Machine Learning PS4:ML trading Algo

YiTao Hu

26/04/2020

```
#import data and libraries
library('lfe')
library('stargazer')
library(readr)
library(data.table)
library(foreign)
library(knitr)
library(ggplot2)
library(reshape2)
library(PerformanceAnalytics)
library(knitr)
StockRetAcct_DT= as.data.table(read.dta("StockRetAcct_insample.dta")) #set the key as firm ID and year
setkey(x = StockRetAcct_DT,FirmID,year)
StockRetAcct_DT=na.omit(StockRetAcct_DT)
```

Question 1

a

```
#create the new feasures
Features_df=StockRetAcct_DT[,list(
                                  lnIssue=lnIssue,
                                  lnIssue2=lnIssue^2,
                                  lnIssueME=lnIssue*lnME,
                                  lnProf=lnProf,
                                  lnProf2=lnProf^2,
                                  lnProfME=lnProf*lnME,
                                  lnInv=lnInv,
                                  lnInv2=lnInv^2,
                                  lnInv=lnInv*lnME,
                                  lnME=lnME,
                                  lnME2=lnME^2
                                  ),
                             by=list(FirmID,year)]

Returns=StockRetAcct_DT[,1:4,]
```

i

Note in order to use the formula written in 48 page, we need to first normalize all our features cross-sectionally (because we assume zero mean and 1 variance.), and then using the formula to compute feature-based long-short portfolio returns. The method is equivalent to using Fama-MacBeth Reregression for Portfolio construction after reviewing notes of Lecture 1.

```
#define a function to normalize the features
normalize=function(vec){
```

```
  return((vec-mean(vec))/sd(vec))
}
FactorRet=data.frame(matrix(NA,35,12))
i=1
#cross-sectionally normalize
for (t in min(Features_df$year):max(Features_df$year)){
  Xt=as.matrix(Features_df[Features_df$year==t,3:ncol(Features_df)])
  #get excess Return
  Rt=exp(Returns[year==t,,]$lnAnnRet)-exp(Returns[year==t,,]$lnRf)
  #normalize
  Xt=apply(Xt, 2, FUN=normalize)
  #insert 1s
  Xt=cbind(1,Xt)
  #compute factor returns
  FactorRet[i,]=t(Xt)%*%Rt/nrow(Xt)
  i=i+1
}
colnames(FactorRet)=c(colnames(Features_df[,3:ncol(Features_df),]),'Mkt')
row.names(FactorRet)=1980:2014
kable(colMeans(FactorRet),caption = 'Mean returns of Factors')
```

Table 1: Mean returns of Factors

x
0.0941350
-0.0162063
-0.0139938
-0.0164889
0.0176816
0.0005300
0.0161245
-0.0209282
-0.0202035
-0.0211280
-0.0054069
-0.0054646