Subset Selection

William Staudenmeier

April 28, 2018

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(tidyr)
library(caret)
```

```
## Warning: package 'caret' was built under R version 3.4.4
```

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

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

```r
library(DMwR)
```

```
## Warning: package 'DMwR' was built under R version 3.4.4
```

```
## Loading required package: grid
```

```r
library(ROCR)
```

```
## Warning: package 'ROCR' was built under R version 3.4.4
```

```
## Loading required package: gplots
```

```
## Warning: package 'gplots' was built under R version 3.4.4
```

```
##
## Attaching package: 'gplots'
```

```
## The following object is masked from 'package:stats':
##
##     lowess
```

```r
require(caTools)
```

```
## Loading required package: caTools
```

```r
library(ISLR)
library(ggplot2)
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
```

```
##      select
```
```r
library(leaps)
```
```
## Warning: package 'leaps' was built under R version 3.4.4
```
```r
library(bestglm)
library(glmnet)
```
```
## Loading required package: Matrix
```
```
##
## Attaching package: 'Matrix'
```
```
## The following object is masked from 'package:tidyr':
##
##      expand
```
```
## Loading required package: foreach
```
```
## Loaded glmnet 2.0-13
```
```r
library(Matrix)


train = read.csv("expert variables after ks.csv",stringsAsFactors = F)



best.subset <- regsubsets(Fraud ~ Amount.1 + Amount_week_diff+ Amount_month_diff +
                            same_Cardnum_diff_Merch_Description_1 +
                            same_Cardnum_diff_amount_1  +
                            same_Cardnum_diff_Merchant_State_1 +
                            same_Cardnum_diff_Merchant_Zip_1  +
                            same_amount_3 +
                            same_amount_5 +
                            same_amount_7 +
                            same_amount_14 +
                            same_amount_30  +
                            same_amount_diff_Merch_Description_3 +
                            same_amount_diff_Merch_Description_5 +
                            same_amount_diff_Merch_Description_7 +
                            same_amount_diff_Merch_Description_14 +
                            same_amount_diff_Merch_Description_30 +
                            same_amount_diff_Cardnum_3 +
                            same_amount_diff_Cardnum_5 +
                            same_amount_diff_Cardnum_7 +
                            same_amount_diff_Cardnum_14 +
                            same_amount_diff_Cardnum_30 +
                            same_amount_diff_Merchant_State_3 +
                            same_amount_diff_Merchant_State_5.y +
                            same_amount_diff_Merchant_State_7 +
                            same_amount_diff_Merchant_State_14 +
                            same_amount_diff_Merchant_State_30 +
                            same_amount_diff_Merchant_Zip_3 +
                            same_amount_diff_Merchant_Zip_7 +
                            same_amount_diff_Merchant_Zip_14 +
```

```
                            same_amount_diff_Merchant_Zip_30, train, method="exhaustive")
```

```
## Warning in leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax,
## force.in = force.in, : 5 linear dependencies found
```

```
## Reordering variables and trying again:
```

```
summary(best.subset)
```

```
## Subset selection object
## Call: regsubsets.formula(Fraud ~ Amount.1 + Amount_week_diff + Amount_month_diff +
##     same_Cardnum_1 + same_Cardnum_diff_Merch_Description_1 +
##     same_Cardnum_diff_amount_1 + same_Cardnum_diff_Merchant_State_1 +
##     same_Cardnum_diff_Merchant_Zip_1 + same_amount_3 + same_amount_5 +
##     same_amount_7 + same_amount_14 + same_amount_30 + same_amount_diff_Merch_Description_3 +
##     same_amount_diff_Merch_Description_5 + same_amount_diff_Merch_Description_7 +
##     same_amount_diff_Merch_Description_14 + same_amount_diff_Merch_Description_30 +
##     same_amount_diff_Cardnum_3 + same_amount_diff_Cardnum_5 +
##     same_amount_diff_Cardnum_7 + same_amount_diff_Cardnum_14 +
##     same_amount_diff_Cardnum_30 + same_amount_diff_Merchant_State_3 +
##     same_amount_diff_Merchant_State_5.y + same_amount_diff_Merchant_State_7 +
##     same_amount_diff_Merchant_State_14 + same_amount_diff_Merchant_State_30 +
##     same_amount_diff_Merchant_Zip_3 + same_amount_diff_Merchant_Zip_7 +
##     same_amount_diff_Merchant_Zip_14 + same_amount_diff_Merchant_Zip_30,
##     train, method = "exhaustive")
## 32 Variables  (and intercept)
##                                         Forced in Forced out
## Amount.1                                    FALSE      FALSE
## Amount_week_diff                            FALSE      FALSE
## Amount_month_diff                           FALSE      FALSE
## same_Cardnum_1                              FALSE      FALSE
## same_Cardnum_diff_Merch_Description_1       FALSE      FALSE
## same_Cardnum_diff_amount_1                  FALSE      FALSE
## same_Cardnum_diff_Merchant_State_1          FALSE      FALSE
## same_Cardnum_diff_Merchant_Zip_1            FALSE      FALSE
## same_amount_3                               FALSE      FALSE
## same_amount_5                               FALSE      FALSE
## same_amount_7                               FALSE      FALSE
## same_amount_14                              FALSE      FALSE
## same_amount_diff_Merch_Description_3        FALSE      FALSE
## same_amount_diff_Merch_Description_5        FALSE      FALSE
## same_amount_diff_Merch_Description_7        FALSE      FALSE
## same_amount_diff_Merch_Description_14       FALSE      FALSE
## same_amount_diff_Cardnum_3                  FALSE      FALSE
## same_amount_diff_Cardnum_5                  FALSE      FALSE
## same_amount_diff_Cardnum_7                  FALSE      FALSE
## same_amount_diff_Cardnum_14                 FALSE      FALSE
## same_amount_diff_Merchant_State_3           FALSE      FALSE
## same_amount_diff_Merchant_State_5.y         FALSE      FALSE
## same_amount_diff_Merchant_State_7           FALSE      FALSE
## same_amount_diff_Merchant_State_14          FALSE      FALSE
## same_amount_diff_Merchant_Zip_3             FALSE      FALSE
## same_amount_diff_Merchant_Zip_7             FALSE      FALSE
## same_amount_diff_Merchant_Zip_14            FALSE      FALSE
## same_amount_30                              FALSE      FALSE
```

```
## same_amount_diff_Merch_Description_30     FALSE     FALSE
## same_amount_diff_Cardnum_30               FALSE     FALSE
## same_amount_diff_Merchant_State_30        FALSE     FALSE
## same_amount_diff_Merchant_Zip_30          FALSE     FALSE
## 1 subsets of each size up to 9
## Selection Algorithm: exhaustive
##            Amount.1 Amount_week_diff Amount_month_diff same_Cardnum_1
## 1  ( 1 ) " "        "*"              " "               " "
## 2  ( 1 ) " "        "*"              " "               " "
## 3  ( 1 ) " "        "*"              " "               " "
## 4  ( 1 ) " "        "*"              " "               "*"
## 5  ( 1 ) " "        "*"              " "               " "
## 6  ( 1 ) " "        "*"              " "               "*"
## 7  ( 1 ) " "        "*"              " "               "*"
## 8  ( 1 ) " "        "*"              " "               "*"
## 9  ( 1 ) " "        "*"              " "               "*"
##            same_Cardnum_diff_Merch_Description_1 same_Cardnum_diff_amount_1
## 1  ( 1 ) " "                                     " "
## 2  ( 1 ) " "                                     "*"
## 3  ( 1 ) "*"                                     "*"
## 4  ( 1 ) " "                                     "*"
## 5  ( 1 ) "*"                                     "*"
## 6  ( 1 ) "*"                                     "*"
## 7  ( 1 ) "*"                                     "*"
## 8  ( 1 ) "*"                                     "*"
## 9  ( 1 ) "*"                                     "*"
##            same_Cardnum_diff_Merchant_State_1
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
## 4  ( 1 ) " "
## 5  ( 1 ) "*"
## 6  ( 1 ) "*"
## 7  ( 1 ) "*"
## 8  ( 1 ) "*"
## 9  ( 1 ) "*"
##            same_Cardnum_diff_Merchant_Zip_1 same_amount_3 same_amount_5
## 1  ( 1 ) " "                                " "           " "
## 2  ( 1 ) " "                                " "           " "
## 3  ( 1 ) " "                                " "           " "
## 4  ( 1 ) "*"                                " "           " "
## 5  ( 1 ) "*"                                " "           " "
## 6  ( 1 ) "*"                                " "           " "
## 7  ( 1 ) "*"                                " "           " "
## 8  ( 1 ) "*"                                " "           "*"
## 9  ( 1 ) "*"                                " "           "*"
##            same_amount_7 same_amount_14 same_amount_30
## 1  ( 1 ) " "             " "            " "
## 2  ( 1 ) " "             " "            " "
## 3  ( 1 ) " "             " "            " "
## 4  ( 1 ) " "             " "            " "
## 5  ( 1 ) " "             " "            " "
## 6  ( 1 ) " "             " "            " "
## 7  ( 1 ) " "             " "            " "
```

```
## 8  ( 1 ) " "             " "             " "
## 9  ( 1 ) " "             " "             " "
##            same_amount_diff_Merch_Description_3
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
## 4  ( 1 ) " "
## 5  ( 1 ) " "
## 6  ( 1 ) " "
## 7  ( 1 ) "*"
## 8  ( 1 ) "*"
## 9  ( 1 ) "*"
##            same_amount_diff_Merch_Description_5
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
## 4  ( 1 ) " "
## 5  ( 1 ) " "
## 6  ( 1 ) " "
## 7  ( 1 ) " "
## 8  ( 1 ) " "
## 9  ( 1 ) " "
##            same_amount_diff_Merch_Description_7
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
## 4  ( 1 ) " "
## 5  ( 1 ) " "
## 6  ( 1 ) " "
## 7  ( 1 ) " "
## 8  ( 1 ) " "
## 9  ( 1 ) " "
##            same_amount_diff_Merch_Description_14
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
## 4  ( 1 ) " "
## 5  ( 1 ) " "
## 6  ( 1 ) " "
## 7  ( 1 ) " "
## 8  ( 1 ) " "
## 9  ( 1 ) " "
##            same_amount_diff_Merch_Description_30 same_amount_diff_Cardnum_3
## 1  ( 1 ) " "                                     " "
## 2  ( 1 ) " "                                     " "
## 3  ( 1 ) " "                                     " "
## 4  ( 1 ) " "                                     " "
## 5  ( 1 ) " "                                     " "
## 6  ( 1 ) " "                                     " "
## 7  ( 1 ) " "                                     " "
## 8  ( 1 ) " "                                     " "
## 9  ( 1 ) " "                                     "*"
##            same_amount_diff_Cardnum_5 same_amount_diff_Cardnum_7
## 1  ( 1 ) " "                          " "
```

```
## 2  ( 1 ) " "                                " "
## 3  ( 1 ) " "                                " "
## 4  ( 1 ) " "                                " "
## 5  ( 1 ) " "                                " "
## 6  ( 1 ) " "                                " "
## 7  ( 1 ) " "                                " "
## 8  ( 1 ) " "                                " "
## 9  ( 1 ) " "                                " "
##          same_amount_diff_Cardnum_14 same_amount_diff_Cardnum_30
## 1  ( 1 ) " "                                " "
## 2  ( 1 ) " "                                " "
## 3  ( 1 ) " "                                " "
## 4  ( 1 ) " "                                " "
## 5  ( 1 ) " "                                " "
## 6  ( 1 ) " "                                " "
## 7  ( 1 ) " "                                " "
## 8  ( 1 ) " "                                " "
## 9  ( 1 ) " "                                " "
##          same_amount_diff_Merchant_State_3
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
## 4  ( 1 ) " "
## 5  ( 1 ) " "
## 6  ( 1 ) " "
## 7  ( 1 ) " "
## 8  ( 1 ) " "
## 9  ( 1 ) " "
##          same_amount_diff_Merchant_State_5.y
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
## 4  ( 1 ) " "
## 5  ( 1 ) " "
## 6  ( 1 ) " "
## 7  ( 1 ) " "
## 8  ( 1 ) " "
## 9  ( 1 ) " "
##          same_amount_diff_Merchant_State_7
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
## 4  ( 1 ) " "
## 5  ( 1 ) " "
## 6  ( 1 ) " "
## 7  ( 1 ) " "
## 8  ( 1 ) " "
## 9  ( 1 ) " "
##          same_amount_diff_Merchant_State_14
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
## 4  ( 1 ) " "
## 5  ( 1 ) " "
```

```
## 6  ( 1 ) " "
## 7  ( 1 ) " "
## 8  ( 1 ) " "
## 9  ( 1 ) " "
##           same_amount_diff_Merchant_State_30
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
## 4  ( 1 ) " "
## 5  ( 1 ) " "
## 6  ( 1 ) " "
## 7  ( 1 ) " "
## 8  ( 1 ) " "
## 9  ( 1 ) " "
##           same_amount_diff_Merchant_Zip_3 same_amount_diff_Merchant_Zip_7
## 1  ( 1 ) " "                               " "
## 2  ( 1 ) " "                               " "
## 3  ( 1 ) " "                               " "
## 4  ( 1 ) " "                               " "
## 5  ( 1 ) " "                               " "
## 6  ( 1 ) " "                               " "
## 7  ( 1 ) " "                               " "
## 8  ( 1 ) " "                               " "
## 9  ( 1 ) " "                               " "
##           same_amount_diff_Merchant_Zip_14 same_amount_diff_Merchant_Zip_30
## 1  ( 1 ) " "                                " "
## 2  ( 1 ) " "                                " "
## 3  ( 1 ) " "                                " "
## 4  ( 1 ) " "                                " "
## 5  ( 1 ) " "                                " "
## 6  ( 1 ) " "                                " "
## 7  ( 1 ) " "                                " "
## 8  ( 1 ) " "                                " "
## 9  ( 1 ) " "                                " "
```

```r
best.subset.summary = summary(best.subset)

best.subset.by.adjr2 = which.max(best.subset.summary$adjr2)
best.subset.by.adjr2
```

```
## [1] 9
```

```r
best.subset.by.cp = which.min(best.subset.summary$cp)
best.subset.by.cp
```

```
## [1] 9
```

```r
best.subset.by.bic = which.min(best.subset.summary$bic)
best.subset.by.bic
```

```
## [1] 8
```