satander.R

spoor

Sun Apr 09 07:13:30 2017

```r
# This R environment comes with all of CRAN preinstalled, as well as many other helpful packages
# The environment is defined by the kaggle/rstats docker image: https://github.com/kaggle/docker-rstats
# For example, here's several helpful packages to load in

library(readr) # CSV file I/O, e.g. the read_csv function
library(xgboost)

# Reading the data
dat_train <- read.csv("C:/Users/spoor/Desktop/Marketing Analytics/satander data/train.csv", stringsAsFac
dat_test <- read.csv("C:/Users/spoor/Desktop/Marketing Analytics/satander data/test.csv", stringsAsFacto

# Mergin the test and train data
dat_test$TARGET <- NA
all_dat <- rbind(dat_train, dat_test)

# Removing the constant variables
train_names <- names(dat_train)[-1]
for (i in train_names)
{
  if (class(all_dat[[i]]) == "integer")
  {
    u <- unique(all_dat[[i]])
    if (length(u) == 1)
    {
      all_dat[[i]] <- NULL
    }
  }
}

#Removing duplicate columns
train_names <- names(all_dat)[-1]
fac <- data.frame(fac = integer())

for(i in 1:length(train_names))
{
  if(i != length(train_names))
  {
    for (k in (i+1):length(train_names))
    {
      if(identical(all_dat[,i], all_dat[,k]) == TRUE)
      {
        fac <- rbind(fac, data.frame(fac = k))
      }
    }
  }
}
same <- unique(fac$fac)
```

1

```r
all_dat <- all_dat[,-same]

#Removing hghly correlated variables
cor_v<-abs(cor(all_dat))
diag(cor_v)<-0
cor_v[upper.tri(cor_v)] <- 0
cor_f <- as.data.frame(which(cor_v > 0.85, arr.ind = T))
all_dat <- all_dat[,-unique(cor_f$row)]

# Splitting the data for model
train <- all_dat[1:nrow(dat_train), ]
test <- all_dat[-(1:nrow(dat_train)), ]


#Building the model
set.seed(88)
param <- list("objective" = "binary:logistic",booster = "gbtree",
              "eval_metric" = "auc",colsample_bytree = 0.85, subsample = 0.95)

y <- as.numeric(train$TARGET)

#AUC was highest in 310th round during cross validation
xgbmodel <- xgboost(data = as.matrix(train[,-c(1,151)]), params = param,
                    nrounds = 310, max.depth = 5, eta = 0.03,
                    label = y, maximize = T)
```

```
## [1]   train-auc:0.816359
## [2]   train-auc:0.824705
## [3]   train-auc:0.827291
## [4]   train-auc:0.831310
## [5]   train-auc:0.833136
## [6]   train-auc:0.834932
## [7]   train-auc:0.836569
## [8]   train-auc:0.836739
## [9]   train-auc:0.836965
## [10]  train-auc:0.837208
## [11]  train-auc:0.837340
## [12]  train-auc:0.837201
## [13]  train-auc:0.837643
## [14]  train-auc:0.837997
## [15]  train-auc:0.838619
## [16]  train-auc:0.838743
## [17]  train-auc:0.839454
## [18]  train-auc:0.839469
## [19]  train-auc:0.839366
## [20]  train-auc:0.839579
## [21]  train-auc:0.839689
## [22]  train-auc:0.839773
## [23]  train-auc:0.839777
## [24]  train-auc:0.840389
## [25]  train-auc:0.840632
## [26]  train-auc:0.841279
## [27]  train-auc:0.841277
## [28]  train-auc:0.841095
```

```
## [29] train-auc:0.841437
## [30] train-auc:0.841390
## [31] train-auc:0.841564
## [32] train-auc:0.841757
## [33] train-auc:0.841851
## [34] train-auc:0.841935
## [35] train-auc:0.842697
## [36] train-auc:0.842981
## [37] train-auc:0.843242
## [38] train-auc:0.843239
## [39] train-auc:0.843561
## [40] train-auc:0.843821
## [41] train-auc:0.843957
## [42] train-auc:0.844568
## [43] train-auc:0.844754
## [44] train-auc:0.844905
## [45] train-auc:0.844695
## [46] train-auc:0.844721
## [47] train-auc:0.845101
## [48] train-auc:0.845126
## [49] train-auc:0.844855
## [50] train-auc:0.845518
## [51] train-auc:0.846078
## [52] train-auc:0.846371
## [53] train-auc:0.847019
## [54] train-auc:0.847509
## [55] train-auc:0.847646
## [56] train-auc:0.847762
## [57] train-auc:0.847975
## [58] train-auc:0.848129
## [59] train-auc:0.848380
## [60] train-auc:0.848694
## [61] train-auc:0.848830
## [62] train-auc:0.849018
## [63] train-auc:0.849300
## [64] train-auc:0.849418
## [65] train-auc:0.849504
## [66] train-auc:0.849628
## [67] train-auc:0.849771
## [68] train-auc:0.849916
## [69] train-auc:0.849980
## [70] train-auc:0.850157
## [71] train-auc:0.850391
## [72] train-auc:0.850575
## [73] train-auc:0.850752
## [74] train-auc:0.850844
## [75] train-auc:0.851032
## [76] train-auc:0.851184
## [77] train-auc:0.851323
## [78] train-auc:0.851307
## [79] train-auc:0.851536
## [80] train-auc:0.851599
## [81] train-auc:0.852025
## [82] train-auc:0.852211
```

```
## [83]  train-auc:0.852385
## [84]  train-auc:0.852419
## [85]  train-auc:0.852675
## [86]  train-auc:0.852707
## [87]  train-auc:0.852809
## [88]  train-auc:0.852833
## [89]  train-auc:0.852808
## [90]  train-auc:0.852892
## [91]  train-auc:0.853151
## [92]  train-auc:0.853348
## [93]  train-auc:0.853581
## [94]  train-auc:0.853686
## [95]  train-auc:0.853763
## [96]  train-auc:0.854025
## [97]  train-auc:0.854246
## [98]  train-auc:0.854478
## [99]  train-auc:0.854644
## [100]     train-auc:0.854769
## [101]     train-auc:0.854905
## [102]     train-auc:0.855044
## [103]     train-auc:0.855131
## [104]     train-auc:0.855279
## [105]     train-auc:0.855368
## [106]     train-auc:0.855515
## [107]     train-auc:0.855787
## [108]     train-auc:0.855967
## [109]     train-auc:0.856089
## [110]     train-auc:0.856142
## [111]     train-auc:0.856307
## [112]     train-auc:0.856562
## [113]     train-auc:0.856654
## [114]     train-auc:0.856860
## [115]     train-auc:0.856957
## [116]     train-auc:0.857112
## [117]     train-auc:0.857259
## [118]     train-auc:0.857432
## [119]     train-auc:0.857846
## [120]     train-auc:0.858111
## [121]     train-auc:0.858319
## [122]     train-auc:0.858465
## [123]     train-auc:0.858659
## [124]     train-auc:0.858837
## [125]     train-auc:0.858938
## [126]     train-auc:0.859094
## [127]     train-auc:0.859245
## [128]     train-auc:0.859430
## [129]     train-auc:0.859583
## [130]     train-auc:0.859722
## [131]     train-auc:0.860088
## [132]     train-auc:0.860196
## [133]     train-auc:0.860402
## [134]     train-auc:0.860504
## [135]     train-auc:0.860599
## [136]     train-auc:0.860854
```

```
## [137]    train-auc:0.860971
## [138]    train-auc:0.861126
## [139]    train-auc:0.861323
## [140]    train-auc:0.861476
## [141]    train-auc:0.861669
## [142]    train-auc:0.861833
## [143]    train-auc:0.862056
## [144]    train-auc:0.862246
## [145]    train-auc:0.862354
## [146]    train-auc:0.862626
## [147]    train-auc:0.862755
## [148]    train-auc:0.862886
## [149]    train-auc:0.863010
## [150]    train-auc:0.863202
## [151]    train-auc:0.863371
## [152]    train-auc:0.863499
## [153]    train-auc:0.863541
## [154]    train-auc:0.863667
## [155]    train-auc:0.863909
## [156]    train-auc:0.863997
## [157]    train-auc:0.864082
## [158]    train-auc:0.864212
## [159]    train-auc:0.864417
## [160]    train-auc:0.864503
## [161]    train-auc:0.864594
## [162]    train-auc:0.864774
## [163]    train-auc:0.864964
## [164]    train-auc:0.865014
## [165]    train-auc:0.865080
## [166]    train-auc:0.865221
## [167]    train-auc:0.865394
## [168]    train-auc:0.865558
## [169]    train-auc:0.865682
## [170]    train-auc:0.865773
## [171]    train-auc:0.865914
## [172]    train-auc:0.866068
## [173]    train-auc:0.866121
## [174]    train-auc:0.866219
## [175]    train-auc:0.866331
## [176]    train-auc:0.866460
## [177]    train-auc:0.866571
## [178]    train-auc:0.866674
## [179]    train-auc:0.866836
## [180]    train-auc:0.866964
## [181]    train-auc:0.867119
## [182]    train-auc:0.867302
## [183]    train-auc:0.867400
## [184]    train-auc:0.867496
## [185]    train-auc:0.867619
## [186]    train-auc:0.867722
## [187]    train-auc:0.867855
## [188]    train-auc:0.868017
## [189]    train-auc:0.868158
## [190]    train-auc:0.868227
```

```
## [191]     train-auc:0.868337
## [192]     train-auc:0.868453
## [193]     train-auc:0.868582
## [194]     train-auc:0.868710
## [195]     train-auc:0.868786
## [196]     train-auc:0.868860
## [197]     train-auc:0.868980
## [198]     train-auc:0.869054
## [199]     train-auc:0.869181
## [200]     train-auc:0.869242
## [201]     train-auc:0.869337
## [202]     train-auc:0.869462
## [203]     train-auc:0.869536
## [204]     train-auc:0.869670
## [205]     train-auc:0.869759
## [206]     train-auc:0.869835
## [207]     train-auc:0.869990
## [208]     train-auc:0.870092
## [209]     train-auc:0.870171
## [210]     train-auc:0.870286
## [211]     train-auc:0.870341
## [212]     train-auc:0.870492
## [213]     train-auc:0.870646
## [214]     train-auc:0.870696
## [215]     train-auc:0.870872
## [216]     train-auc:0.871057
## [217]     train-auc:0.871140
## [218]     train-auc:0.871203
## [219]     train-auc:0.871370
## [220]     train-auc:0.871464
## [221]     train-auc:0.871603
## [222]     train-auc:0.871662
## [223]     train-auc:0.871702
## [224]     train-auc:0.871741
## [225]     train-auc:0.871804
## [226]     train-auc:0.871895
## [227]     train-auc:0.871948
## [228]     train-auc:0.872043
## [229]     train-auc:0.872110
## [230]     train-auc:0.872191
## [231]     train-auc:0.872249
## [232]     train-auc:0.872355
## [233]     train-auc:0.872363
## [234]     train-auc:0.872408
## [235]     train-auc:0.872459
## [236]     train-auc:0.872516
## [237]     train-auc:0.872564
## [238]     train-auc:0.872703
## [239]     train-auc:0.872738
## [240]     train-auc:0.872770
## [241]     train-auc:0.872848
## [242]     train-auc:0.872880
## [243]     train-auc:0.872949
## [244]     train-auc:0.873144
```

```
## [245]    train-auc:0.873170
## [246]    train-auc:0.873208
## [247]    train-auc:0.873373
## [248]    train-auc:0.873449
## [249]    train-auc:0.873627
## [250]    train-auc:0.873788
## [251]    train-auc:0.873879
## [252]    train-auc:0.873908
## [253]    train-auc:0.873942
## [254]    train-auc:0.874010
## [255]    train-auc:0.874064
## [256]    train-auc:0.874144
## [257]    train-auc:0.874172
## [258]    train-auc:0.874251
## [259]    train-auc:0.874372
## [260]    train-auc:0.874432
## [261]    train-auc:0.874503
## [262]    train-auc:0.874589
## [263]    train-auc:0.874687
## [264]    train-auc:0.874731
## [265]    train-auc:0.874744
## [266]    train-auc:0.874842
## [267]    train-auc:0.874864
## [268]    train-auc:0.874966
## [269]    train-auc:0.875071
## [270]    train-auc:0.875094
## [271]    train-auc:0.875158
## [272]    train-auc:0.875245
## [273]    train-auc:0.875283
## [274]    train-auc:0.875304
## [275]    train-auc:0.875400
## [276]    train-auc:0.875464
## [277]    train-auc:0.875582
## [278]    train-auc:0.875792
## [279]    train-auc:0.875842
## [280]    train-auc:0.875868
## [281]    train-auc:0.875895
## [282]    train-auc:0.875948
## [283]    train-auc:0.875999
## [284]    train-auc:0.876046
## [285]    train-auc:0.876193
## [286]    train-auc:0.876226
## [287]    train-auc:0.876278
## [288]    train-auc:0.876390
## [289]    train-auc:0.876455
## [290]    train-auc:0.876487
## [291]    train-auc:0.876511
## [292]    train-auc:0.876679
## [293]    train-auc:0.876701
## [294]    train-auc:0.876788
## [295]    train-auc:0.876905
## [296]    train-auc:0.876956
## [297]    train-auc:0.876995
## [298]    train-auc:0.877007
```

```
## [299]     train-auc:0.877166
## [300]     train-auc:0.877282
## [301]     train-auc:0.877315
## [302]     train-auc:0.877342
## [303]     train-auc:0.877434
## [304]     train-auc:0.877528
## [305]     train-auc:0.877573
## [306]     train-auc:0.877606
## [307]     train-auc:0.877648
## [308]     train-auc:0.877741
## [309]     train-auc:0.877792
## [310]     train-auc:0.877889
```

```r
#Prediction
res <- predict(xgbmodel, newdata = data.matrix(test[,-c(1,151)]))
res <- data.frame(ID = test$ID, TARGET = res)

write.csv(res, "submission.csv", row.names = FALSE)
```