log_reg.R

abby_surmeier

Mon Aug 6 10:11:43 2018

```
# log regression using sonar dataset
library(mlbench)
data(Sonar)
summary(Sonar)
```

```
##        V1                V2                V3                V4
##  Min.   :0.00150   Min.   :0.00060   Min.   :0.00150   Min.   :0.00580
##  1st Qu.:0.01335   1st Qu.:0.01645   1st Qu.:0.01895   1st Qu.:0.02438
##  Median :0.02280   Median :0.03080   Median :0.03430   Median :0.04405
##  Mean   :0.02916   Mean   :0.03844   Mean   :0.04383   Mean   :0.05389
##  3rd Qu.:0.03555   3rd Qu.:0.04795   3rd Qu.:0.05795   3rd Qu.:0.06450
##  Max.   :0.13710   Max.   :0.23390   Max.   :0.30590   Max.   :0.42640
##        V5                V6                V7                V8
##  Min.   :0.00670   Min.   :0.01020   Min.   :0.0033   Min.   :0.00550
##  1st Qu.:0.03805   1st Qu.:0.06703   1st Qu.:0.0809   1st Qu.:0.08042
##  Median :0.06250   Median :0.09215   Median :0.1070   Median :0.11210
##  Mean   :0.07520   Mean   :0.10457   Mean   :0.1217   Mean   :0.13480
##  3rd Qu.:0.10028   3rd Qu.:0.13412   3rd Qu.:0.1540   3rd Qu.:0.16960
##  Max.   :0.40100   Max.   :0.38230   Max.   :0.3729   Max.   :0.45900
##        V9               V10               V11               V12
##  Min.   :0.00750   Min.   :0.0113   Min.   :0.0289   Min.   :0.0236
##  1st Qu.:0.09703   1st Qu.:0.1113   1st Qu.:0.1293   1st Qu.:0.1335
##  Median :0.15225   Median :0.1824   Median :0.2248   Median :0.2490
##  Mean   :0.17800   Mean   :0.2083   Mean   :0.2360   Mean   :0.2502
##  3rd Qu.:0.23342   3rd Qu.:0.2687   3rd Qu.:0.3016   3rd Qu.:0.3312
##  Max.   :0.68280   Max.   :0.7106   Max.   :0.7342   Max.   :0.7060
##       V13               V14               V15               V16
##  Min.   :0.0184   Min.   :0.0273   Min.   :0.0031   Min.   :0.0162
##  1st Qu.:0.1661   1st Qu.:0.1752   1st Qu.:0.1646   1st Qu.:0.1963
##  Median :0.2640   Median :0.2811   Median :0.2817   Median :0.3047
##  Mean   :0.2733   Mean   :0.2966   Mean   :0.3202   Mean   :0.3785
##  3rd Qu.:0.3513   3rd Qu.:0.3862   3rd Qu.:0.4529   3rd Qu.:0.5357
##  Max.   :0.7131   Max.   :0.9970   Max.   :1.0000   Max.   :0.9988
##       V17               V18               V19               V20
##  Min.   :0.0349   Min.   :0.0375   Min.   :0.0494   Min.   :0.0656
##  1st Qu.:0.2059   1st Qu.:0.2421   1st Qu.:0.2991   1st Qu.:0.3506
##  Median :0.3084   Median :0.3683   Median :0.4350   Median :0.5425
##  Mean   :0.4160   Mean   :0.4523   Mean   :0.5048   Mean   :0.5630
##  3rd Qu.:0.6594   3rd Qu.:0.6791   3rd Qu.:0.7314   3rd Qu.:0.8093
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##       V21               V22               V23               V24
##  Min.   :0.0512   Min.   :0.0219   Min.   :0.0563   Min.   :0.0239
##  1st Qu.:0.3997   1st Qu.:0.4069   1st Qu.:0.4502   1st Qu.:0.5407
##  Median :0.6177   Median :0.6649   Median :0.6997   Median :0.6985
##  Mean   :0.6091   Mean   :0.6243   Mean   :0.6470   Mean   :0.6727
##  3rd Qu.:0.8170   3rd Qu.:0.8320   3rd Qu.:0.8486   3rd Qu.:0.8722
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
```

```
##       V25              V26              V27              V28
##  Min.   :0.0240   Min.   :0.0921   Min.   :0.0481   Min.   :0.0284
##  1st Qu.:0.5258   1st Qu.:0.5442   1st Qu.:0.5319   1st Qu.:0.5348
##  Median :0.7211   Median :0.7545   Median :0.7456   Median :0.7319
##  Mean   :0.6754   Mean   :0.6999   Mean   :0.7022   Mean   :0.6940
##  3rd Qu.:0.8737   3rd Qu.:0.8938   3rd Qu.:0.9171   3rd Qu.:0.9003
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##       V29              V30              V31              V32
##  Min.   :0.0144   Min.   :0.0613   Min.   :0.0482   Min.   :0.0404
##  1st Qu.:0.4637   1st Qu.:0.4114   1st Qu.:0.3456   1st Qu.:0.2814
##  Median :0.6808   Median :0.6071   Median :0.4904   Median :0.4296
##  Mean   :0.6421   Mean   :0.5809   Mean   :0.5045   Mean   :0.4390
##  3rd Qu.:0.8521   3rd Qu.:0.7352   3rd Qu.:0.6420   3rd Qu.:0.5803
##  Max.   :1.0000   Max.   :1.0000   Max.   :0.9657   Max.   :0.9306
##       V33              V34              V35              V36
##  Min.   :0.0477   Min.   :0.0212   Min.   :0.0223   Min.   :0.0080
##  1st Qu.:0.2579   1st Qu.:0.2176   1st Qu.:0.1794   1st Qu.:0.1543
##  Median :0.3912   Median :0.3510   Median :0.3127   Median :0.3211
##  Mean   :0.4172   Mean   :0.4032   Mean   :0.3926   Mean   :0.3848
##  3rd Qu.:0.5561   3rd Qu.:0.5961   3rd Qu.:0.5934   3rd Qu.:0.5565
##  Max.   :1.0000   Max.   :0.9647   Max.   :1.0000   Max.   :1.0000
##       V37              V38              V39              V40
##  Min.   :0.0351   Min.   :0.0383   Min.   :0.0371   Min.   :0.0117
##  1st Qu.:0.1601   1st Qu.:0.1743   1st Qu.:0.1740   1st Qu.:0.1865
##  Median :0.3063   Median :0.3127   Median :0.2835   Median :0.2781
##  Mean   :0.3638   Mean   :0.3397   Mean   :0.3258   Mean   :0.3112
##  3rd Qu.:0.5189   3rd Qu.:0.4405   3rd Qu.:0.4349   3rd Qu.:0.4244
##  Max.   :0.9497   Max.   :1.0000   Max.   :0.9857   Max.   :0.9297
##       V41              V42              V43              V44
##  Min.   :0.0360   Min.   :0.0056   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.1631   1st Qu.:0.1589   1st Qu.:0.1552   1st Qu.:0.1269
##  Median :0.2595   Median :0.2451   Median :0.2225   Median :0.1777
##  Mean   :0.2893   Mean   :0.2783   Mean   :0.2465   Mean   :0.2141
##  3rd Qu.:0.3875   3rd Qu.:0.3842   3rd Qu.:0.3245   3rd Qu.:0.2717
##  Max.   :0.8995   Max.   :0.8246   Max.   :0.7733   Max.   :0.7762
##       V45               V46               V47               V48
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.09448   1st Qu.:0.06855   1st Qu.:0.06425   1st Qu.:0.04512
##  Median :0.14800   Median :0.12135   Median :0.10165   Median :0.07810
##  Mean   :0.19723   Mean   :0.16063   Mean   :0.12245   Mean   :0.09142
##  3rd Qu.:0.23155   3rd Qu.:0.20037   3rd Qu.:0.15443   3rd Qu.:0.12010
##  Max.   :0.70340   Max.   :0.72920   Max.   :0.55220   Max.   :0.33390
##       V49               V50               V51                V52
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.000000   Min.   :0.000800
##  1st Qu.:0.02635   1st Qu.:0.01155   1st Qu.:0.008425   1st Qu.:0.007275
##  Median :0.04470   Median :0.01790   Median :0.013900   Median :0.011400
##  Mean   :0.05193   Mean   :0.02042   Mean   :0.016069   Mean   :0.013420
##  3rd Qu.:0.06853   3rd Qu.:0.02527   3rd Qu.:0.020825   3rd Qu.:0.016725
##  Max.   :0.19810   Max.   :0.08250   Max.   :0.100400   Max.   :0.070900
##       V53                V54               V55
##  Min.   :0.000500   Min.   :0.001000   Min.   :0.00060
##  1st Qu.:0.005075   1st Qu.:0.005375   1st Qu.:0.00415
##  Median :0.009550   Median :0.009300   Median :0.00750
##  Mean   :0.010709   Mean   :0.010941   Mean   :0.00929
```

```
##  3rd Qu.:0.014900   3rd Qu.:0.014500   3rd Qu.:0.01210
##  Max.   :0.039000   Max.   :0.035200   Max.   :0.04470
##       V56                V57                V58
##  Min.   :0.000400   Min.   :0.00030   Min.   :0.000300
##  1st Qu.:0.004400   1st Qu.:0.00370   1st Qu.:0.003600
##  Median :0.006850   Median :0.00595   Median :0.005800
##  Mean   :0.008222   Mean   :0.00782   Mean   :0.007949
##  3rd Qu.:0.010575   3rd Qu.:0.01043   3rd Qu.:0.010350
##  Max.   :0.039400   Max.   :0.03550   Max.   :0.044000
##       V59                V60           Class
##  Min.   :0.000100   Min.   :0.000600   M:111
##  1st Qu.:0.003675   1st Qu.:0.003100   R: 97
##  Median :0.006400   Median :0.005300
##  Mean   :0.007941   Mean   :0.006507
##  3rd Qu.:0.010325   3rd Qu.:0.008525
##  Max.   :0.036400   Max.   :0.043900
```

```
str(Sonar)
```

```
## 'data.frame':    208 obs. of  61 variables:
##  $ V1   : num  0.02 0.0453 0.0262 0.01 0.0762 0.0286 0.0317 0.0519 0.0223 0.0164 ...
##  $ V2   : num  0.0371 0.0523 0.0582 0.0171 0.0666 0.0453 0.0956 0.0548 0.0375 0.0173 ...
##  $ V3   : num  0.0428 0.0843 0.1099 0.0623 0.0481 ...
##  $ V4   : num  0.0207 0.0689 0.1083 0.0205 0.0394 ...
##  $ V5   : num  0.0954 0.1183 0.0974 0.0205 0.059 ...
##  $ V6   : num  0.0986 0.2583 0.228 0.0368 0.0649 ...
##  $ V7   : num  0.154 0.216 0.243 0.11 0.121 ...
##  $ V8   : num  0.16 0.348 0.377 0.128 0.247 ...
##  $ V9   : num  0.3109 0.3337 0.5598 0.0598 0.3564 ...
##  $ V10  : num  0.211 0.287 0.619 0.126 0.446 ...
##  $ V11  : num  0.1609 0.4918 0.6333 0.0881 0.4152 ...
##  $ V12  : num  0.158 0.655 0.706 0.199 0.395 ...
##  $ V13  : num  0.2238 0.6919 0.5544 0.0184 0.4256 ...
##  $ V14  : num  0.0645 0.7797 0.532 0.2261 0.4135 ...
##  $ V15  : num  0.066 0.746 0.648 0.173 0.453 ...
##  $ V16  : num  0.227 0.944 0.693 0.213 0.533 ...
##  $ V17  : num  0.31 1 0.6759 0.0693 0.7306 ...
##  $ V18  : num  0.3 0.887 0.755 0.228 0.619 ...
##  $ V19  : num  0.508 0.802 0.893 0.406 0.203 ...
##  $ V20  : num  0.48 0.782 0.862 0.397 0.464 ...
##  $ V21  : num  0.578 0.521 0.797 0.274 0.415 ...
##  $ V22  : num  0.507 0.405 0.674 0.369 0.429 ...
##  $ V23  : num  0.433 0.396 0.429 0.556 0.573 ...
##  $ V24  : num  0.555 0.391 0.365 0.485 0.54 ...
##  $ V25  : num  0.671 0.325 0.533 0.314 0.316 ...
##  $ V26  : num  0.641 0.32 0.241 0.533 0.229 ...
##  $ V27  : num  0.71 0.327 0.507 0.526 0.7 ...
##  $ V28  : num  0.808 0.277 0.853 0.252 1 ...
##  $ V29  : num  0.679 0.442 0.604 0.209 0.726 ...
##  $ V30  : num  0.386 0.203 0.851 0.356 0.472 ...
##  $ V31  : num  0.131 0.379 0.851 0.626 0.51 ...
##  $ V32  : num  0.26 0.295 0.504 0.734 0.546 ...
##  $ V33  : num  0.512 0.198 0.186 0.612 0.288 ...
##  $ V34  : num  0.7547 0.2341 0.2709 0.3497 0.0981 ...
##  $ V35  : num  0.854 0.131 0.423 0.395 0.195 ...
```

```
##  $ V36  : num  0.851 0.418 0.304 0.301 0.418 ...
##  $ V37  : num  0.669 0.384 0.612 0.541 0.46 ...
##  $ V38  : num  0.61 0.106 0.676 0.881 0.322 ...
##  $ V39  : num  0.494 0.184 0.537 0.986 0.283 ...
##  $ V40  : num  0.274 0.197 0.472 0.917 0.243 ...
##  $ V41  : num  0.051 0.167 0.465 0.612 0.198 ...
##  $ V42  : num  0.2834 0.0583 0.2587 0.5006 0.2444 ...
##  $ V43  : num  0.282 0.14 0.213 0.321 0.185 ...
##  $ V44  : num  0.4256 0.1628 0.2222 0.3202 0.0841 ...
##  $ V45  : num  0.2641 0.0621 0.2111 0.4295 0.0692 ...
##  $ V46  : num  0.1386 0.0203 0.0176 0.3654 0.0528 ...
##  $ V47  : num  0.1051 0.053 0.1348 0.2655 0.0357 ...
##  $ V48  : num  0.1343 0.0742 0.0744 0.1576 0.0085 ...
##  $ V49  : num  0.0383 0.0409 0.013 0.0681 0.023 0.0264 0.0507 0.0285 0.0777 0.0092 ...
##  $ V50  : num  0.0324 0.0061 0.0106 0.0294 0.0046 0.0081 0.0159 0.0178 0.0439 0.0198 ...
##  $ V51  : num  0.0232 0.0125 0.0033 0.0241 0.0156 0.0104 0.0195 0.0052 0.0061 0.0118 ...
##  $ V52  : num  0.0027 0.0084 0.0232 0.0121 0.0031 0.0045 0.0201 0.0081 0.0145 0.009 ...
##  $ V53  : num  0.0065 0.0089 0.0166 0.0036 0.0054 0.0014 0.0248 0.012 0.0128 0.0223 ...
##  $ V54  : num  0.0159 0.0048 0.0095 0.015 0.0105 0.0038 0.0131 0.0045 0.0145 0.0179 ...
##  $ V55  : num  0.0072 0.0094 0.018 0.0085 0.011 0.0013 0.007 0.0121 0.0058 0.0084 ...
##  $ V56  : num  0.0167 0.0191 0.0244 0.0073 0.0015 0.0089 0.0138 0.0097 0.0049 0.0068 ...
##  $ V57  : num  0.018 0.014 0.0316 0.005 0.0072 0.0057 0.0092 0.0085 0.0065 0.0032 ...
##  $ V58  : num  0.0084 0.0049 0.0164 0.0044 0.0048 0.0027 0.0143 0.0047 0.0093 0.0035 ...
##  $ V59  : num  0.009 0.0052 0.0095 0.004 0.0107 0.0051 0.0036 0.0048 0.0059 0.0056 ...
##  $ V60  : num  0.0032 0.0044 0.0078 0.0117 0.0094 0.0062 0.0103 0.0053 0.0022 0.004 ...
##  $ Class: Factor w/ 2 levels "M","R": 2 2 2 2 2 2 2 2 2 2 ...
```

```
apply(Sonar, 2, var)
```

```
## Warning in FUN(newX[, i], ...): NAs introduced by coercion
```

```
##           V1           V2           V3           V4           V5
## 5.285821e-04 1.086357e-03 1.476720e-03 2.164851e-03 3.086043e-03
##           V6           V7           V8           V9          V10
## 3.493453e-03 3.817743e-03 7.250894e-03 1.401553e-02 1.806759e-02
##          V11          V12          V13          V14          V15
## 1.761051e-02 1.962016e-02 1.987031e-02 2.705183e-02 4.220017e-02
##          V16          V17          V18          V19          V20
## 5.412602e-02 6.952560e-02 6.839759e-02 6.655799e-02 6.898649e-02
##          V21          V22          V23          V24          V25
## 6.647019e-02 6.547617e-02 6.258773e-02 5.717656e-02 5.998896e-02
##          V26          V27          V28          V29          V30
## 5.627700e-02 6.034734e-02 5.625864e-02 5.771995e-02 4.873029e-02
##          V31          V32          V33          V34          V35
## 4.579276e-02 4.546998e-02 4.264758e-02 5.347266e-02 6.714950e-02
##          V36          V37          V38          V39          V40
## 6.975989e-02 5.755796e-02 4.535763e-02 3.963069e-02 3.191994e-02
##          V41          V42          V43          V44          V45
## 2.927901e-02 2.846923e-02 1.931916e-02 1.776648e-02 2.299119e-02
##          V46          V47          V48          V49          V50
## 1.793937e-02 7.560777e-03 3.895846e-03 1.292688e-03 1.867326e-04
##          V51          V52          V53          V54          V55
## 1.442040e-04 9.281370e-05 4.984837e-05 5.330175e-05 5.023450e-05
##          V56          V57          V58          V59          V60
## 3.290296e-05 3.347196e-05 4.185884e-05 3.820949e-05 2.531140e-05
```

```
##        Class
##           NA
```

```
#80/20 split
train_index <- sample(1:nrow(Sonar), 166)
train <- Sonar[train_index,]
test <- Sonar[-train_index,]

#Log regression
set.seed(123)
library(caret)
library(LogicReg)
ctrl <- trainControl(method = "repeatedcv", number = 10, repeats = 3)
log_mod <- train(Class ~., data = train, method = "glm", family = "binomial"(link = 'logit'), preProcess
```

```
## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```r
log_pred <- predict(log_mod, newdata = test)
confusionMatrix(log_pred, test$Class)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  M  R
##          M 21  7
##          R  3 11
##
##                Accuracy : 0.7619
##                  95% CI : (0.6055, 0.8795)
##     No Information Rate : 0.5714
##     P-Value [Acc > NIR] : 0.008242
##
##                   Kappa : 0.5
##  Mcnemar's Test P-Value : 0.342782
```

```
##
##             Sensitivity : 0.8750
##             Specificity : 0.6111
##          Pos Pred Value : 0.7500
##          Neg Pred Value : 0.7857
##              Prevalence : 0.5714
##          Detection Rate : 0.5000
##    Detection Prevalence : 0.6667
##       Balanced Accuracy : 0.7431
##
##        'Positive' Class : M
##
```