codetools_0.2-9 stringr_1.0.0    kernlab_0.9-19   stringi_0.5-5
## [29] pls_2.4-3       lubridate.data   RcppEigen_0.3.2.9.0 nnet_7.3-9
## [33] splines_3.2.2   knitr_1.11       pROC_1.7.3       Hmisc_3.17-0
## [37] ipred_0.9-5     Cubist_0.0.18    randomForest_4.6-12 party_1.0-25
## [41] Formula_1.2-1   survival_2.38-3  multcomp_1.4-4   TH.data_1.0-6
## [45] sandwich_2.3-4  zoo_1.7-12       RWeka_0.4-26     RWekajars_3.7.12-1
## [49] partykit_1.0-5  mvtnorm_1.0-3    gtools_3.5.0     ellipse_0.3-8
## [53] Rmisc_1.5       plyrmisc_1.0-1   latticeExtra_0.6-26 RColorBrewer_1.1-2
## [57] cluster_2.1-1   gam_1.12         acepack_1.3-3.3  MuMIn_1.15.6
## [61] xtable_1.8-2    pbkrtest_0.4-2   lme4_1.1-7       Rcpp_0.12.0
## [65] RcppEigen_0.3.2.9.0 SparseM_1.7      MatrixModels_0.4-1 Matrix_1.2-3
## [69] tseriesChaos_0.1-13.1 fseries_1.0-14   fNonlinear_3011.81 timeSeries_3011.93
## [73] timeDate_3011.99 mnormt_1.5-3     lsmeans_2.15-1   digest_0.6.8
## [77] RcppArmadillo_0.4.650.1.0 pamr_1.55        survcomp_1.18.0  prodlim_1.5.5
## [81] impute_1.44.0    preprocessCore_1.28.0 GSEABase_1.28.0  graph_1.46.0
## [85] BiocGenerics_0.14.0 Biobase_2.28.0   boot_1.3-17      abind_1.4-3
## [89] tree_1.0-37      mlbench_2.1-1    maxstat_0.7-25   glmnet_2.0-2
## [93] foreign_0.8-66   glmnetUtils_1.1  gbm_2.1.1        GenSA_1.1.6
## [97] earth_4.4.0      elasticnet_1.1   dynamicTreeCut_1.63-1 deepnet_0.2
## [101] AUC_0.3.0       ada_2.0-5        adabag_4.1       accuracy_0.1-1
## [105] Metrics_0.1.1   penalizedLDA_1.00 ipw_1.0-9        HDtest_2.0
## [109] energy_1.6.2    Rdistance_1.1    ape_3.4          gtx_0.0.6
## [113] gdata_2.17.0    GPfit_1.0-8      fastICA_1.2-0    earth.bigmem_4.4.0-beta
## [117] plotmo_3.1.0    TeachingDemos_2.10 diagnosis_0.3    ROCR_1.0-7
## [121] spatstat_1.42-1 spatstat.data_1.1-1 rpart_4.1-10     RANN_2.4.1
## [125] spam_1.3-0      pester_0.1.2     RGCCA_2.1        impact_1.0
## [129] GenKern_1.1     fda_2.4.4        expsmooth_2.3    dtw_1.18-1
## [133] dlm_1.1-3       diagram_1.6      CVTuning_1.3.3   clue_0.3-47
## [137] bootstrap_2.3   ANN2_1.0         AnomalyDetection_1.0.0 ALSM_1.0
## [141] ALL.wt.mRNA_1.0 ALL.wt.miRNA_1.0 ALL.mRNA_1.0     ALL.miRNA_1.0
## [145] ALL_1.8.0       AgeDataCorrection_1.0-1 ADGofTest_0.3    ACCLIMATE_1.0
## [149] ACCLIMATE.data_1.0 ABCoptim_0.14    ABC.RAP_0.1      ABC.EX_1.0
## [153] ABC.DE3_1.0     ABC.DE2_1.0      ABC.DE1_1.0      ABC.ARS_1.0
## [157] ABC.MCMC_1.0    ABC.EL_1.0       ABC.ABC_1.0      ABCdata_1.0
## [161] ABCDE_1.0       ABC.RAP.data_1.0 ABC.EX.data_1.0  ABC.DE3.data_1.0
## [165] ABC.DE2.data_1.0 ABC.DE1.data_1.0 ABC.ARS.data_1.0 ABC.MCMC.data_1.0
## [169] ABC.EL.data_1.0 ABC.ABC.data_1.0 ABCdata.data_1.0 ABCDE.data_1.0
## [173] ACCLIMATE.data_1.0 ALL.wt.mRNA.data_1.0 ALL.wt.miRNA.data_1.0 ALL.mRNA.data_1.0
## [177] ALL.miRNA.data_1.0 ALL.data_1.8.0 ADGofTest.data_0.3 AgeDataCorrection.data_1.0-1
## [181] ANN2.data_1.0   AnomalyDetection.data_1.0.0 ALSM.data_1.0 ACCLIMATE.data_1.0
## [185] ABCoptim.data_0.14 ABC.RAP.data_1.0 ABC.EX.data_1.0 ABC.DE3.data_1.0
## [189] ABC.DE2.data_1.0 ABC.DE1.data_1.0 ABC.ARS.data_1.0 ABC.MCMC.data_1.0
## [193] ABC.EL.data_1.0 ABC.ABC.data_1.0 ABCdata.data_1.0 ABCDE.data_1.0
```

7.2 Bibliography

Logit

Logit regression models have been extended from the binomial to the multinomial case in many ways and are fairly well understood. Agresti (2002) provides a good overview of the theory behind logit models and Hosmer and Lemeshow (2000) provides a more applied treatment.

Let $\mathbf{x}_i$ denote the ith observation, $i = 1, \ldots, N$, with intercept $x_{i0} = 1$. Let K be the number of classes and let y_i be the response for $\mathbf{x}_i$. Then, with the last class used as a baseline, the multinomial logit model is

$$\log \frac{\Pr(y_i = k \mid \mathbf{x}_i)}{\Pr(y_i = K \mid \mathbf{x}_i)} = \beta_{k0} + \beta_k^T \mathbf{x}_i.$$

Regularized versions of logit models have been developed as well, for example, the lasso (Tibshirani, 1996) and the elastic net (Zou and Hastie, 2005). Regularized logit models have been developed using the lasso (Shevade and Keerthi, 2003), the elastic net (Friedman, Hastie, and Tibshirani, 2010), and the fused lasso (Tibshirani et al., 2005). Let $\mathbf{\beta} = (\beta_{10}, \ldots, \beta_{K0}, \beta_1^T, \ldots, \beta_K^T)^T$ be the vector of all parameters in the model, and let $\mathbf{\beta}_{-0}$ be the vector of all parameters except the intercepts. Then the lasso-regularized multinomial logit model maximizes the penalized log-likelihood

$$\ell(\mathbf{\beta}) - \lambda \sum_{k=1}^{K} \|\beta_k\|_1$$

for some $\lambda \geq 0$, and the elastic net-regularized multinomial logit model maximizes the penalized log-likelihood

$$\ell(\mathbf{\beta}) - \lambda \sum_{k=1}^{K} \left(\alpha \|\beta_k\|_1 + (1-\alpha)\|\beta_k\|_2^2\right)$$

for some $\lambda \geq 0$ and $0 \leq \alpha \leq 1$.

Let $\mathbf{D}_k$ be the $(K-1) \times (K-1)$ diagonal matrix with ith element $\Pr(y_i = k \mid \mathbf{x}_i)(1 - \Pr(y_i = k \mid \mathbf{x}_i))$, let $\mathbf{W}_k$ be the $(K-1) \times (K-1)$ weight matrix $\mathbf{W}_k = \mathbf{D}_k - \mathbf{D}_k \mathbf{X} (\mathbf{X}^T \mathbf{D}_k \mathbf{X})^{-1} \mathbf{X}^T \mathbf{D}_k$, and let $\mathbf{z}_k$ be the $(K-1)$-vector $\mathbf{z}_k = \mathbf{D}_k^{-1}(\mathbf{y}_k - \mathbf{p}_k) + \mathbf{X}\beta_k$. Then the Newton-Raphson update for the unpenalized multinomial logit model maximizes the quadratic form

$$\sum_{k=1}^{K} \mathbf{z}_k^T \mathbf{W}_k \mathbf{X} \beta_k$$

Friedman, Hastie, and Tibshirani (2010) show that the Newton-Raphson update for the elastic net-regularized multinomial logit model maximizes the penalized quadratic form

$$\sum_{k=1}^{K} \mathbf{z}_k^T \mathbf{W}_k \mathbf{X} \beta_k - \lambda \left(\alpha \|\beta_k\|_1 + (1-\alpha)\|\beta_k\|_2^2\right)$$

Friedman, Hastie, and Tibshirani (2010) show that the coordinate descent update for the elastic net-regularized multinomial logit model maximizes the penalized quadratic form

$$\sum_{k=1}^{K} \mathbf{z}_k^T \mathbf{W}_k \mathbf{X} \beta_k - \lambda \left(\alpha \|\beta_k\|_1 + (1-\alpha)\|\beta_k\|_2^2\right)$$