jack-eda.R

lajackso

Sun Feb 07 12:23:15 2016

```
# Log Loss function
# calculate logloss; p is a probability vector; y is a vector of zero or one.
logloss <- function(p,y) {
      xp <- function(pv) {return(max(min(pv,1-10^{-15}),10^{-15}))}
      p <- sapply(p,xp)
      loss <- -sum(y*log(p)+(1-y)*log(1-p))/length(p)
      return(loss)
}
# example log loss
p<-c(0.4,0.4,.55,.55,.55); y<-c(0,0,1,1,1);
logloss(p,y)
```

```
## [1] 0.5630324
```

```
# Log Loss of all predictions 0.5 (Kaggle Leaderboard: All 0.5 Benchmark = 0.69315)
logloss(rep(0.5,2500+7500),c(rep(0,2500),rep(1,7500)))
```

```
## [1] 0.6931472
```

```
################

# Get data
directory <- "C:/Users/lajackso/Documents/GitHub/cardif/"
fileData <- paste0(directory,"train.csv")
data <- read.csv(fileData)

# rows and columns
dim(data)
```

```
## [1] 114321    133
```

```
# fraction complete cases
sum(complete.cases(data))/dim(data)[1]
```

```
## [1] 0.5472398
```

```
# identify numeric columns (convert others to factors)
numericCols <- which(sapply(data, is.numeric))
data[,-(numericCols)] <- lapply(data[,-(numericCols)] , factor)
# summary for numerical columns
summary(data[,numericCols])
```

```
##        ID              target             v1              v2
##  Min.   :     3   Min.   :0.0000   Min.   : 0.00   Min.   : 0.00
```

```
##  1st Qu.: 57280   1st Qu.:1.0000   1st Qu.: 0.91   1st Qu.: 5.32
##  Median :114189   Median :1.0000   Median : 1.47   Median : 7.02
##  Mean   :114229   Mean   :0.7612   Mean   : 1.63   Mean   : 7.46
##  3rd Qu.:171206   3rd Qu.:1.0000   3rd Qu.: 2.14   3rd Qu.: 9.47
##  Max.   :228713   Max.   :1.0000   Max.   :20.00   Max.   :20.00
##                                    NA's   :49832   NA's   :49796
##        v4               v5               v6               v7
##  Min.   : 0.00   Min.   : 0.00   Min.   : 0.00   Min.   : 0.00
##  1st Qu.: 3.49   1st Qu.: 7.61   1st Qu.: 2.07   1st Qu.: 2.10
##  Median : 4.21   Median : 8.67   Median : 2.41   Median : 2.45
##  Mean   : 4.15   Mean   : 8.74   Mean   : 2.44   Mean   : 2.48
##  3rd Qu.: 4.83   3rd Qu.: 9.77   3rd Qu.: 2.78   3rd Qu.: 2.83
##  Max.   :20.00   Max.   :20.00   Max.   :20.00   Max.   :20.00
##  NA's   :49796   NA's   :48624   NA's   :49832   NA's   :49832
##        v8               v9              v10              v11
##  Min.   : 0.00   Min.   : 0.00   Min.   : 0.000   Min.   : 0.00
##  1st Qu.: 0.09   1st Qu.: 7.85   1st Qu.: 1.050   1st Qu.:15.00
##  Median : 0.39   Median : 9.06   Median : 1.313   Median :15.50
##  Mean   : 1.50   Mean   : 9.03   Mean   : 1.883   Mean   :15.45
##  3rd Qu.: 1.63   3rd Qu.:10.23   3rd Qu.: 2.101   3rd Qu.:15.95
##  Max.   :20.00   Max.   :20.00   Max.   :18.534   Max.   :20.00
##  NA's   :48619   NA's   :49851   NA's   :84       NA's   :49836
##       v12              v13              v14                v15
##  Min.   : 0.000   Min.   : 0.00   Min.   :-0.000001   Min.   : 0.00
##  1st Qu.: 6.322   1st Qu.: 3.07   1st Qu.:11.256017   1st Qu.: 1.61
##  Median : 6.613   Median : 3.59   Median :11.967825   Median : 1.99
##  Mean   : 6.881   Mean   : 3.80   Mean   :12.094279   Mean   : 2.08
##  3rd Qu.: 7.020   3rd Qu.: 4.29   3rd Qu.:12.715774   3rd Qu.: 2.42
##  Max.   :18.711   Max.   :20.00   Max.   :20.000000   Max.   :20.00
##  NA's   :86       NA's   :49832   NA's   :4           NA's   :49836
##       v16              v17              v18              v19
##  Min.   : 0.00   Min.   : 0.00   Min.   : 0.00   Min.   : 0.00
##  1st Qu.: 3.86   1st Qu.: 2.70   1st Qu.: 0.51   1st Qu.: 0.17
##  Median : 4.93   Median : 3.55   Median : 0.77   Median : 0.20
##  Mean   : 4.92   Mean   : 3.83   Mean   : 0.84   Mean   : 0.22
##  3rd Qu.: 5.96   3rd Qu.: 4.51   3rd Qu.: 1.07   3rd Qu.: 0.24
##  Max.   :20.00   Max.   :20.00   Max.   :20.00   Max.   :20.00
##  NA's   :49895   NA's   :49796   NA's   :49832   NA's   :49843
##       v20              v21              v23              v25
##  Min.   : 1.52   Min.   : 0.1062   Min.   : 0.00   Min.   : 0.04
##  1st Qu.:17.33   1st Qu.: 6.4155   1st Qu.: 0.00   1st Qu.: 0.15
##  Median :18.04   Median : 7.0454   Median : 0.00   Median : 0.47
##  Mean   :17.77   Mean   : 7.0297   Mean   : 1.09   Mean   : 1.70
##  3rd Qu.:18.54   3rd Qu.: 7.6706   3rd Qu.: 0.00   3rd Qu.: 1.95
##  Max.   :20.00   Max.   :19.2961   Max.   :20.00   Max.   :20.00
##  NA's   :49840   NA's   :611       NA's   :50675   NA's   :48619
##       v26              v27              v28              v29
##  Min.   : 0.00   Min.   : 0.00   Min.   : 0.00   Min.   : 0.00
##  1st Qu.: 1.51   1st Qu.: 2.20   1st Qu.: 3.49   1st Qu.: 7.43
##  Median : 1.83   Median : 2.67   Median : 5.04   Median : 8.30
##  Mean   : 1.88   Mean   : 2.74   Mean   : 5.09   Mean   : 8.21
##  3rd Qu.: 2.18   3rd Qu.: 3.22   3rd Qu.: 6.57   3rd Qu.: 9.09
##  Max.   :20.00   Max.   :20.00   Max.   :19.85   Max.   :20.00
##  NA's   :49832   NA's   :49832   NA's   :49832   NA's   :49832
```

```
##       v32             v33             v34              v35
##  Min.   : 0.00   Min.   : 0.00   Min.   : 0.000   Min.   : 0.00
##  1st Qu.: 1.26   1st Qu.: 1.47   1st Qu.: 5.054   1st Qu.: 7.25
##  Median : 1.56   Median : 1.95   Median : 6.537   Median : 8.07
##  Mean   : 1.62   Mean   : 2.16   Mean   : 6.406   Mean   : 8.12
##  3rd Qu.: 1.90   3rd Qu.: 2.63   3rd Qu.: 7.703   3rd Qu.: 8.94
##  Max.   :17.56   Max.   :20.00   Max.   :20.000   Max.   :20.00
##  NA's   :49832   NA's   :49832   NA's   :111      NA's   :49832
##       v36             v37             v38               v39
##  Min.   : 0.00   Min.   : 0.00   Min.   : 0.00000   Min.   : 0.00
##  1st Qu.:11.77   1st Qu.: 0.40   1st Qu.: 0.00000   1st Qu.: 0.13
##  Median :13.77   Median : 0.64   Median : 0.00000   Median : 0.38
##  Mean   :13.38   Mean   : 0.74   Mean   : 0.09093   Mean   : 1.24
##  3rd Qu.:15.32   3rd Qu.: 0.95   3rd Qu.: 0.00000   3rd Qu.: 1.19
##  Max.   :20.00   Max.   :20.00   Max.   :12.00000   Max.   :19.92
##  NA's   :48624   NA's   :49843                      NA's   :49836
##       v40              v41             v42             v43
##  Min.   : 0.000   Min.   : 0.00   Min.   : 0.00   Min.   : 0.00
##  1st Qu.: 8.408   1st Qu.: 6.54   1st Qu.:12.34   1st Qu.: 1.79
##  Median :10.334   Median : 7.20   Median :12.93   Median : 2.15
##  Mean   :10.466   Mean   : 7.18   Mean   :12.92   Mean   : 2.22
##  3rd Qu.:12.765   3rd Qu.: 7.83   3rd Qu.:13.49   3rd Qu.: 2.56
##  Max.   :20.000   Max.   :20.00   Max.   :20.00   Max.   :20.00
##  NA's   :111      NA's   :49832   NA's   :49832   NA's   :49836
##       v44             v45             v46             v48
##  Min.   : 0.00   Min.   : 0.00   Min.   : 0.07   Min.   : 0.00
##  1st Qu.: 9.58   1st Qu.: 7.83   1st Qu.: 0.12   1st Qu.:11.22
##  Median :10.78   Median : 9.16   Median : 0.44   Median :12.41
##  Mean   :10.80   Mean   : 9.14   Mean   : 1.63   Mean   :12.54
##  3rd Qu.:12.02   3rd Qu.:10.42   3rd Qu.: 1.82   3rd Qu.:13.78
##  Max.   :19.83   Max.   :20.00   Max.   :20.00   Max.   :20.00
##  NA's   :49796   NA's   :49832   NA's   :48619   NA's   :49796
##       v49             v50              v51             v53
##  Min.   : 0.00   Min.   : 0.0000   Min.   : 0.00   Min.   : 0.00
##  1st Qu.: 7.47   1st Qu.: 0.6588   1st Qu.: 5.60   1st Qu.:15.28
##  Median : 8.02   Median : 1.2119   Median : 7.13   Median :15.77
##  Mean   : 8.02   Mean   : 1.5043   Mean   : 7.20   Mean   :15.71
##  3rd Qu.: 8.56   3rd Qu.: 2.0072   3rd Qu.: 8.64   3rd Qu.:16.22
##  Max.   :20.00   Max.   :20.0000   Max.   :20.00   Max.   :20.00
##  NA's   :49832   NA's   :86        NA's   :50678   NA's   :49836
##       v54             v55             v57             v58
##  Min.   : 0.01   Min.   : 0.00   Min.   : 0.00   Min.   : 0.00
##  1st Qu.: 0.09   1st Qu.: 0.98   1st Qu.: 3.65   1st Qu.: 1.50
##  Median : 0.31   Median : 1.37   Median : 4.07   Median : 5.33
##  Mean   : 1.25   Mean   : 1.56   Mean   : 4.08   Mean   : 7.70
##  3rd Qu.: 1.41   3rd Qu.: 1.94   3rd Qu.: 4.49   3rd Qu.:13.96
##  Max.   :20.00   Max.   :20.00   Max.   :20.00   Max.   :20.00
##  NA's   :48619   NA's   :49832   NA's   :49832   NA's   :49836
##       v59             v60             v61             v62
##  Min.   : 0.00   Min.   : 0.00   Min.   : 0.00   Min.   :0.000
##  1st Qu.: 9.06   1st Qu.: 1.36   1st Qu.:13.60   1st Qu.:1.000
##  Median :10.54   Median : 1.67   Median :15.08   Median :1.000
##  Mean   :10.59   Mean   : 1.71   Mean   :14.58   Mean   :1.031
##  3rd Qu.:12.03   3rd Qu.: 2.01   3rd Qu.:16.11   3rd Qu.:1.000
```

```
##  Max.   :20.00   Max.   :20.00   Max.   :18.85   Max.   :7.000
##  NA's   :49796   NA's   :49832   NA's   :49796
##       v63             v64             v65             v67
##  Min.   : 0.05   Min.   : 0.00   Min.   : 0.66   Min.   : 0.00
##  1st Qu.: 0.14   1st Qu.: 4.79   1st Qu.:15.03   1st Qu.: 8.58
##  Median : 0.46   Median : 6.11   Median :16.26   Median : 9.31
##  Mean   : 1.69   Mean   : 6.34   Mean   :15.85   Mean   : 9.29
##  3rd Qu.: 1.85   3rd Qu.: 7.52   3rd Qu.:17.16   3rd Qu.: 9.99
##  Max.   :20.00   Max.   :20.00   Max.   :20.00   Max.   :20.00
##  NA's   :48619   NA's   :49796   NA's   :49840   NA's   :49832
##       v68             v69             v70             v72
##  Min.   : 1.50   Min.   : 0.00   Min.   : 0.43   Min.   : 0.000
##  1st Qu.:17.08   1st Qu.: 8.39   1st Qu.:10.80   1st Qu.: 1.000
##  Median :18.27   Median : 9.52   Median :12.49   Median : 1.000
##  Mean   :17.56   Mean   : 9.45   Mean   :12.27   Mean   : 1.432
##  3rd Qu.:18.91   3rd Qu.:10.54   3rd Qu.:13.99   3rd Qu.: 2.000
##  Max.   :20.00   Max.   :20.00   Max.   :19.82   Max.   :12.000
##  NA's   :49836   NA's   :49895   NA's   :48636
##       v73             v76             v77             v78
##  Min.   : 0.00   Min.   : 0.00   Min.   : 0.00   Min.   : 0.00
##  1st Qu.: 1.90   1st Qu.: 1.63   1st Qu.: 6.50   1st Qu.:12.31
##  Median : 2.33   Median : 2.17   Median : 7.38   Median :13.33
##  Mean   : 2.43   Mean   : 2.41   Mean   : 7.31   Mean   :13.33
##  3rd Qu.: 2.85   3rd Qu.: 2.81   3rd Qu.: 8.16   3rd Qu.:14.39
##  Max.   :20.00   Max.   :20.00   Max.   :15.97   Max.   :20.00
##  NA's   :49836   NA's   :49796   NA's   :49832   NA's   :49895
##       v80             v81             v82             v83
##  Min.   : 0.00   Min.   : 0.00   Min.   : 0.00   Min.   : 0.00
##  1st Qu.: 1.45   1st Qu.: 5.98   1st Qu.: 3.39   1st Qu.: 1.43
##  Median : 2.09   Median : 7.52   Median : 3.69   Median : 1.94
##  Mean   : 2.21   Mean   : 7.29   Mean   : 6.21   Mean   : 2.17
##  3rd Qu.: 2.86   3rd Qu.: 8.78   3rd Qu.: 8.79   3rd Qu.: 2.67
##  Max.   :20.00   Max.   :20.00   Max.   :20.00   Max.   :20.00
##  NA's   :49851   NA's   :48624   NA's   :48624   NA's   :49832
##       v84             v85             v86             v87
##  Min.   : 0.00   Min.   : 0.00   Min.   : 0.00   Min.   : 0.87
##  1st Qu.: 0.96   1st Qu.: 1.84   1st Qu.: 0.94   1st Qu.: 8.06
##  Median : 1.42   Median : 2.59   Median : 1.16   Median :10.00
##  Mean   : 1.61   Mean   : 2.82   Mean   : 1.22   Mean   :10.18
##  3rd Qu.: 2.07   3rd Qu.: 3.57   3rd Qu.: 1.42   3rd Qu.:12.23
##  Max.   :20.00   Max.   :20.00   Max.   :17.56   Max.   :19.84
##  NA's   :49832   NA's   :50682   NA's   :49832   NA's   :48663
##       v88             v89             v90             v92
##  Min.   : 0.00   Min.   : 0.02   Min.   :0.00    Min.   :0.00
##  1st Qu.: 1.18   1st Qu.: 0.10   1st Qu.:0.86    1st Qu.:0.44
##  Median : 1.76   Median : 0.33   Median :0.97    Median :0.54
##  Mean   : 1.92   Mean   : 1.52   Mean   :0.97    Mean   :0.58
##  3rd Qu.: 2.46   3rd Qu.: 1.75   3rd Qu.:1.06    3rd Qu.:0.68
##  Max.   :20.00   Max.   :20.00   Max.   :6.31    Max.   :8.92
##  NA's   :49832   NA's   :48619   NA's   :49836   NA's   :49843
##       v93             v94             v95             v96
##  Min.   : 0.00   Min.   : 0.00   Min.   :0.00    Min.   : 0.00
##  1st Qu.: 4.55   1st Qu.: 3.33   1st Qu.:0.50    1st Qu.: 5.76
##  Median : 5.30   Median : 3.74   Median :0.62    Median : 6.51
```

```
##  Mean   : 5.48   Mean   : 3.85   Mean   :0.67    Mean   : 6.46
##  3rd Qu.: 6.22   3rd Qu.: 4.23   3rd Qu.:0.77    3rd Qu.: 7.23
##  Max.   :20.00   Max.   :19.02   Max.   :9.07    Max.   :20.00
##  NA's   :49832   NA's   :49832   NA's   :49843   NA's   :49832
##       v97             v98             v99             v100
##  Min.   : 0.00   Min.   : 0.00   Min.   : 0.00   Min.   : 0.00
##  1st Qu.: 6.32   1st Qu.: 6.13   1st Qu.: 0.93   1st Qu.: 5.75
##  Median : 7.45   Median : 7.64   Median : 1.24   Median :14.48
##  Mean   : 7.62   Mean   : 7.67   Mean   : 1.25   Mean   :12.09
##  3rd Qu.: 8.78   3rd Qu.: 9.06   3rd Qu.: 1.55   3rd Qu.:18.32
##  Max.   :20.00   Max.   :19.06   Max.   :20.00   Max.   :20.00
##  NA's   :49843   NA's   :48654   NA's   :49832   NA's   :49836
##       v101            v102            v103            v104
##  Min.   : 0.00   Min.   : 0.00   Min.   : 0.00   Min.   : 0.00
##  1st Qu.: 5.26   1st Qu.: 1.79   1st Qu.: 4.50   1st Qu.: 2.14
##  Median : 6.62   Median : 2.46   Median : 5.13   Median : 2.51
##  Mean   : 6.87   Mean   : 2.89   Mean   : 5.30   Mean   : 2.64
##  3rd Qu.: 8.24   3rd Qu.: 3.41   3rd Qu.: 5.87   3rd Qu.: 2.95
##  Max.   :20.00   Max.   :20.00   Max.   :18.78   Max.   :20.00
##  NA's   :49796   NA's   :51316   NA's   :49832   NA's   :49832
##       v105            v106            v108            v109
##  Min.   : 0.00   Min.   : 0.00   Min.   : 0.00   Min.   : 0.00
##  1st Qu.: 0.06   1st Qu.:10.05   1st Qu.: 1.53   1st Qu.: 1.84
##  Median : 0.24   Median :12.09   Median : 1.98   Median : 3.09
##  Mean   : 1.08   Mean   :11.79   Mean   : 2.15   Mean   : 4.18
##  3rd Qu.: 1.02   3rd Qu.:13.77   3rd Qu.: 2.54   3rd Qu.: 5.15
##  Max.   :20.00   Max.   :20.00   Max.   :20.00   Max.   :20.00
##  NA's   :48658   NA's   :49796   NA's   :48624   NA's   :48624
##       v111            v114            v115            v116
##  Min.   : 0.00   Min.   : 0.00   Min.   : 0.00   Min.   : 0.00
##  1st Qu.: 2.32   1st Qu.:12.00   1st Qu.: 9.46   1st Qu.: 1.85
##  Median : 3.11   Median :14.04   Median :10.48   Median : 2.22
##  Mean   : 3.37   Mean   :13.57   Mean   :10.55   Mean   : 2.29
##  3rd Qu.: 4.12   3rd Qu.:15.37   3rd Qu.:11.61   3rd Qu.: 2.65
##  Max.   :20.00   Max.   :20.00   Max.   :20.00   Max.   :20.00
##  NA's   :49832   NA's   :30      NA's   :49895   NA's   :49836
##       v117            v118            v119            v120
##  Min.   : 0.00   Min.   : 0.00   Min.   : 0.00   Min.   : 0.00
##  1st Qu.: 5.76   1st Qu.: 6.98   1st Qu.: 0.32   1st Qu.: 0.78
##  Median : 8.07   Median : 8.14   Median : 1.46   Median : 1.14
##  Mean   : 8.30   Mean   : 8.36   Mean   : 3.17   Mean   : 1.29
##  3rd Qu.:10.50   3rd Qu.: 9.57   3rd Qu.: 4.17   3rd Qu.: 1.65
##  Max.   :20.00   Max.   :20.00   Max.   :20.00   Max.   :10.39
##  NA's   :48624   NA's   :49843   NA's   :50680   NA's   :49836
##       v121            v122            v123            v124
##  Min.   : 0.00   Min.   : 0.00   Min.   : 0.02   Min.   : 0.00
##  1st Qu.: 1.79   1st Qu.: 5.65   1st Qu.: 1.96   1st Qu.: 0.02
##  Median : 2.44   Median : 6.75   Median : 2.74   Median : 0.14
##  Mean   : 2.74   Mean   : 6.82   Mean   : 3.55   Mean   : 0.92
##  3rd Qu.: 3.38   3rd Qu.: 7.91   3rd Qu.: 4.08   3rd Qu.: 0.87
##  Max.   :20.00   Max.   :20.00   Max.   :19.69   Max.   :20.00
##  NA's   :49840   NA's   :49851   NA's   :50678   NA's   :48619
##       v126            v127            v128            v129
##  Min.   : 0.00   Min.   : 0.00   Min.   : 0.00   Min.   : 0.0000
```

```
##  1st Qu.: 1.42   1st Qu.: 2.10   1st Qu.: 1.39   1st Qu.: 0.0000
##  Median : 1.61   Median : 2.96   Median : 1.80   Median : 0.0000
##  Mean   : 1.67   Mean   : 3.24   Mean   : 2.03   Mean   : 0.3101
##  3rd Qu.: 1.84   3rd Qu.: 4.11   3rd Qu.: 2.39   3rd Qu.: 0.0000
##  Max.   :15.63   Max.   :20.00   Max.   :20.00   Max.   :11.0000
##  NA's   :49832   NA's   :49832   NA's   :48624
##       v130            v131
##  Min.   : 0.00   Min.   : 0.00
##  1st Qu.: 1.11   1st Qu.: 1.01
##  Median : 1.56   Median : 1.59
##  Mean   : 1.93   Mean   : 1.74
##  3rd Qu.: 2.33   3rd Qu.: 2.26
##  Max.   :20.00   Max.   :20.00
##  NA's   :49843   NA's   :49895
```

```
# summary for factor columns
summary(data[,-(numericCols)])
```

```
##  v3              v22         v24            v30        v31
##   :  3457   AGDF   :  2386   A: 3789          :60110    : 3457
##  A:   227   YGJ    :  2119   B: 8150   C      :32178   A:88347
##  B:    53   QKI    :   668   C:20872   G      : 8728   B:18947
##  C:110584   PWR    :   649   D:26333   D      : 5225   C: 3570
##                     :   500   E:55177   E      : 2973
##             HZE    :   423             F      : 2589
##             (Other):107576             (Other): 2518
##       v47             v52             v56        v66             v71
##  C      :55425   J      :11103   BW     :11351   A:70353   F      :75094
##  I      :39071   I      :10260   DI     :10256   B:18264   B      :30255
##  E      : 5301   F      : 9806   AS     : 8832   C:25704   C      : 8947
##  F      : 4322   C      : 9681   BZ     : 7174             I      :   16
##  G      : 3946   D      : 9607          : 6882             G      :    5
##  D      : 3157   L      : 9578   AW     : 6369             A      :    1
##  (Other): 3099   (Other):54286   (Other):63457             (Other):    3
##  v74         v75             v79             v91             v107
##  A:    45   A:   18   C      :34561   A      :27079   E      :27079
##  B:113560   B:39192   B      :25801   G      :24545   C      :24545
##  C:   716   C:   24   E      :25257   C      :23157   D      :23157
##             D:75087   D      : 5302   B      :22683   B      :22683
##                       I      : 4561   F      :13418   A      :13418
##                       K      : 4308   E      : 3206   F      : 3206
##                       (Other):14531   (Other):  233   (Other):  233
##  v110            v112            v113            v125
##  A:55688   F      :21671          :55304   BM     : 5759
##  B:55426   I      :10224   G      :16252   AK     : 5337
##  C: 3207   A      : 9545   M      : 7374   BJ     : 4465
##            N      : 9086   AC     : 5956   CG     : 3826
##            D      : 7327   AF     : 3568   AP     : 3410
##            H      : 5651   I      : 2605   BY     : 3311
##            (Other):50817   (Other):23262   (Other):88213
```

```
# fractionof ones in target
sum(data$target)/dim(data)[1]
```

```
## [1] 0.7611987
```

```
# partition data for later exploration (by target fractions)
library(caret)
```

```
## Loading required package: lattice
## Loading required package: ggplot2
```

```
set.seed(1959)
inTrain = createDataPartition(data$target, p = 2/3, list = FALSE)
training = data[inTrain,]
testing = data[-inTrain,]
```