Demo April 18: ~~Classification~~ Predictions

Abhishek Nandy

April 18, 2016

Election Prediction example

Think back to the 2008 elections, when it seemed Hillary Clinton's momentum was picking up again; the data set concerns all the counties in states where primaries had already been held The unit of observation, then, is a county and variables included various demographic measures (age and ethnic makeup, education level, religious breakdown), political measures (did the county go to Bush or Kerry in 04) and economic factors (unemployment rate, the amount of construction in the county), and so on

```
set.seed(55)
require(ElemStatLearn)
```

```
## Loading required package: ElemStatLearn
```

```
primaries <- read.csv("~/Documents/Stat3022/Computing/primaries.csv")
head(primaries)
```

```
##   fips county_name state_postal region election_date racetype tvotes
## 1 1001     Autauga           AL      S        2/5/08  Primary   4118
## 2 1003     Baldwin           AL      S        2/5/08  Primary  12085
## 3 1005     Barbour           AL      S        2/5/08  Primary   3823
## 4 1007        Bibb           AL      S        2/5/08  Primary   1751
## 5 1009      Blount           AL      S        2/5/08  Primary   3471
## 6 1011     Bullock           AL      S        2/5/08  Primary   2540
##   clinton obama edwards      margin  winner POP05_SQMI popUnder30_00
## 1    1760  2268      57  0.12336086   obama       78.9          30.3
## 2    6259  5450     227 -0.06694249 clinton       97.0          27.2
## 3    1322  2393      60  0.28014648   obama       32.3          31.4
## 4     922   755      44 -0.09537407 clinton       33.4          32.8
## 5    2735   617      75 -0.61019879 clinton       80.7          30.2
## 6     471  2032       6  0.61456693   obama       18.5          31.6
##   pop65up_00 presVote04   kerry04    Bush04 pres04margin pres04winner
## 1       10.2      20081 0.2369404 0.7567352   0.51979483         bush
## 2       15.5      69320 0.2250289 0.7641518   0.53912291         bush
## 3       13.3      10777 0.4483623 0.5473694   0.09900715         bush
## 4       11.6       7600 0.2748684 0.7200000   0.44513158         bush
## 5       12.9      21504 0.1831287 0.8085007   0.62537202         bush
## 6       13.2       4717 0.6805173 0.3167267  -0.36379055        kerry
##    pop06  pop00 hisp06 white06 black06 indian06 asian06 hawaii06 mixed06
## 1  49730  43911    827   39368    8559      203     271       12     490
## 2 169162 141423   4176  145687   16301      770     679       42    1507
## 3  28171  29044    953   13793   13035      115      84        6     185
## 4  21482  19939    304   16266    4705       63      25        2     117
## 5  56436  51229   3752   51031     872      238     125        7     411
## 6  10906  11613    752    2413    7615       38      22        1      65
##   pct_less_30k pct_more_100k pct_hs_grad pct_labor_force pct_homeowner
```

```
## 1       0.3418    0.07594540      0.7872          0.6515        0.8083
## 2       0.3582    0.09493099      0.8202          0.5982        0.7958
## 3       0.5682    0.05032592      0.6465          0.4802        0.7316
## 4       0.4746    0.03521604      0.6319          0.5289        0.8019
## 5       0.4149    0.05012270      0.7045          0.6057        0.8349
## 6       0.6361    0.02579514      0.6048          0.4160        0.7444
##   unempFeb07 unempFeb08 unempChg pctUnins00 subForPctHomes poverty05
## 1        3.0        3.6      0.6       12.8           0.07      10.4
## 2        3.1        3.4      0.3       13.8           0.00      11.4
## 3        5.0        7.1      2.1       19.4           6.40      22.4
## 4        3.7        4.2      0.5       16.5           0.10      16.6
## 5        3.1        3.3      0.2       15.7           0.04      11.4
## 6        7.6        7.3     -0.3       22.9           0.00      38.2
##   median_hhi05    Catholic So.Bapt.Conv Un.Methodist    E.L.C.A.
## 1        45019 0.034118752    0.3372261   0.07567951 0.000000000
## 2        42804 0.074650144    0.1979062   0.07045544 0.001068262
## 3        29534 0.009125973    0.2608995   0.06236655 0.000000000
## 4        34212 0.000000000    0.3862480   0.02367233 0.000000000
## 5        40588 0.007369081    0.3558717   0.07118219 0.000000000
## 6        21728 0.002987878    0.1464914   0.03585453 0.000000000
##   Construction Manufacturing FinancialActivities GoodsProducing
## 1     6.797467     17.577510           5.366229       26.77624
## 2    10.558143      9.230177           7.923872       21.28236
## 3     2.501616     45.067103           3.379843       51.27552
## 4    20.750603     15.723631           3.477562       42.55710
## 5     9.754604     22.763883           4.300316       34.12934
## 6     0.000000      0.000000           3.523767       56.86010
##   ServiceProviding
## 1         73.22376
## 2         78.71764
## 3         48.72448
## 4         57.44290
## 5         65.87066
## 6         43.13990
```

The description of the variables were inadequate.

```
colnames(primaries)
```

```
##  [1] "fips"              "county_name"       "state_postal"
##  [4] "region"            "election_date"     "racetype"
##  [7] "tvotes"            "clinton"           "obama"
## [10] "edwards"           "margin"            "winner"
## [13] "POP05_SQMI"        "popUnder30_00"     "pop65up_00"
## [16] "presVote04"        "kerry04"           "Bush04"
## [19] "pres04margin"      "pres04winner"      "pop06"
## [22] "pop00"             "hisp06"            "white06"
## [25] "black06"           "indian06"          "asian06"
## [28] "hawaii06"          "mixed06"           "pct_less_30k"
## [31] "pct_more_100k"     "pct_hs_grad"       "pct_labor_force"
## [34] "pct_homeowner"     "unempFeb07"        "unempFeb08"
## [37] "unempChg"          "pctUnins00"        "subForPctHomes"
## [40] "poverty05"         "median_hhi05"      "Catholic"
## [43] "So.Bapt.Conv"      "Un.Methodist"      "E.L.C.A."
## [46] "Construction"      "Manufacturing"     "FinancialActivities"
## [49] "GoodsProducing"    "ServiceProviding"
```

"state_postal" → state names (two digit)
"tvotes" → total votes
"racetype" → primary / caucus
"clinton" → votes for clinton
"obama" → votes for Obama.
"edwards" → votes for Edwards
"margin" → win margin
"winner" → obama / clinton / edwards.
"pop65up_00" → percentage of pop. above 65 age
"pct_less_30k" → percentage of people earning less than 30k
"unempFeb08" → number / percentage of unemployed people.

↓
Goods producing Industry.

```
tail(primaries)
```

```
##       fips county_name state_postal region election_date racetype tvotes
## 2445 56035    Sublette           WY      W        3/8/08   Caucus     78
## 2446 56037  Sweetwater           WY      W        3/8/08   Caucus    596
## 2447 56039       Teton           WY      W        3/8/08   Caucus   1150
## 2448 56041       Uinta           WY      W        3/8/08   Caucus    168
## 2449 56043    Washakie           WY      W        3/8/08   Caucus     98
## 2450 56045      Weston           WY      W        3/8/08   Caucus     68
##      clinton obama edwards     margin  winner POP05_SQMI popUnder30_00
## 2445      27    48       0  0.2692308   obama        1.4          24.7
## 2446     342   254       0 -0.1476510 clinton        3.6          29.7
## 2447     231   919       0  0.5982609   obama        4.5          38.6
## 2448      45   123       0  0.4642857   obama        9.6          28.5
## 2449      40    58       0  0.1836735   obama        3.6          23.6
## 2450      40    28       0 -0.1764706 clinton        2.8          24.7
##      pop65up_00 presVote04   kerry04    Bush04 pres04margin pres04winner
## 2445       12.0       3651 0.1999452 0.7797864   0.57984114         bush
## 2446        8.0      16272 0.3200590 0.6546829   0.33462389         bush
## 2447        6.9      11359 0.5257505 0.4510960  -0.07465446        kerry
## 2448        7.0       8081 0.2246009 0.7525059   0.52790496         bush
## 2449       15.9       4114 0.2078269 0.7778318   0.57000486         bush
## 2450       15.6       3392 0.1709906 0.8074882   0.63649764         bush
##      pop06 pop00 hisp06 white06 black06 indian06 asian06 hawaii06 mixed06
## 2445  7359  5952    196    7027      16       44      19        5      52
## 2446 38763 37504   4171   32976     347      353     353       29     534
## 2447 19288 18359   2253   16620      39       87     135        5     149
## 2448 20213 19710   1318   18351      20      176      65       11     272
## 2449  7819  8265   1029    6577       8       62      53        0      90
## 2450  6762  6643    175    6395       9       93      15        1      74
##      pct_less_30k pct_more_100k pct_hs_grad pct_labor_force pct_homeowner
## 2445       0.3651    0.07166948      0.8897          0.6974        0.7351
## 2446       0.3005    0.08259482      0.8737          0.7057        0.7509
## 2447       0.2299    0.18758119      0.9471          0.7941        0.5484
## 2448       0.3320    0.05084746      0.8476          0.7127        0.7516
## 2449       0.4375    0.07346691      0.8562          0.6687        0.7312
## 2450       0.4619    0.05452480      0.8520          0.6001        0.7797
##      unempFeb07 unempFeb08 unempChg pctUnins00 subForPctHomes poverty05
## 2445        1.7        1.6     -0.1       13.4             0       7.2
## 2446        2.4        2.6      0.2       15.0             0       7.6
## 2447        2.2        2.3      0.1       10.0             0       6.1
## 2448        3.0        3.5      0.5       15.5             0      10.5
## 2449        4.4        4.5      0.1       14.8             0      12.6
## 2450        3.5        3.6      0.1       13.7             0       9.4
##      median_hhi05   Catholic So.Bapt.Conv Un.Methodist    E.L.C.A.
## 2445        52315 0.13125000  0.049155405  0.000000000 0.000000000
## 2446        57211 0.26839125  0.044107091  0.005662936 0.007922793
## 2447        60314 0.09528245  0.005369569  0.003123117 0.010136431
## 2448        48557 0.09320231  0.040066863  0.000000000 0.008307162
## 2449        41584 0.14489082  0.060803474  0.065508505 0.039087948
## 2450        40736 0.12071042  0.011739916  0.080523781 0.052829621
##      Construction Manufacturing FinancialActivities GoodsProducing
## 2445    16.705169     1.7326018            2.974300       54.32429
```

```
## 2446     9.959176     6.2691857           4.383050       42.67065
## 2447    13.685720     0.7518666           5.465895       14.91493
## 2448    16.410946     3.5058220           5.170457       32.13250
## 2449     9.949833    13.0330268           6.323161       34.04055
## 2450     7.336489     7.7667814           5.249570       33.58434
##      ServiceProviding
## 2445         45.67571
## 2446         57.32935
## 2447         85.08507
## 2448         67.86750
## 2449         65.95945
## 2450         66.41566
```

Data pre-processing

Suppose our problem is like this. Imagine the year is 2008. Just like the current situation, we know the results of the primary elections that have already occured. We want to predict the results for the upcoming primaries/caucuses. We are bulding a model based on demographic data which are a big contribution to the election outcomes. We know the demographic data data for the states where the upcoming elections will be held. So we want to predict who will win those states. But since we are doing this with 2008 data, we already know what happened. So we can cross examine how good our model is, if our predictions match with the reality.

```
summary(primaries)
```

```
##       fips                county_name    state_postal  region    election_date
##  Min.   : 1001   Washington:  24   TX     : 453   MW: 643   2/5/08 :1090
##  1st Qu.:17108   Franklin  :  20   GA     : 159   NE: 162   3/4/08 : 560
##  Median :29370   Jefferson :  20   VA     : 134   S :1316   2/12/08: 159
##  Mean   :32045   Jackson   :  18   MO     : 115   W : 329   2/19/08: 111
##  3rd Qu.:48214   Madison   :  16   IL     : 102             2/9/08 : 103
##  Max.   :56045   Lincoln   :  14   IA     :  99             1/3/08 :  99
##                  (Other)   :2338   (Other):1388             (Other): 328
##     racetype        tvotes            clinton           obama
##  Caucus : 573   Min.   :      1   Min.   :     0   Min.   :     0.0
##  Primary:1877   1st Qu.:    800   1st Qu.:   339   1st Qu.:   222.5
##                 Median :   2470   Median :  1172   Median :   895.5
##                 Mean   :  11792   Mean   :  5543   Mean   :  5564.2
##                 3rd Qu.:   6587   3rd Qu.:  3430   3rd Qu.:  2786.8
##                 Max.   :1271094   Max.   :699743   Max.   :728328.0
##
##     edwards            margin              winner       POP05_SQMI
##  Min.   :    0.0   Min.   :-1.00000   clinton:1369   Min.   :    0.1
##  1st Qu.:   10.0   1st Qu.:-0.31093   obama  :1061   1st Qu.:   20.0
##  Median :   66.0   Median :-0.05856   NA's   :  20   Median :   44.6
##  Mean   :  409.4   Mean   :-0.07687                  Mean   :  282.7
##  3rd Qu.:  231.0   3rd Qu.: 0.17218                  3rd Qu.:  117.4
##  Max.   :26896.0   Max.   : 1.00000                  Max.   :57173.0
##                                                      NA's   :1
##  popUnder30_00     pop65up_00      presVote04         kerry04
##  Min.   :11.90   Min.   : 1.80   Min.   :    80   Min.   :0.07098
##  1st Qu.:26.10   1st Qu.:11.80   1st Qu.:  5628   1st Qu.:0.30322
##  Median :28.60   Median :14.10   Median :  11698   Median :0.38985
```

```
##  Mean   :28.95   Mean   :14.42   Mean   :  45243   Mean   :0.39375
##  3rd Qu.:31.10   3rd Qu.:16.70   3rd Qu.:  30745   3rd Qu.:0.47250
##  Max.   :60.00   Max.   :34.70   Max.   :3023280   Max.   :0.89184
##  NA's   :1       NA's   :1       NA's   :1         NA's   :1
##      Bush04        pres04margin      pres04winner     pop06
##  Min.   :0.0934   Min.   :-0.79844   bush :1956   Min.   :     60
##  1st Qu.:0.5168   1st Qu.: 0.04435   kerry: 493   1st Qu.:  13466
##  Median :0.5999   Median : 0.21053   NA's :   1   Median :  28785
##  Mean   :0.5970   Mean   : 0.20326                Mean   : 113964
##  3rd Qu.:0.6883   3rd Qu.: 0.38562                3rd Qu.:  73630
##  Max.   :0.9197   Max.   : 0.84790                Max.   :9948081
##  NA's   :1        NA's   :1                       NA's   :1
##      pop00             hisp06           white06           black06
##  Min.   :     65   Min.   :      7   Min.   :     47   Min.   :     0
##  1st Qu.:  13208   1st Qu.:    313   1st Qu.:   9650   1st Qu.:   143
##  Median :  27634   Median :   1092   Median :  22234   Median :  1339
##  Mean   : 106558   Mean   :  20339   Mean   :  71788   Mean   : 14307
##  3rd Qu.:  69666   3rd Qu.:   4870   3rd Qu.:  59440   3rd Qu.:  6747
##  Max.   :9547850   Max.   :4706994   Max.   :2900342   Max.   :1363924
##  NA's   :1         NA's   :1         NA's   :1         NA's   :1
##     indian06          asian06          hawaii06          mixed06
##  Min.   :    0.0   Min.   :      0   Min.   :    0.0   Min.   :     0
##  1st Qu.:   42.0   1st Qu.:     38   1st Qu.:    2.0   1st Qu.:    99
##  Median :  116.0   Median :    140   Median :    7.0   Median :   260
##  Mean   :  717.5   Mean   :   5172   Mean   :  127.5   Mean   :  1513
##  3rd Qu.:  431.0   3rd Qu.:    738   3rd Qu.:   26.0   3rd Qu.:   878
##  Max.   :57535.0   Max.   :1273057   Max.   :24259.0   Max.   :138724
##  NA's   :1         NA's   :1         NA's   :1         NA's   :1
##   pct_less_30k    pct_more_100k      pct_hs_grad     pct_labor_force
##  Min.   :0.0841   Min.   :0.00000   Min.   :0.3470   Min.   :0.3192
##  1st Qu.:0.3653   1st Qu.:0.04110   1st Qu.:0.7008   1st Qu.:0.5597
##  Median :0.4381   Median :0.05454   Median :0.7804   Median :0.6092
##  Mean   :0.4298   Mean   :0.07016   Mean   :0.7651   Mean   :0.6044
##  3rd Qu.:0.5058   3rd Qu.:0.07908   3rd Qu.:0.8340   3rd Qu.:0.6547
##  Max.   :0.7366   Max.   :0.37562   Max.   :0.9696   Max.   :0.8609
##  NA's   :2        NA's   :2         NA's   :2        NA's   :2
##  pct_homeowner      unempFeb07       unempFeb08      unempChg
##  Min.   :0.1954   Min.   : 1.700   Min.   : 1.6   Min.   :-3.8000
##  1st Qu.:0.7036   1st Qu.: 4.100   1st Qu.: 4.1   1st Qu.:-0.4000
##  Median :0.7530   Median : 5.100   Median : 5.3   Median : 0.1000
##  Mean   :0.7389   Mean   : 5.572   Mean   : 5.7   Mean   : 0.1286
##  3rd Qu.:0.7906   3rd Qu.: 6.600   3rd Qu.: 6.8   3rd Qu.: 0.7000
##  Max.   :0.8954   Max.   :21.900   Max.   :23.2   Max.   : 9.3000
##  NA's   :2        NA's   :9        NA's   :9      NA's   :9
##    pctUnins00    subForPctHomes      poverty05     median_hhi05
##  Min.   : 3.80   Min.   :  0.0000   Min.   : 2.5   Min.   :17843
##  1st Qu.:11.50   1st Qu.:  0.0000   1st Qu.:10.9   1st Qu.:32414
##  Median :15.00   Median :  0.0300   Median :14.8   Median :37440
##  Mean   :15.36   Mean   :  0.9652   Mean   :15.7   Mean   :39552
##  3rd Qu.:18.50   3rd Qu.:  0.2400   3rd Qu.:19.3   3rd Qu.:44056
##  Max.   :38.00   Max.   :187.9700   Max.   :46.4   Max.   :98245
##  NA's   :2       NA's   :8
##     Catholic        So.Bapt.Conv      Un.Methodist        E.L.C.A.
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.000000
```

NA's, get rid of them.

```
##  1st Qu.:0.02063   1st Qu.:0.00758   1st Qu.:0.02709   1st Qu.:0.000000
##  Median :0.08525   Median :0.11802   Median :0.05360   Median :0.001943
##  Mean   :0.13613   Mean   :0.15829   Mean   :0.06024   Mean   :0.025437
##  3rd Qu.:0.20094   3rd Qu.:0.27330   3rd Qu.:0.08281   3rd Qu.:0.015252
##  Max.   :0.94682   Max.   :0.96139   Max.   :0.33857   Max.   :0.773274
##  NA's   :3         NA's   :3         NA's   :3         NA's   :3
##   Construction Manufacturing FinancialActivities GoodsProducing
##  Min.   :  0   Min.   :  0   Min.   :  0         Min.   :  0
##  1st Qu.:  4   1st Qu.:  6   1st Qu.:  4         1st Qu.: 21
##  Median :  6   Median : 14   Median :  5         Median : 29
##  Mean   :Inf   Mean   :Inf   Mean   :Inf         Mean   :Inf
##  3rd Qu.:  9   3rd Qu.: 23   3rd Qu.:  6         3rd Qu.: 39
##  Max.   :Inf   Max.   :Inf   Max.   :Inf         Max.   :Inf
##  NA's   :3     NA's   :3     NA's   :1           NA's   :5
##  ServiceProviding
##  Min.   :  0
##  1st Qu.: 61
##  Median : 71
##  Mean   :Inf
##  3rd Qu.: 79
##  Max.   :Inf
##  NA's   :2
```

```
spamdata <- spam
lapply(primaries, class)
```

```
## $fips
## [1] "integer"
##
## $county_name
## [1] "factor"
##
## $state_postal
## [1] "factor"
##
## $region
## [1] "factor"
##
## $election_date
## [1] "factor"
##
## $racetype
## [1] "factor"
##
## $tvotes
## [1] "integer"
##
## $clinton
## [1] "integer"
##
## $obama
## [1] "integer"
##
## $edwards
```

```
## [1] "integer"
##
## $margin
## [1] "numeric"
##
## $winner
## [1] "factor"
##
## $POP05_SQMI
## [1] "numeric"
##
## $popUnder30_00
## [1] "numeric"
##
## $pop65up_00
## [1] "numeric"
##
## $presVote04
## [1] "integer"
##
## $kerry04
## [1] "numeric"
##
## $Bush04
## [1] "numeric"
##
## $pres04margin
## [1] "numeric"
##
## $pres04winner
## [1] "factor"
##
## $pop06
## [1] "integer"
##
## $pop00
## [1] "integer"
##
## $hisp06
## [1] "integer"
##
## $white06
## [1] "integer"
##
## $black06
## [1] "integer"
##
## $indian06
## [1] "integer"
##
## $asian06
## [1] "integer"
##
## $hawaii06
```

```
## [1] "integer"
##
## $mixed06
## [1] "integer"
##
## $pct_less_30k
## [1] "numeric"
##
## $pct_more_100k
## [1] "numeric"
##
## $pct_hs_grad
## [1] "numeric"
##
## $pct_labor_force
## [1] "numeric"
##
## $pct_homeowner
## [1] "numeric"
##
## $unempFeb07
## [1] "numeric"
##
## $unempFeb08
## [1] "numeric"
##
## $unempChg
## [1] "numeric"
##
## $pctUnins00
## [1] "numeric"
##
## $subForPctHomes
## [1] "numeric"
##
## $poverty05
## [1] "numeric"
##
## $median_hhi05
## [1] "integer"
##
## $Catholic
## [1] "numeric"
##
## $So.Bapt.Conv
## [1] "numeric"
##
## $Un.Methodist
## [1] "numeric"
##
## $E.L.C.A.
## [1] "numeric"
##
## $Construction
```

```
## [1] "numeric"
##
## $Manufacturing
## [1] "numeric"
##
## $FinancialActivities
## [1] "numeric"
##
## $GoodsProducing
## [1] "numeric"
##
## $ServiceProviding
## [1] "numeric"
```

```
primaries <- na.omit(primaries)
colSums(is.na(primaries))
```

→ is.na(primaries) returns a data frame with same size as primaries with all entries as 0 except corresponding to the ~~data~~ NA entries where it puts 1. just like this

Data		isonaData
1 NA	→	0 1
2 5		0 0
3 NA		0 1

```
##               fips        county_name       state_postal
##                  0                  0                  0
##             region      election_date           racetype
##                  0                  0                  0
##             tvotes            clinton              obama
##                  0                  0                  0
##            edwards             margin             winner
##                  0                  0                  0
##          POP05_SQMI       popUnder30_00        pop65up_00
##                  0                  0                  0
##         presVote04            kerry04              Bush04
##                  0                  0                  0
##        pres04margin       pres04winner              pop06
##                  0                  0                  0
##               pop00             hisp06             white06
##                  0                  0                  0
##             black06           indian06             asian06
##                  0                  0                  0
##            hawaii06           mixed06        pct_less_30k
##                  0                  0                  0
##       pct_more_100k        pct_hs_grad    pct_labor_force
##                  0                  0                  0
##       pct_homeowner         unempFeb07         unempFeb08
##                  0                  0                  0
##            unempChg         pctUnins00      subForPctHomes
##                  0                  0                  0
##           poverty05       median_hhi05            Catholic
##                  0                  0                  0
##        So.Bapt.Conv       Un.Methodist            E.L.C.A.
##                  0                  0                  0
##        Construction      Manufacturing FinancialActivities
##                  0                  0                  0
##      GoodsProducing    ServiceProviding
##                  0                  0
```

↑ All entries are 0 means no more NA left

```
primaries <- primaries[order(primaries[,5]),]
unique(primaries[,5])
```

```
##  [1] 1/15/08 1/19/08 1/26/08 1/29/08 1/3/08  1/8/08  2/12/08 2/19/08
##  [9] 2/5/08  2/9/08  3/11/08 3/4/08  3/8/08
## 13 Levels: 1/15/08 1/19/08 1/26/08 1/29/08 1/3/08 1/8/08 ... 3/8/08
```

```
training <- primaries[- c(which(primaries[,5] == '3/4/08'), which(primaries[,5] == '3/8/08'),
                          which(primaries[,5] == '3/11/08')),]
test <- primaries[ c(which(primaries[,5] == '3/4/08'), which(primaries[,5] == '3/8/08'),
                          which(primaries[,5] == '3/11/08')),]

trainingdata <- training[, - c(1,2,3,4,5,6,8,9,10,11,20)]
testdata <- test[,- c(1,2,3,4,5,6,8,9,10,11,20)]
```

Model fitting, Model selection and Classification

Model: Elastic Net

```
X <- as.matrix(trainingdata[,-2])
Y <- trainingdata[,2]
length(Y)
```

↖ 2nd column is the response.

```
## [1] 1761
```

```
nrow(X)
```

```
## [1] 1761
```

```
require(glmnet)
```

```
## Loading required package: glmnet
```

```
## Loading required package: Matrix
```

```
## Loading required package: foreach
```

```
## Loaded glmnet 2.0-5
```

```
fit1 <- cv.glmnet(X,Y, family='binomial')
```

```
coef(fit1, s = "lambda.min")
```

```
## 39 x 1 sparse Matrix of class "dgCMatrix"
##                                1
## (Intercept)         9.932430e+00
## tvotes              8.478728e-06
```

```
## POP05_SQMI           -5.376695e-05
## popUnder30_00        -4.727286e-02
## pop65up_00           -1.055329e-01
## presVote04            3.007864e-05
## kerry04              -1.842320e+00
## Bush04               -5.279103e+00
## pres04margin          .
## pop06                 .
## pop00                -1.073614e-05
## hisp06               -9.677522e-06
## white06              -9.930614e-06
## black06               1.922761e-05
## indian06             -8.021641e-05
## asian06               7.345968e-06
## hawaii06              1.031218e-03
## mixed06              -1.946963e-04
## pct_less_30k         -1.367590e+01
## pct_more_100k         2.573166e+00
## pct_hs_grad           2.987648e+00
## pct_labor_force       2.307700e+00
## pct_homeowner        -3.841388e+00
## unempFeb07           -1.435538e-01
## unempFeb08            .
## unempChg              5.114218e-03
## pctUnins00            5.964317e-02
## subForPctHomes        1.276953e-02
## poverty05             1.545141e-01
## median_hhi05         -4.295788e-05
## Catholic             -1.727965e+00
## So.Bapt.Conv         -2.885346e+00
## Un.Methodist          1.970734e+00
## E.L.C.A.              5.940516e+00
## Construction          5.588922e-02
## Manufacturing        -1.177984e-02
## FinancialActivities  -5.719260e-03
## GoodsProducing        2.473142e-02
## ServiceProviding      9.226284e-03
```

Let us see the performance in training data

```
show <- function(tt){
  print(tt)
  cat(paste("Misclassification rate =",round(1-sum(diag(tt))/sum(tt),2),"\n"))
  invisible()}

nx <-  as.matrix(trainingdata[,-2])
nrow(trainingdata)
```

```
## [1] 1761
```

```
nrow(testdata)
```

```
## [1] 646
```

```
show(with(trainingdata, table(actual=Y,
predicted= predict(fit1,newx=nx,s= "lambda.min" , type = "class" ) )))
```

```
##          predicted
## actual    clinton obama
##   clinton     686   210
##   obama       217   648
## Misclassification rate = 0.24
```

Let us see the performance in test data

```
nx <-  as.matrix(testdata[,-2])
show(with(testdata, table(actual=testdata[,2],
predicted= predict(fit1,newx=nx,s= "lambda.min" , type = "class" ) )))
```

```
##          predicted
## actual    clinton obama
##   clinton     340   125
##   obama        86    95
## Misclassification rate = 0.33
```

Miselassification rate in training data. is very high.

This could suggest.

(1.) The predictors are not good enough i.e. we need better predictors.

(2.) The model is not good enough. This could mean a lot of things; the log(odds) is not linearly related to the covariates, i.e. the relationship.

$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$ is not a good way to desribe the relationship.

Maybe $\log\left(\frac{p}{1-p}\right) = g(X_1, \cdots, X_p)$. where g is much more. complicated.

(3) There is too much "randomness."

3. Misclassification rate in test data is even higher. This could suggest that the model is overfit. (4).

This together with (1) suggests that, on one hand we need good predictors, and on the other hand, lack of good predictors are being compensated by using more unnecessary predictors.

Next we try to add some covariates that we originally left out to avoid complications to see if situation improves.

April - 20

Adding some categorical covariates that were left out

we will add racetype and region to our model.

```
head(primaries)
```

```
##      fips county_name state_postal region election_date racetype tvotes
## 863 26001      Alcona           MI     MW       1/15/08  Primary    616
## 864 26003       Alger           MI     MW       1/15/08  Primary    479
## 865 26005     Allegan           MI     MW       1/15/08  Primary   3762
## 866 26007      Alpena           MI     MW       1/15/08  Primary   1505
## 867 26009      Antrim           MI     MW       1/15/08  Primary    998
## 868 26011      Arenac           MI     MW       1/15/08  Primary    729
##     clinton obama edwards     margin  winner POP05_SQMI popUnder30_00
## 863     425     0       0 -0.6899351 clinton       16.8          18.8
## 864     305     0       0 -0.6367432 clinton       10.5          27.5
## 865    2233     0       0 -0.5935673 clinton      135.0          29.3
## 866    1001     0       0 -0.6651163 clinton       52.2          25.6
## 867     584     0       0 -0.5851703 clinton       46.9          23.8
## 868     509     0       0 -0.6982167 clinton       46.2          26.5
##     pop65up_00 presVote04   kerry04    Bush04 pres04margin pres04winner
## 863       24.5       6531 0.4395958 0.5499923  0.110396570         bush
## 864       17.2       4765 0.5026233 0.4864638 -0.016159496        kerry
## 865       11.1      53907 0.3590443 0.6311240  0.272079693         bush
## 866       17.1      15211 0.4869502 0.5039116  0.016961410         bush
## 867       17.5      13619 0.3724209 0.6152434  0.242822527         bush
## 868       16.6       8216 0.4961052 0.4954966 -0.000608569        kerry
##      pop06  pop00 hisp06 white06 black06 indian06 asian06 hawaii06 mixed06
## 863  11759  11706    102   11400      45       76      35        1     100
## 864   9665   9844    125    8350     600      339      36        2     213
## 865 113501 106133   7280  102305    1491      532     819       15    1059
## 866  30067  31294    215   29173     131      143     152        1     252
## 867  24463  23268    311   23398      92      275      47       19     321
## 868  17024  17302    313   15888     334      163      56        1     269
##     pct_less_30k pct_more_100k pct_hs_grad pct_labor_force pct_homeowner
## 863       0.4738    0.04106375      0.7975          0.4435        0.8954
## 864       0.4037    0.04450882      0.8155          0.5105        0.8243
## 865       0.2824    0.08589358      0.8232          0.6902        0.8287
## 866       0.4333    0.04115866      0.8310          0.6037        0.7926
## 867       0.3740    0.06797061      0.8461          0.5907        0.8500
## 868       0.4632    0.04931670      0.7678          0.5234        0.8429
##     unempFeb07 unempFeb08 unempChg pctUnins00 subForPctHomes poverty05
## 863       13.0       13.1      0.1       13.6           0.00      15.4
## 864        9.8       11.0      1.2       11.5           0.00      13.4
## 865        7.2        7.6      0.4        9.9           0.26       9.6
## 866       10.2       10.2      0.0       10.7           0.00      13.7
## 867        9.3        9.8      0.5       11.4           0.00      10.3
## 868       12.4       12.9      0.5       13.4           6.15      17.6
##     median_hhi05   Catholic So.Bapt.Conv Un.Methodist    E.L.C.A.
## 863        33974 0.09053674  0.000000000   0.03225531 0.021332878
## 864        37825 0.31940783  0.000000000   0.02169945 0.085074021
## 865        47662 0.11349075  0.000000000   0.03117399 0.007372356
## 866        35628 0.43590726  0.007536565   0.02529220 0.110493709
## 867        45232 0.05664215  0.003980961   0.03729987 0.006317611
## 868        34179 0.17974405  0.004806300   0.04736812 0.015055880
```

```
##     Construction Manufacturing FinancialActivities GoodsProducing
## 863     7.052779      15.81037            4.390472       24.07753
## 864     5.062305      28.64486            5.140187       35.29595
## 865     6.951944      40.96969            2.100339       53.76219
## 866     6.377853      17.28291            6.553741       24.47138
## 867     7.322243      23.02653            3.840966       34.38885
## 868     2.647953      14.41181            2.966286       20.43843
##     ServiceProviding
## 863         75.92247
## 864         64.70405
## 865         46.23781
## 866         75.52862
## 867         65.61115
## 868         79.56157
```

```
dummy1 <- dummy2 <- dummy3 <- rep(0, times = nrow(primaries))
dummy1[which(primaries$region == 'NE')] <- 1
dummy2[which(primaries$region == 'S')] <- 1
dummy3[which(primaries$region == 'w')] <- 1
primaries <- cbind(primaries, dummy1,dummy2,dummy3)
class(primaries)
```

```
## [1] "data.frame"
```

```
head(primaries)
```

```
##      fips county_name state_postal region election_date racetype tvotes
## 863 26001      Alcona           MI     MW       1/15/08  Primary    616
## 864 26003       Alger           MI     MW       1/15/08  Primary    479
## 865 26005     Allegan           MI     MW       1/15/08  Primary   3762
## 866 26007      Alpena           MI     MW       1/15/08  Primary   1505
## 867 26009      Antrim           MI     MW       1/15/08  Primary    998
## 868 26011      Arenac           MI     MW       1/15/08  Primary    729
##     clinton obama edwards     margin  winner POP05_SQMI popUnder30_00
## 863     425     0       0 -0.6899351 clinton       16.8          18.8
## 864     305     0       0 -0.6367432 clinton       10.5          27.5
## 865    2233     0       0 -0.5935673 clinton      135.0          29.3
## 866    1001     0       0 -0.6651163 clinton       52.2          25.6
## 867     584     0       0 -0.5851703 clinton       46.9          23.8
## 868     509     0       0 -0.6982167 clinton       46.2          26.5
##     pop65up_00 presVote04   kerry04    Bush04 pres04margin pres04winner
## 863       24.5       6531 0.4395958 0.5499923  0.110396570         bush
## 864       17.2       4765 0.5026233 0.4864638 -0.016159496        kerry
## 865       11.1      53907 0.3590443 0.6311240  0.272079693         bush
## 866       17.1      15211 0.4869502 0.5039116  0.016961410         bush
## 867       17.5      13619 0.3724209 0.6152434  0.242822527         bush
## 868       16.6       8216 0.4961052 0.4954966 -0.000608569        kerry
##      pop06  pop00 hisp06 white06 black06 indian06 asian06 hawaii06 mixed06
## 863  11759  11706    102   11400      45       76       35        1      100
## 864   9665   9844    125    8350     600      339       36        2      213
## 865 113501 106133   7280  102305    1491      532      819       15     1059
## 866  30067  31294    215   29173     131      143      152        1      252
## 867  24463  23268    311   23398      92      275       47       19      321
```

```
## 868  17024  17302    313   15888     334      163      56        1     269
##     pct_less_30k pct_more_100k pct_hs_grad pct_labor_force pct_homeowner
## 863       0.4738    0.04106375      0.7975          0.4435        0.8954
## 864       0.4037    0.04450882      0.8155          0.5105        0.8243
## 865       0.2824    0.08589358      0.8232          0.6902        0.8287
## 866       0.4333    0.04115866      0.8310          0.6037        0.7926
## 867       0.3740    0.06797061      0.8461          0.5907        0.8500
## 868       0.4632    0.04931670      0.7678          0.5234        0.8429
##     unempFeb07 unempFeb08 unempChg pctUnins00 subForPctHomes poverty05
## 863       13.0       13.1      0.1       13.6           0.00      15.4
## 864        9.8       11.0      1.2       11.5           0.00      13.4
## 865        7.2        7.6      0.4        9.9           0.26       9.6
## 866       10.2       10.2      0.0       10.7           0.00      13.7
## 867        9.3        9.8      0.5       11.4           0.00      10.3
## 868       12.4       12.9      0.5       13.4           6.15      17.6
##     median_hhi05   Catholic So.Bapt.Conv Un.Methodist    E.L.C.A.
## 863        33974 0.09053674  0.000000000   0.03225531 0.021332878
## 864        37825 0.31940783  0.000000000   0.02169945 0.085074021
## 865        47662 0.11349075  0.000000000   0.03117399 0.007372356
## 866        35628 0.43590726  0.007536565   0.02529220 0.110493709
## 867        45232 0.05664215  0.003980961   0.03729987 0.006317611
## 868        34179 0.17974405  0.004806300   0.04736812 0.015055880
##     Construction Manufacturing FinancialActivities GoodsProducing
## 863     7.052779       15.81037            4.390472       24.07753
## 864     5.062305       28.64486            5.140187       35.29595
## 865     6.951944       40.96969            2.100339       53.76219
## 866     6.377853       17.28291            6.553741       24.47138
## 867     7.322243       23.02653            3.840966       34.38885
## 868     2.647953       14.41181            2.966286       20.43843
##     ServiceProviding dummy1 dummy2 dummy3
## 863         75.92247      0      0      0
## 864         64.70405      0      0      0
## 865         46.23781      0      0      0
## 866         75.52862      0      0      0
## 867         65.61115      0      0      0
## 868         79.56157      0      0      0
```

```
training1 <- primaries[- c(which(primaries[,5] == '3/4/08'), which(primaries[,5] == '3/8/08'),
                           which(primaries[,5] == '3/11/08')),]
test1 <- primaries[ c(which(primaries[,5] == '3/4/08'), which(primaries[,5] == '3/8/08'),
                           which(primaries[,5] == '3/11/08')),]
head(training1)
```

```
##      fips county_name state_postal region election_date racetype tvotes
## 863 26001      Alcona           MI     MW       1/15/08  Primary    616
## 864 26003       Alger           MI     MW       1/15/08  Primary    479
## 865 26005     Allegan           MI     MW       1/15/08  Primary   3762
## 866 26007      Alpena           MI     MW       1/15/08  Primary   1505
## 867 26009      Antrim           MI     MW       1/15/08  Primary    998
## 868 26011      Arenac           MI     MW       1/15/08  Primary    729
##     clinton obama edwards     margin  winner POP05_SQMI popUnder30_00
## 863     425     0       0 -0.6899351 clinton       16.8          18.8
## 864     305     0       0 -0.6367432 clinton       10.5          27.5
## 865    2233     0       0 -0.5935673 clinton      135.0          29.3
```

```
## 866    1001     0       0 -0.6651163 clinton       52.2          25.6
## 867     584     0       0 -0.5851703 clinton       46.9          23.8
## 868     509     0       0 -0.6982167 clinton       46.2          26.5
##     pop65up_00 presVote04   kerry04    Bush04 pres04margin pres04winner
## 863       24.5       6531 0.4395958 0.5499923  0.110396570         bush
## 864       17.2       4765 0.5026233 0.4864638 -0.016159496        kerry
## 865       11.1      53907 0.3590443 0.6311240  0.272079693         bush
## 866       17.1      15211 0.4869502 0.5039116  0.016961410         bush
## 867       17.5      13619 0.3724209 0.6152434  0.242822527         bush
## 868       16.6       8216 0.4961052 0.4954966 -0.000608569        kerry
##      pop06  pop00 hisp06 white06 black06 indian06 asian06 hawaii06 mixed06
## 863  11759  11706    102   11400      45       76      35        1     100
## 864   9665   9844    125    8350     600      339      36        2     213
## 865 113501 106133   7280  102305    1491      532     819       15    1059
## 866  30067  31294    215   29173     131      143     152        1     252
## 867  24463  23268    311   23398      92      275      47       19     321
## 868  17024  17302    313   15888     334      163      56        1     269
##     pct_less_30k pct_more_100k pct_hs_grad pct_labor_force pct_homeowner
## 863       0.4738    0.04106375      0.7975          0.4435        0.8954
## 864       0.4037    0.04450882      0.8155          0.5105        0.8243
## 865       0.2824    0.08589358      0.8232          0.6902        0.8287
## 866       0.4333    0.04115866      0.8310          0.6037        0.7926
## 867       0.3740    0.06797061      0.8461          0.5907        0.8500
## 868       0.4632    0.04931670      0.7678          0.5234        0.8429
##     unempFeb07 unempFeb08 unempChg pctUnins00 subForPctHomes poverty05
## 863       13.0       13.1      0.1       13.6           0.00      15.4
## 864        9.8       11.0      1.2       11.5           0.00      13.4
## 865        7.2        7.6      0.4        9.9           0.26       9.6
## 866       10.2       10.2      0.0       10.7           0.00      13.7
## 867        9.3        9.8      0.5       11.4           0.00      10.3
## 868       12.4       12.9      0.5       13.4           6.15      17.6
##     median_hhi05   Catholic So.Bapt.Conv Un.Methodist    E.L.C.A.
## 863        33974 0.09053674  0.000000000   0.03225531 0.021332878
## 864        37825 0.31940783  0.000000000   0.02169945 0.085074021
## 865        47662 0.11349075  0.000000000   0.03117399 0.007372356
## 866        35628 0.43590726  0.007536565   0.02529220 0.110493709
## 867        45232 0.05664215  0.003980961   0.03729987 0.006317611
## 868        34179 0.17974405  0.004806300   0.04736812 0.015055880
##     Construction Manufacturing FinancialActivities GoodsProducing
## 863     7.052779      15.81037            4.390472       24.07753
## 864     5.062305      28.64486            5.140187       35.29595
## 865     6.951944      40.96969            2.100339       53.76219
## 866     6.377853      17.28291            6.553741       24.47138
## 867     7.322243      23.02653            3.840966       34.38885
## 868     2.647953      14.41181            2.966286       20.43843
##     ServiceProviding dummy1 dummy2 dummy3
## 863         75.92247      0      0      0
## 864         64.70405      0      0      0
## 865         46.23781      0      0      0
## 866         75.52862      0      0      0
## 867         65.61115      0      0      0
## 868         79.56157      0      0      0
```

```
trainingdata1 <- training1[, - c(1,2,3,4,5,8,9,10,11,20)]
testdata1 <- test1[,- c(1,2,3,4,5,8,9,10,11,20)]
trainingdata1 <- within(trainingdata1, {racetype <- as.numeric(racetype)})
head(trainingdata1)
```

```
##     racetype tvotes  winner POP05_SQMI popUnder30_00 pop65up_00 presVote04
## 863        2    616 clinton       16.8          18.8       24.5       6531
## 864        2    479 clinton       10.5          27.5       17.2       4765
## 865        2   3762 clinton      135.0          29.3       11.1      53907
## 866        2   1505 clinton       52.2          25.6       17.1      15211
## 867        2    998 clinton       46.9          23.8       17.5      13619
## 868        2    729 clinton       46.2          26.5       16.6       8216
##       kerry04    Bush04 pres04margin  pop06  pop00 hisp06 white06 black06
## 863 0.4395958 0.5499923  0.110396570  11759  11706    102   11400      45
## 864 0.5026233 0.4864638 -0.016159496   9665   9844    125    8350     600
## 865 0.3590443 0.6311240  0.272079693 113501 106133   7280  102305    1491
## 866 0.4869502 0.5039116  0.016961410  30067  31294    215   29173     131
## 867 0.3724209 0.6152434  0.242822527  24463  23268    311   23398      92
## 868 0.4961052 0.4954966 -0.000608569  17024  17302    313   15888     334
##     indian06 asian06 hawaii06 mixed06 pct_less_30k pct_more_100k
## 863       76      35         1     100       0.4738    0.04106375
## 864      339      36         2     213       0.4037    0.04450882
## 865      532     819        15    1059       0.2824    0.08589358
## 866      143     152         1     252       0.4333    0.04115866
## 867      275      47        19     321       0.3740    0.06797061
## 868      163      56         1     269       0.4632    0.04931670
##     pct_hs_grad pct_labor_force pct_homeowner unempFeb07 unempFeb08
## 863      0.7975          0.4435        0.8954       13.0       13.1
## 864      0.8155          0.5105        0.8243        9.8       11.0
## 865      0.8232          0.6902        0.8287        7.2        7.6
## 866      0.8310          0.6037        0.7926       10.2       10.2
## 867      0.8461          0.5907        0.8500        9.3        9.8
## 868      0.7678          0.5234        0.8429       12.4       12.9
##     unempChg pctUnins00 subForPctHomes poverty05 median_hhi05   Catholic
## 863      0.1       13.6           0.00      15.4        33974 0.09053674
## 864      1.2       11.5           0.00      13.4        37825 0.31940783
## 865      0.4        9.9           0.26       9.6        47662 0.11349075
## 866      0.0       10.7           0.00      13.7        35628 0.43590726
## 867      0.5       11.4           0.00      10.3        45232 0.05664215
## 868      0.5       13.4           6.15      17.6        34179 0.17974405
##     So.Bapt.Conv Un.Methodist    E.L.C.A. Construction Manufacturing
## 863  0.000000000   0.03225531 0.021332878     7.052779      15.81037
## 864  0.000000000   0.02169945 0.085074021     5.062305      28.64486
## 865  0.000000000   0.03117399 0.007372356     6.951944      40.96969
## 866  0.007536565   0.02529220 0.110493709     6.377853      17.28291
## 867  0.003980961   0.03729987 0.006317611     7.322243      23.02653
## 868  0.004806300   0.04736812 0.015055880     2.647953      14.41181
##     FinancialActivities GoodsProducing ServiceProviding dummy1 dummy2
## 863            4.390472       24.07753         75.92247      0      0
## 864            5.140187       35.29595         64.70405      0      0
## 865            2.100339       53.76219         46.23781      0      0
## 866            6.553741       24.47138         75.52862      0      0
## 867            3.840966       34.38885         65.61115      0      0
```

```
## 868            2.966286       20.43843         79.56157       0       0
##     dummy3
## 863      0
## 864      0
## 865      0
## 866      0
## 867      0
## 868      0
```

```
testdata1 <- within(testdata1, {racetype <- as.numeric(racetype)})
```

```
X <- as.matrix(trainingdata1[,-3])
Y <- trainingdata1[,3]
length(Y)
```

```
## [1] 1761
```

```
nrow(X)
```

```
## [1] 1761
```

```
require(glmnet)
fit1 <- cv.glmnet(X,Y, family='binomial')
```

```
coef(fit1, s = "lambda.min")
```

```
## 43 x 1 sparse Matrix of class "dgCMatrix"
##                                    1
## (Intercept)          3.534501e+01
## racetype            -9.956649e-01
## tvotes               1.660793e-05
## POP05_SQMI          -8.670877e-06
## popUnder30_00       -2.773802e-02
## pop65up_00          -9.810798e-02
## presVote04           2.466928e-05
## kerry04             -2.538128e+01
## Bush04              -2.929783e+01
## pres04margin         .
## pop06               -1.043383e-09
## pop00               -9.393411e-06
## hisp06              -1.548317e-05
## white06             -8.040002e-06
## black06              1.945375e-05
## indian06            -7.444297e-05
## asian06              5.551581e-06
## hawaii06             1.127357e-03
## mixed06             -2.285613e-04
## pct_less_30k        -1.272157e+01
## pct_more_100k        6.293939e+00
## pct_hs_grad          7.651533e-01
## pct_labor_force      1.783347e+00
```

```
## pct_homeowner        -3.861450e+00
## unempFeb07           -1.120635e-01
## unempFeb08            .
## unempChg              1.018407e-02
## pctUnins00            4.744463e-02
## subForPctHomes        1.168637e-02
## poverty05             1.430445e-01
## median_hhi05         -2.665598e-05
## Catholic             -1.008624e+00
## So.Bapt.Conv         -2.500310e+00
## Un.Methodist          3.821955e+00
## E.L.C.A.              2.897967e+00
## Construction          5.284962e-02
## Manufacturing        -8.505009e-03
## FinancialActivities -1.618908e-02
## GoodsProducing        2.039738e-02
## ServiceProviding      1.358407e-02
## dummy1               -1.825483e+00
## dummy2                2.893767e-02
## dummy3                .
```

Midwest is baseline.

The dummy 3. is not needed in the model, means in this case the odds. do not change. if you look at western voters compared to midwes

Let us see the performance in training data

```
show <- function(tt){
  print(tt)
  cat(paste("Misclassification rate =",round(1-sum(diag(tt))/sum(tt),2),"\n"))
  invisible()}

nx <-  as.matrix(trainingdata1[,-3])
nrow(trainingdata1)
```

```
## [1] 1761
```

```
nrow(testdata1)
```

```
## [1] 646
```

```
show(with(trainingdata1, table(actual=Y,
predicted= predict(fit1,newx=nx,s= "lambda.min" , type = "class" ) )))
```

```
##          predicted
## actual    clinton obama
##   clinton     700   196
##   obama       213   652
## Misclassification rate = 0.23
```

Let us see the performance in test data

```
nx <-  as.matrix(testdata1[,-3])
show(with(testdata1, table(actual=testdata1[,3],
predicted= predict(fit1,newx=nx,s= "lambda.min" , type = "class" ) )))
```

```
##          predicted
## actual    clinton obama
##   clinton     317   148
##   obama        73   108
## Misclassification rate = 0.34
```

```
require(randomForest)
```

```
## Loading required package: randomForest
```

```
## randomForest 4.6-12
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
rf <- randomForest(X,Y)
```

Random forest on training data

```
show(with(testdata1, table(actual=trainingdata1[,3], predicted=predict(rf))))
```

```
##          predicted
## actual    clinton obama
##   clinton     723   173
##   obama       131   734
## Misclassification rate = 0.17
```

Random Forest on test data

```
show(with(testdata1, table(actual=testdata1[,3],
predicted=predict(rf, newdata= testdata1[,-3]))))
```

```
##          predicted
## actual    clinton obama
##   clinton     338   127
##   obama        30   151
## Misclassification rate = 0.24
```

instead of who wins

Exercise: Predict win margin, using the same set of predictors.