scad table 1

Chaorui Zhang

2021/4/27

library(MASS)  
library(lars)

## Loaded lars 1.2

library(msgps)  
library(ncvreg)

## Warning: package 'ncvreg' was built under R version 4.0.5

## generate data

the correlation between and is we choose and and /n’’ and

generate\_data <-function (sigma\_epsilon,n){  
 beta = c(3,1.5,0,0,2,0,0,0)  
 p = length(beta)  
 rho=0.5  
 sigma\_x = matrix(rho,p,p)  
 for(i in seq(p)){  
 for(j in seq(p)){  
 sigma\_x[i,j]=rho\*\*abs(i-j)  
 }  
 }  
 x = mvrnorm(n = n,mu = rep(0,p),Sigma = sigma\_x)  
 y = x%\*%beta + sigma\_epsilon\*rnorm(n=n,mean=0,sd=1)  
 lars(x=x,y=y,type = 'lasso')  
 return(list(x,y,beta))  
 # sigma\_epsl  
}  
RME <- function (real,fore){  
 rme = 0  
 for(i in seq(length(real))){  
 if(real[i]!=0){  
 rme=rme+abs(real[i]-fore[i])/real[i]  
 }else{  
 if(fore[i]!=0){  
 rme =rme+1  
 }  
 }  
 }  
 rme = rme/length(real)  
 # rme = mean(abs(real-fore)/real)  
 return(rme)  
}  
correct\_or\_not <-function (real,fore){  
 correct = 0  
 incorrect = 0  
 for(i in seq(length(real))){  
 if(real[i]==0){  
 if(fore[i]==0){  
 correct = correct+1  
 }  
 }else{  
 if(fore[i]==0){  
 incorrect = incorrect+1  
 }  
 }  
 }  
 return(c(correct,incorrect))  
}

table\_1 <- function (test\_time,sigma\_epsilon,n){  
 options (warn = -1)  
 test\_time = 100  
 lasso\_rme\_list =c()  
 lasso\_correct\_list =c()  
 lasso\_incorrect\_list =c()  
 #lambda is 0.5  
 scad\_0.5\_rme\_list =c()  
 scad\_0.5\_correct\_list =c()  
 scad\_0.5\_incorrect\_list =c()  
 #lambda is choose by cv  
 scad\_cv\_rme\_list =c()  
 scad\_cv\_correct\_list =c()  
 scad\_cv\_incorrect\_list =c()  
 ridge\_rme\_list =c()  
 ridge\_correct\_list =c()  
 ridge\_incorrect\_list =c()  
  
 for(tt in seq(test\_time)){  
 data =generate\_data(sigma\_epsilon,n)  
 beta = data[[3]]  
 lar1 = lars(x=data[[1]],y=data[[2]],type = 'lasso')  
 lasso\_beta\_pre = lar1$beta[which.min(lar1$Cp),]  
 lasso\_rme\_list = c(lasso\_rme\_list,RME(beta,lasso\_beta\_pre))  
 lasso\_correct\_list = c(lasso\_correct\_list,correct\_or\_not(beta,lasso\_beta\_pre)[1])  
 lasso\_incorrect\_list = c(lasso\_incorrect\_list,correct\_or\_not(beta,lasso\_beta\_pre)[2])  
 scad\_0.5 <- ncvreg(X=data[[1]],y=data[[2]],family="gaussian",penalty = "SCAD",lambda = 0.5)  
 scad\_0.5\_beta\_pre = scad\_0.5$beta[2:length(scad\_0.5$beta)]  
 scad\_0.5\_rme\_list = c(scad\_0.5\_rme\_list,RME(beta,scad\_0.5\_beta\_pre))  
 scad\_0.5\_correct\_list = c(scad\_0.5\_correct\_list,correct\_or\_not(beta,scad\_0.5\_beta\_pre)[1])  
 scad\_0.5\_incorrect\_list = c(scad\_0.5\_incorrect\_list,correct\_or\_not(beta,scad\_0.5\_beta\_pre)[2])  
 scad <- cv.ncvreg(X=data[[1]],y=data[[2]],family=c("gaussian"),penalty = "SCAD")  
 scad\_cv <- ncvreg(data[[1]],data[[2]],family="gaussian",penalty = "SCAD",lambda =scad$lambda.min)  
 scad\_cv\_beta\_pre = scad\_cv$beta[2:length(scad\_0.5$beta)]  
 scad\_cv\_rme\_list = c(scad\_cv\_rme\_list,RME(beta,scad\_cv\_beta\_pre))  
 scad\_cv\_correct\_list = c(scad\_cv\_correct\_list,correct\_or\_not(beta,scad\_cv\_beta\_pre)[1])  
 scad\_cv\_incorrect\_list = c(scad\_cv\_incorrect\_list,correct\_or\_not(beta,scad\_cv\_beta\_pre)[2])  
 ridge =lm.ridge(formula = y~.,data=data.frame(y=data[[2]],x=data[[1]]))  
 ridge\_beta\_pre = ridge$coef  
 ridge\_rme\_list = c(ridge\_rme\_list,RME(beta,ridge\_beta\_pre))  
 ridge\_correct\_list = c(ridge\_correct\_list,correct\_or\_not(beta,ridge\_beta\_pre)[1])  
 ridge\_incorrect\_list = c(ridge\_incorrect\_list,correct\_or\_not(beta,ridge\_beta\_pre)[2])  
  
 }  
  
 df = data.frame('Method'=0,'MRME'=0,'correct'=0,'incorrect'=0)  
 df[1,]=list('SCAD 0.5',median(scad\_0.5\_rme\_list),mean(scad\_0.5\_correct\_list),mean(scad\_0.5\_incorrect\_list))  
 df[2,]=list('SCAD cv',median(scad\_cv\_rme\_list),mean(scad\_cv\_correct\_list),mean(scad\_cv\_incorrect\_list))  
 df[3,]=list('LASSO',median(lasso\_rme\_list),mean(lasso\_correct\_list),mean(lasso\_incorrect\_list))  
  
 df[5,]=list('ridge',median(ridge\_rme\_list),mean(ridge\_correct\_list),mean(ridge\_incorrect\_list))  
 print(df)  
}  
table\_1(100,3,40)

## Method MRME correct incorrect  
## 1 SCAD 0.5 0.2379470 4.01 0.08  
## 2 SCAD cv 0.2665739 3.51 0.15  
## 3 LASSO 0.3311490 3.10 0.02  
## 4 <NA> NA NA NA  
## 5 ridge 0.7141705 0.00 0.00

data =generate\_data(3,40)  
scad <- cv.ncvreg(X=data[[1]],y=data[[2]],family=c("gaussian"),penalty = "SCAD")  
# scad\_cv\_min\_lambda <-ncvreg(X=data[[1]],y=data[[2]],family=c("gaussian"),penalty = "SCAD",lambda = 0.5)  
# scad\_cv\_min\_lambda$beta[]  
scad$lambda.min

## [1] 0.1927161