HW5_SOLUTION

gustavoc@msu.edu

12/12/2021

Code

```
  library(BGLR)
  data(wheat)
  head(wheat.Y)
```

```
##               1           2           4          5
## 775   1.6716295 -1.72746986 -1.89028479  0.0509159
## 2166 -0.2527028  0.40952243  0.30938553 -1.7387588
## 2167  0.3418151 -0.64862633 -0.79955921 -1.0535691
## 2465  0.7854395  0.09394919  0.57046773  0.5517574
## 3881  0.9983176 -0.28248062  1.61868192 -0.1142848
## 3889  2.3360969  0.62647587  0.07353311  0.7195856
```

```
  dim(wheat.X)
```

```
## [1]  599 1279
```

```
  X=scale(wheat.X,center=TRUE,scale=TRUE)
  N=nrow(X)
  y=wheat.Y[,2] # picks one phenotype

  nTST=50


  COR=matrix(nrow=nTST,ncol=2,NA)
  colnames(COR)=c('Lasso','Forward')

  DF=COR

  library(BGDataExt)
```

```
##
## Attaching package: 'BGDataExt'
```

```
## The following object is masked from 'package:graphics':
##
##     segments
```

```
  library(glmnet)
```

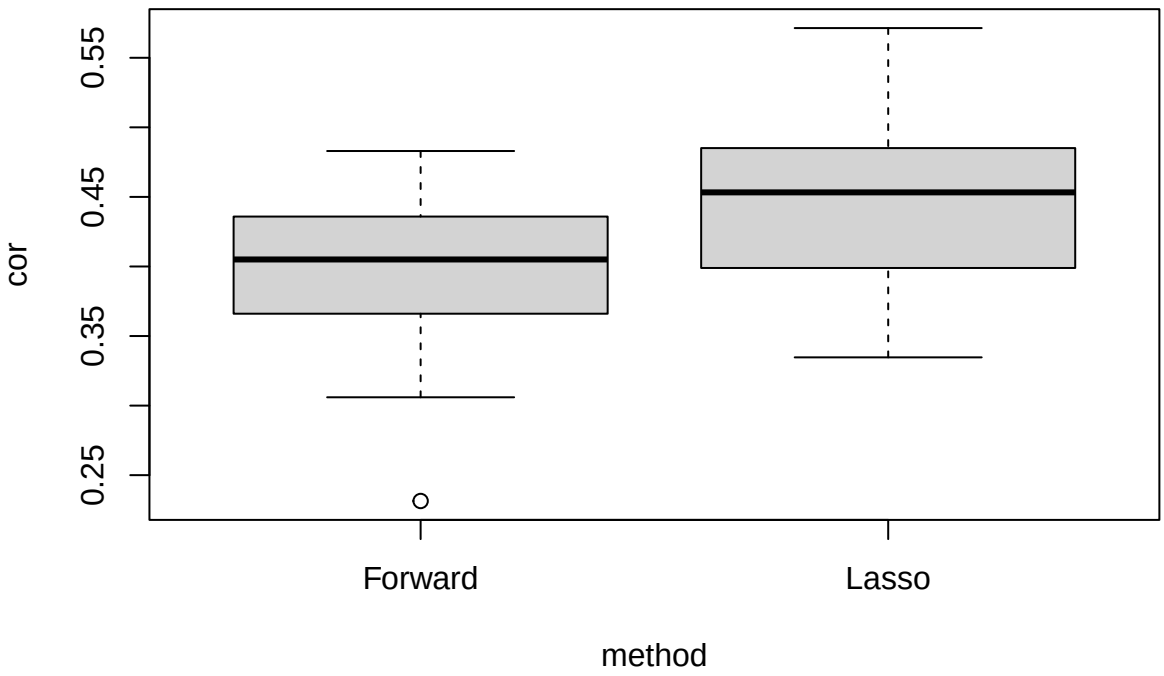
```
## Loading required package: Matrix
```

```
## Loaded glmnet 4.1-3
```

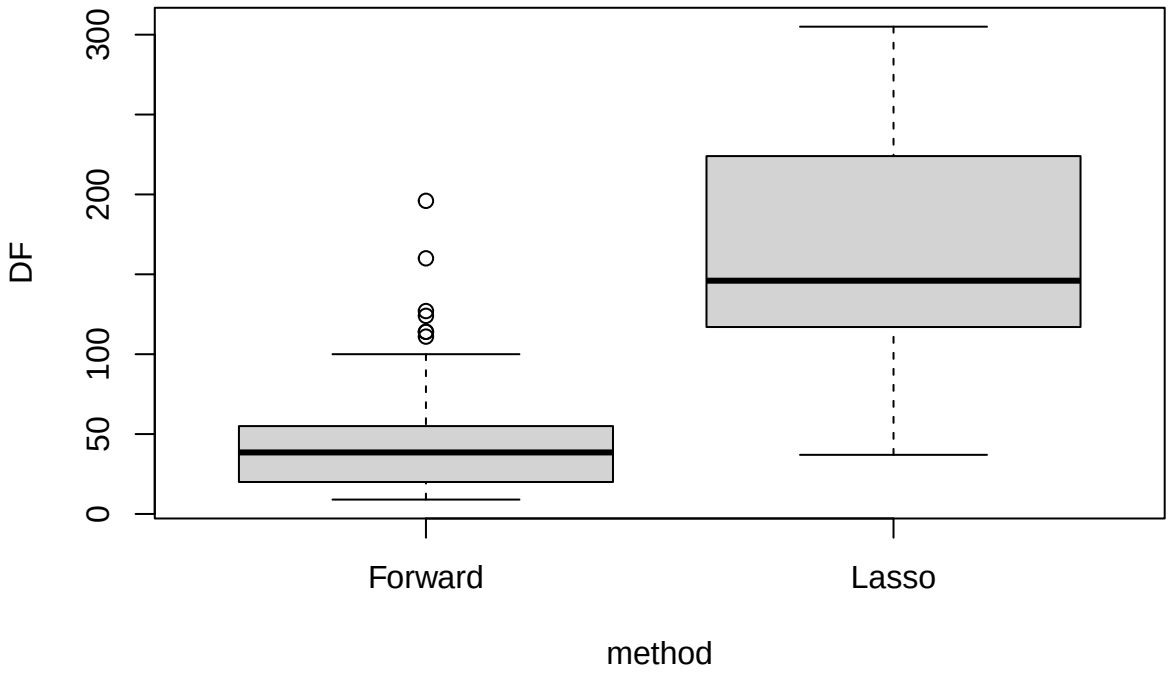
```
for(i in 1:nTST){
  tst<-sample(1:N,size=150,replace=FALSE)
  XTRN<-X[-tst,]
  yTRN<-y[-tst]
    XTST<-X[tst,]
    yTST<-y[tst]

  ## Lasso
   fm=glmnet(y=yTRN,x=XTRN,alpha=1)
   tmpCor=rep(NA,100)
   for(j in 2:100){ tmpCor[j]=cor(yTST,XTST%*%fm$beta[,j])}
   tmp=which.max(tmpCor)
   DF[i,'Lasso']=fm$df[tmp]
   COR[i,'Lasso']=tmpCor[tmp]

  ## Forward
   FM=FWD(y=yTRN,X=XTRN,df=200,verbose=FALSE)
   tmpCor=rep(NA,150)
   XTST=cbind(1,XTST)
   for(j in 2:200){ tmpCor[j]=cor(yTST,XTST%*%FM$B[,j])}
   tmp=which.max(tmpCor)
   DF[i,'Forward']=tmp
   COR[i,'Forward']=tmpCor[tmp]
 }
```

Plots

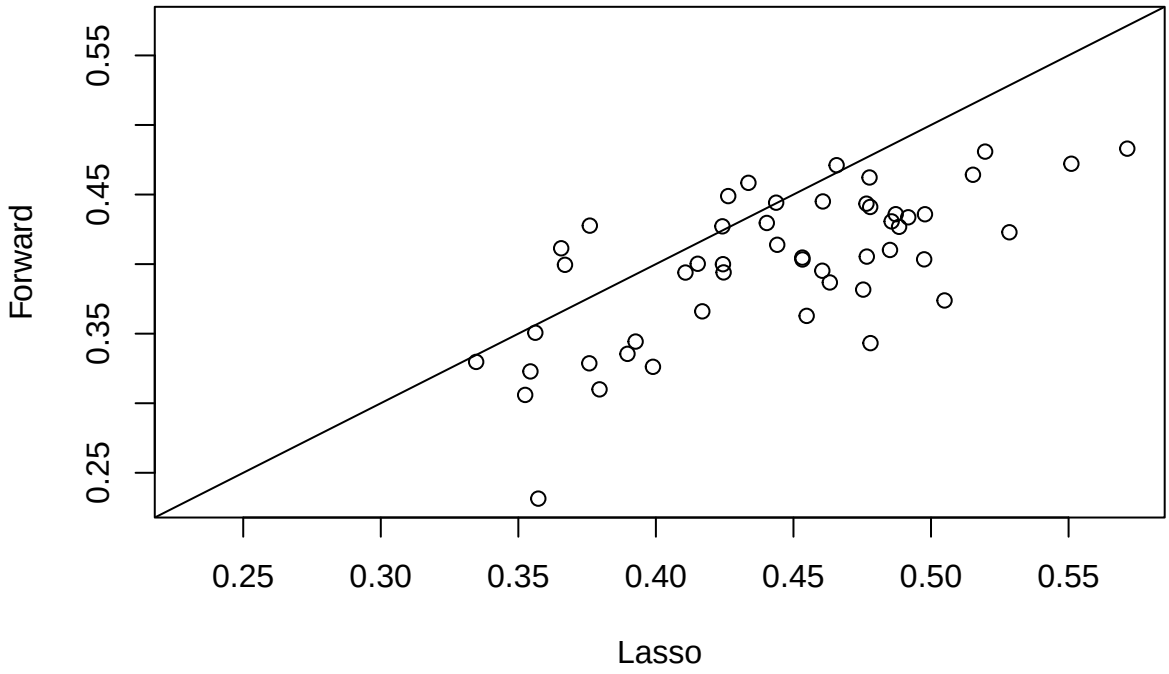
```
TMP=data.frame(method=rep(colnames(COR),each=nrow(COR)),cor=as.vector(COR))
boxplot(cor~method,data=TMP)
```




```
TMP=data.frame(method=rep(colnames(COR),each=nrow(COR)),DF=as.vector(DF))
boxplot(DF~method,data=TMP)
```




```
tmp=range(as.vector(COR),na.rm=T)
plot(COR,xlim=tmp,ylim=tmp);abline(a=0,b=1)
```



Conclussions

- Lasso achieved higher prediction accuracy than the Forward regression in all training-testing partitions.
- The Forward regression achieved its maximum accuracy with smaller DF than Lasso.