

# Caret Package Demo - SPAM Example

*Jiachang (Ernest) Xu*

*6/22/2017*

## SPAM Example: Data Splitting

```
## require caret package for machine learning algorithms
require(caret)

## Loading required package: caret
## Loading required package: lattice
## Loading required package: ggplot2
## require kernlab for spam data
require(kernlab)

## Loading required package: kernlab
##
## Attaching package: 'kernlab'
## The following object is masked from 'package:ggplot2':
##
##      alpha
## data loading
data(spam)
head(spam)

##      make address  all num3d  our over remove internet order mail receive
## 1 0.00      0.64 0.64      0 0.32 0.00      0.00      0.00 0.00 0.00      0.00
## 2 0.21      0.28 0.50      0 0.14 0.28      0.21      0.07 0.00 0.94      0.21
## 3 0.06      0.00 0.71      0 1.23 0.19      0.19      0.12 0.64 0.25      0.38
## 4 0.00      0.00 0.00      0 0.63 0.00      0.31      0.63 0.31 0.63      0.31
## 5 0.00      0.00 0.00      0 0.63 0.00      0.31      0.63 0.31 0.63      0.31
## 6 0.00      0.00 0.00      0 1.85 0.00      0.00      1.85 0.00 0.00      0.00
##      will people report addresses free business email  you credit your font
## 1 0.64      0.00      0.00      0.00 0.32      0.00 1.29 1.93      0.00 0.96      0
## 2 0.79      0.65      0.21      0.14 0.14      0.07 0.28 3.47      0.00 1.59      0
## 3 0.45      0.12      0.00      1.75 0.06      0.06 1.03 1.36      0.32 0.51      0
## 4 0.31      0.31      0.00      0.00 0.31      0.00 0.00 3.18      0.00 0.31      0
## 5 0.31      0.31      0.00      0.00 0.31      0.00 0.00 3.18      0.00 0.31      0
## 6 0.00      0.00      0.00      0.00 0.00      0.00 0.00 0.00      0.00 0.00      0
##      num000 money hp hpl george num650 lab labs telnet num857 data num415
## 1 0.00 0.00 0 0      0      0 0 0      0      0 0      0
## 2 0.43 0.43 0 0      0      0 0 0      0      0 0      0
## 3 1.16 0.06 0 0      0      0 0 0      0      0 0      0
## 4 0.00 0.00 0 0      0      0 0 0      0      0 0      0
## 5 0.00 0.00 0 0      0      0 0 0      0      0 0      0
## 6 0.00 0.00 0 0      0      0 0 0      0      0 0      0
##      num85 technology num1999 parts pm direct cs meeting original project
## 1 0      0      0.00      0 0      0.00 0      0      0.00      0
```

```
## 2      0      0      0.07      0 0      0.00 0      0      0.00      0
## 3      0      0      0.00      0 0      0.06 0      0      0.12      0
## 4      0      0      0.00      0 0      0.00 0      0      0.00      0
## 5      0      0      0.00      0 0      0.00 0      0      0.00      0
## 6      0      0      0.00      0 0      0.00 0      0      0.00      0
##      re  edu table conference charSemicolon charRoundbracket
## 1 0.00 0.00      0      0      0.00      0.000
## 2 0.00 0.00      0      0      0.00      0.132
## 3 0.06 0.06      0      0      0.01      0.143
## 4 0.00 0.00      0      0      0.00      0.137
## 5 0.00 0.00      0      0      0.00      0.135
## 6 0.00 0.00      0      0      0.00      0.223
##      charSquarebracket charExclamation charDollar charHash capitalAve
## 1      0      0.778      0.000      0.000      3.756
## 2      0      0.372      0.180      0.048      5.114
## 3      0      0.276      0.184      0.010      9.821
## 4      0      0.137      0.000      0.000      3.537
## 5      0      0.135      0.000      0.000      3.537
## 6      0      0.000      0.000      0.000      3.000
##      capitalLong capitalTotal type
## 1      61      278 spam
## 2     101     1028 spam
## 3     485     2259 spam
## 4      40      191 spam
## 5      40      191 spam
## 6      15       54 spam
## splitting data to training and testing
inTrain <- createDataPartition(y = spam$type, p = 0.75, list = FALSE)
training <- spam[inTrain, ]
testing <- spam[-inTrain, ]
dim(training)
```

```
## [1] 3451  58
```

## SPAM Example: Fit a Model (Generalized Linear Model)

```
## generalized linear model
set.seed(32343)
suppressWarnings(model.fit <- train(type ~ ., data = training, method = "glm"))
model.fit

## Generalized Linear Model
##
## 3451 samples
##  57 predictor
##  2 classes: 'nonspam', 'spam'
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 3451, 3451, 3451, 3451, 3451, 3451, ...
## Resampling results:
##
## Accuracy   Kappa
```

```
## 0.9168474 0.8258369
```

## SPAM Example: Final Model

```
model.fit$finalModel
```

```
##
## Call:  NULL
##
## Coefficients:
##      (Intercept)          make          address
##      -1.636e+00      -3.781e-01      -1.461e-01
##           all          num3d           our
##           1.850e-01          1.868e+00          6.128e-01
##           over          remove          internet
##           7.700e-01          1.987e+00          5.722e-01
##           order          mail          receive
##           5.130e-01          1.298e-01      -3.628e-01
##           will          people          report
##           -1.068e-01      -8.317e-02          9.529e-02
##           addresses          free          business
##           8.410e-01          9.922e-01          7.765e-01
##           email          you          credit
##           9.400e-02          1.170e-01          2.174e+00
##           your          font          num000
##           2.547e-01          2.106e-01          2.848e+00
##           money          hp          hpl
##           3.493e-01      -1.961e+00      -1.221e+00
##           george          num650          lab
##           -9.425e+00          3.851e-01      -4.324e+00
##           labs          telnet          num857
##           -2.578e-01      -5.449e+00          1.985e+00
##           data          num415          num85
##           -4.398e-01      -1.316e+01      -1.904e+00
##           technology          num1999          parts
##           9.701e-01          1.281e-01      -5.586e-01
##           pm          direct          cs
##           -7.414e-01      -3.307e-01      -4.526e+01
##           meeting          original          project
##           -2.453e+00      -2.742e+00      -1.438e+00
##           re          edu          table
##           -8.179e-01      -1.486e+00      -2.668e+00
##           conference      charSemicolon      charRoundbracket
##           -3.614e+00      -1.263e+00      -1.739e-01
##           charSquarebracket      charExclamation      charDollar
##           -4.183e-01          3.426e-01          5.249e+00
##           charHash          capitalAve          capitalLong
##           2.066e+00      -1.984e-02          1.234e-02
##           capitalTotal
##           6.717e-04
##
## Degrees of Freedom: 3450 Total (i.e. Null); 3393 Residual
## Null Deviance:      4628
```

```
## Residual Deviance: 1391 AIC: 1507
```

## SPAM Example: Prediction

```
## prediction
prediction <- predict(model.fit, newdata = testing)
prediction
```

```
##      [1] spam      spam      spam      spam      spam      nonspam nonspam spam
##      [9] spam      nonspam spam      spam      spam      spam      spam      spam
##     [17] spam      spam      spam      spam      spam      spam      nonspam spam
##     [25] spam      spam      spam      spam      nonspam spam      spam      nonspam
##     [33] spam      spam      spam      spam      spam      spam      spam      spam
##     [41] spam      spam      spam      spam      spam      spam      spam      spam
##     [49] spam      nonspam spam      spam      nonspam spam      spam      spam
##     [57] spam      spam      nonspam spam      spam      spam      spam      spam
##     [65] spam      spam      spam      spam      spam      spam      spam      spam
##     [73] spam      nonspam spam      spam      spam      spam      spam      spam
##     [81] spam      spam      spam      spam      spam      spam      nonspam spam
##     [89] spam      spam      spam      spam      spam      spam      spam      spam
##     [97] spam      spam      spam      spam      spam      spam      nonspam spam
##    [105] spam      spam      spam      nonspam spam      nonspam spam      spam
##    [113] nonspam nonspam spam      spam      spam      spam      spam      spam
##    [121] spam      spam      spam      spam      nonspam spam      spam      spam
##    [129] spam      spam      nonspam spam      spam      spam      spam      spam
##    [137] spam      spam      spam      spam      spam      spam      spam      spam
##    [145] spam      spam      spam      spam      spam      spam      spam      spam
##    [153] spam      spam      spam      spam      spam      spam      nonspam spam
##    [161] spam      spam      spam      spam      spam      spam      spam      spam
##    [169] spam      nonspam spam      spam      spam      spam      spam      nonspam
##    [177] spam      spam      spam      spam      nonspam spam      spam      spam
##    [185] spam      nonspam spam      spam      spam      spam      spam      spam
##    [193] spam      spam      spam      nonspam spam      spam      spam      spam
##    [201] spam      spam      spam      spam      spam      spam      spam      spam
##    [209] spam      spam      spam      spam      spam      spam      spam      spam
##    [217] spam      spam      spam      spam      spam      spam      spam      spam
##    [225] spam      spam      nonspam spam      spam      spam      spam      spam
##    [233] spam      spam      spam      spam      spam      spam      spam      spam
##    [241] spam      spam      spam      spam      spam      nonspam nonspam spam
##    [249] spam      spam      spam      spam      spam      spam      spam      spam
##    [257] spam      spam      spam      spam      spam      spam      spam      spam
##    [265] spam      spam      spam      spam      spam      spam      spam      spam
##    [273] spam      spam      spam      spam      nonspam spam      nonspam spam
##    [281] spam      spam      spam      spam      spam      spam      nonspam spam
##    [289] spam      spam      spam      spam      spam      nonspam spam      nonspam
##    [297] spam      spam      spam      spam      spam      spam      spam      spam
##    [305] spam      spam      nonspam spam      nonspam spam      spam      spam
##    [313] spam      spam      spam      spam      spam      spam      spam      spam
##    [321] spam      spam      spam      spam      spam      nonspam spam      spam
##    [329] spam      spam      spam      spam      nonspam spam      spam      spam
##    [337] spam      spam      spam      spam      spam      spam      nonspam spam
##    [345] spam      spam      spam      spam      spam      spam      spam      spam
##    [353] nonspam spam      nonspam spam      spam      spam      spam      spam
```

[illegible]

```

## [793] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [801] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [809] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [817] nonspam nonspam spam nonspam nonspam nonspam nonspam nonspam nonspam
## [825] nonspam spam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [833] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [841] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [849] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [857] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [865] nonspam nonspam nonspam nonspam nonspam nonspam nonspam spam nonspam
## [873] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [881] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [889] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [897] nonspam nonspam nonspam nonspam nonspam spam nonspam nonspam nonspam
## [905] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [913] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [921] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [929] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [937] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [945] nonspam spam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [953] nonspam nonspam spam nonspam nonspam nonspam nonspam nonspam nonspam
## [961] nonspam nonspam nonspam nonspam spam nonspam nonspam nonspam nonspam
## [969] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [977] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [985] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [993] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [1001] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [1009] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [1017] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [1025] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [1033] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [1041] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [1049] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [1057] nonspam spam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [1065] spam spam spam nonspam nonspam nonspam nonspam nonspam nonspam
## [1073] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [1081] nonspam spam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [1089] nonspam nonspam nonspam nonspam nonspam spam nonspam nonspam nonspam
## [1097] nonspam nonspam nonspam spam nonspam nonspam spam nonspam nonspam
## [1105] nonspam spam nonspam nonspam nonspam spam nonspam nonspam nonspam
## [1113] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [1121] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [1129] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [1137] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## [1145] nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam nonspam
## Levels: nonspam spam

```

## SPAM Example: Confusion Matrix

```
confusionMatrix(data = prediction, testing$type)
```

```
## Confusion Matrix and Statistics
##
```

```

##           Reference
## Prediction nonspam spam
##   nonspam    669   54
##   spam       28  399
##
##           Accuracy : 0.9287
##           95% CI : (0.9123, 0.9429)
##   No Information Rate : 0.6061
##   P-Value [Acc > NIR] : < 2.2e-16
##
##           Kappa : 0.8492
## Mcnemar's Test P-Value : 0.005766
##
##           Sensitivity : 0.9598
##           Specificity : 0.8808
##           Pos Pred Value : 0.9253
##           Neg Pred Value : 0.9344
##           Prevalence : 0.6061
##           Detection Rate : 0.5817
##   Detection Prevalence : 0.6287
##           Balanced Accuracy : 0.9203
##
##           'Positive' Class : nonspam
##

```