Adaptive Lasso for Cox's proportional hazards model

Zhiguo ZHANG

2021/6/7

目录

```
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 4.0.5
```

```
library(lars)
```

```
## Warning: package 'lars' was built under R version 4.0.3
```

```
## Loaded lars 1.2
```

```
library(tsDyn)
```

```
## Warning: package 'tsDyn' was built under R version 4.0.5
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
library(ncvreg)
```

```
## Warning: package 'ncvreg' was built under R version 4.0.4
```

```
library(leaps)
```

```
## Warning: package 'leaps' was built under R version 4.0.4
```

```
library(MASS)
library(glmnet)
```

```
## Warning: package 'glmnet' was built under R version 4.0.4
```

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

```
## Loaded glmnet 4.1-1
```

```
library(survival)
```

```
##
## Attaching package: 'survival'
```

```
## The following object is masked from 'package:ncvreg':
##
##     heart
```

```
library(locfit)
```

```
## Warning: package 'locfit' was built under R version 4.0.4
```

```
## locfit 1.5-9.4    2020-03-24
```

```
library(stats)
```

```
#模型数据
data_xy_model=function(i,n,m,rate){
  set.seed(i*10)
  #产生生存时间
  #产生样本量 n=100, 服从参数lambda= 2 的指数分布的生存时间
  p=9
  v=0.5
  sigma_z=matrix(0,p,p)
  rho=0.5
  for (t in 1:p) {
    for (j in 1:p) {
      sigma_z[t,j]=rho^(abs(t-j))
```

```
    }
  }
  mu=rep(0,p)
  Z=mvrnorm(n,mu=mu,Sigma=sigma_z)
  # eps=rnorm(n,mean=0,sd=sigma_eps)
  beta=list(c(-0.7,-0.7,0,0,0,-0.7,0,0,0),c(-0.4,-0.3,0,0,0,-0.2,0,0,0))
  u=runif(n,0,1)
  T1=(-log(u)/exp(Z%*%beta[[m]]))^(1/v)

  #产生删失时间
  l=seq(0.01,30,0.01)
  for (r in 1:length(l)) {
    C=runif(n,0,l[r])
    prob=sum(C<T1)/n
    if(prob<=rate){
      break
    }
  }

  #产生生存数据
  time=runif(n,0,1)
  state=rbinom(n,1,rate)
  Y=Surv(time,state)
  Y[,1]=T1
  Y[,2]=1
  for (j in 1:n){
    if(Y[j,1]>C[j]){
      Y[j,2]=0
      Y[j,1]=C[j]
    }
  }
  return(list(Z,Y))
}
```

```
#MSE
MSE=function(beta_hat,m){
  p=9
  sigma_z=matrix(0,p,p)
  rho=0.5
```

```
  for (t in 1:p) {
    for (j in 1:p) {
      sigma_z[t,j]=rho^(abs(t-j))
    }
  }
  beta=list(c(-0.7,-0.7,0,0,0,-0.7,0,0,0),c(-0.4,-0.3,0,0,0,-0.2,0,0,0))
  mse=t(beta_hat-beta[[m]])%*%sigma_z%*%(beta_hat-beta[[m]])
  return(mse)
}
```

```
#表1
table1=function(n,rate,m,times){
  lasso_MSE=rep(NA,times)
  lasso_corr=rep(NA,times)
  lasso_incorr=rep(NA,times)

  adalasso_MSE=rep(NA,times)
  adalasso_corr=rep(NA,times)
  adalasso_incorr=rep(NA,times)

  scad_MSE=rep(NA,times)
  scad_corr=rep(NA,times)
  scad_incorr=rep(NA,times)

  for (i in 1:times) {
    data1=data_xy_model(i,n,m,rate)
    Z=data1[[1]]
    Y=data1[[2]]

    #Lasso
    lasso_cv=cv.glmnet(Z,Y,alpha=1,family="cox")
    lasso_lambda=lasso_cv$lambda[which.min(lasso_cv$cvm)]
    lasso=glmnet(Z,Y,alpha=1,lambda=lasso_lambda,family="cox")
    lasso_beta=lasso$beta
    lasso_MSE[i]=MSE(lasso_beta,m)
    lasso_corr[i]=sum(lasso_beta[-c(1,2,6)]==0)
    lasso_incorr[i]=sum(lasso_beta[c(1,2,6)]==0)
```

```
    #adapt lasso
    ols_coef=lm(Y[,1]~Z+0)$coef
    adaptlasso_cv=cv.glmnet(Z,Y,alpha=1,penalty.factor=abs(ols_coef)^(-1),nfolds=5,family="cox")
    adaptlasso_lambda=adaptlasso_cv$lambda.min
    adaptlasso=glmnet(Z,Y,alpha=1,penalty.factor=abs(ols_coef)^(-1),lambda=adaptlasso_lambda,famil
    adalasso_beta=adaptlasso$beta
    adalasso_MSE[i]=MSE(adalasso_beta,m)
    adalasso_corr[i]=sum(adalasso_beta[-c(1,2,6)]==0)
    adalasso_incorr[i]=sum(adalasso_beta[c(1,2,6)]==0)

    #SCAD
    scad=ncvreg(Z, Y[,1],penalty="SCAD",gamma=3.7)
    scad_cv=cv.ncvreg(Z, Y[,1],penalty="SCAD")
    scad_lambda=scad_cv$lambda.min
    scad_beta=coef(scad, lambda=scad_lambda)[-1]
    scad_MSE[i]=MSE(scad_beta,m)
    scad_corr[i]=sum(scad_beta[-c(1,2,6)]==0)
    scad_incorr[i]=sum(scad_beta[c(1,2,6)]==0)

  }
  result=list(N=n,rate=rate,
              LASSO=c(mean(lasso_corr),mean(lasso_incorr),mean(lasso_MSE),sd(lasso_MSE)),
              SCAD=c(mean(scad_corr),mean(scad_incorr),mean(scad_MSE),sd(scad_MSE)),
              ALASSO=c(mean(adalasso_corr),mean(adalasso_incorr),mean(adalasso_MSE),sd(adalasso_MS
  return(result)
}
```

```
#table1 result
times=100
m=1
list(table1(100,0.25,m,times),
table1(200,0.25,m,times),
table1(300,0.25,m,times),
table1(100,0.4,m,times),
table1(200,0.4,m,times),
table1(300,0.4,m,times))
```

```
## [[1]]
## [[1]]$N
```

```
## [1] 100
##
## [[1]]$rate
## [1] 0.25
##
## [[1]]$LASSO
## [1] 3.6700000 0.0000000 0.1339416 0.0973271
##
## [[1]]$SCAD
## [1] 3.740000 0.180000 6.865384 3.009000
##
## [[1]]$ALASSO
## [1] 3.8700000 0.0700000 0.1596218 0.1355845
##
##
## [[2]]
## [[2]]$N
## [1] 200
##
## [[2]]$rate
## [1] 0.25
##
## [[2]]$LASSO
## [1] 3.61000000 0.00000000 0.06471181 0.03838980
##
## [[2]]$SCAD
## [1] 4.100000 0.000000 7.651089 2.886639
##
## [[2]]$ALASSO
## [1] 4.82000000 0.00000000 0.05029614 0.03254845
##
##
## [[3]]
## [[3]]$N
## [1] 300
##
## [[3]]$rate
## [1] 0.25
##
```

```
## [[3]]$LASSO
## [1] 3.37000000 0.00000000 0.04627752 0.02979916
##
## [[3]]$SCAD
## [1] 4.480000 0.000000 7.826984 2.471628
##
## [[3]]$ALASSO
## [1] 4.96000000 0.01000000 0.03543030 0.04569909
##
##
## [[4]]
## [[4]]$N
## [1] 100
##
## [[4]]$rate
## [1] 0.4
##
## [[4]]$LASSO
## [1] 3.7700000 0.0000000 0.1629936 0.1046385
##
## [[4]]$SCAD
## [1] 3.920000 0.280000 2.930250 0.539857
##
## [[4]]$ALASSO
## [1] 4.0400000 0.1400000 0.1936344 0.1660790
##
##
## [[5]]
## [[5]]$N
## [1] 200
##
## [[5]]$rate
## [1] 0.4
##
## [[5]]$LASSO
## [1] 3.66000000 0.00000000 0.07803909 0.04884802
##
## [[5]]$SCAD
## [1] 4.0400000 0.0700000 3.0471856 0.3756278
```

```
##
## [[5]]$ALASSO
## [1] 4.35000000 0.01000000 0.06714791 0.05536094
##
##
## [[6]]
## [[6]]$N
## [1] 300
##
## [[6]]$rate
## [1] 0.4
##
## [[6]]$LASSO
## [1] 3.36000000 0.00000000 0.05054243 0.02985091
##
## [[6]]$SCAD
## [1] 4.1500000 0.0100000 3.1028831 0.3575721
##
## [[6]]$ALASSO
## [1] 4.60000000 0.00000000 0.04056844 0.02917639
```

```r
#表2
table2=function(n,rate,m,times){
  lasso_fre=rep(0,9)
  adalasso_fre=rep(0,9)
  scad_fre=rep(0,9)

  for (i in 1:times) {
    data1=data_xy_model(i=i,n=n,m=m,rate=rate)
    Z=data1[[1]]
    Y=data1[[2]]

    #Lasso
    lasso_cv=cv.glmnet(Z,Y,alpha=1,family="cox")
    lasso_lambda=lasso_cv$lambda[which.min(lasso_cv$cvm)]
    lasso=glmnet(Z,Y,alpha=1,lambda=lasso_lambda,family="cox")
    lasso_beta=lasso$beta
    for (j in 1:9) {
      if(lasso_beta[j]!=0){
```

```
        lasso_fre[j]=lasso_fre[j]+1
      }
    }

    #SCAD
    scad=ncvreg(Z, Y[,1],penalty="SCAD",gamma=3.7)
    scad_cv=cv.ncvreg(Z, Y[,1],penalty="SCAD")
    scad_lambda=scad_cv$lambda.min
    scad_beta=coef(scad, lambda=scad_lambda)[-1]
    for (j in 1:9) {
      if(scad_beta[j]!=0){
        scad_fre[j]=scad_fre[j]+1
      }
    }

    #adapt lasso
    ols_coef=lm(Y[,1]~Z+0)$coef
    adaptlasso_cv=cv.glmnet(Z,Y,alpha=1,penalty.factor=abs(ols_coef)^(-1),nfolds=5,family="cox")
    adaptlasso_lambda=adaptlasso_cv$lambda.min
    adaptlasso=glmnet(Z,Y,alpha=1,penalty.factor=abs(ols_coef)^(-1),lambda=adaptlasso_lambda,famil
    adalasso_beta=adaptlasso$beta
    for (j in 1:9) {
      if(adalasso_beta[j]!=0){
        adalasso_fre[j]=adalasso_fre[j]+1
      }
    }
  }
  result=list(N=n,
              LASSO=c(lasso_fre),
              SCAD=c(scad_fre),
              ALASSO=c(adalasso_fre))
  return(result)
}
```

```
#table2 result
times=100
m=1
list(table2(100,0.25,m,times),
     table2(200,0.25,m,times),
```

```
    table2(300,0.25,m,times))
```

```
## [[1]]
## [[1]]$N
## [1] 100
##
## [[1]]$LASSO
## [1] 100 100  42  35  40 100  35  38  43
##
## [[1]]$SCAD
## [1] 95 90 39 32 31 96 32 37 41
##
## [[1]]$ALASSO
## [1] 98 97 33 38 29 97 37 30 33
##
##
## [[2]]
## [[2]]$N
## [1] 200
##
## [[2]]$LASSO
## [1] 100 100  43  36  40 100  41  41  38
##
## [[2]]$SCAD
## [1] 100 100  31  31  25 100  32  37  34
##
## [[2]]$ALASSO
## [1] 100 100  16  15  14 100  18  17  14
##
##
## [[3]]
## [[3]]$N
## [1] 300
##
## [[3]]$LASSO
## [1] 100 100  47  49  39 100  45  42  41
##
## [[3]]$SCAD
## [1] 100 100  20  30  22 100  29  29  34
```

```
##
## [[3]]$ALASSO
## [1] 100 100  12  18  11  99  20  20  19
```

```r
#表3
table3=function(n,rate,m,times){
  lasso_beta=matrix(NA,times,3)
  adalasso_beta=matrix(NA,times,3)
  scad_beta=matrix(NA,times,3)
  lasso_ase=rep(NA,3)
  scad_ase=rep(NA,3)
  adalasso_ase=matrix(NA,times,3)
  beta=list(c(-0.7,-0.7,0,0,0,-0.7,0,0,0),c(-0.4,-0.3,0,0,0,-0.2,0,0,0))
  for (i in 1:times) {
    data1=data_xy_model(i=i,n=n,m=m,rate=rate)
    Z=data1[[1]]
    Y=data1[[2]]

    #Lasso
    lasso_cv=cv.glmnet(Z,Y,alpha=1,family="cox")
    lasso_lambda=lasso_cv$lambda[which.min(lasso_cv$cvm)]
    lasso=glmnet(Z,Y,alpha=1,lambda=lasso_lambda,family="cox")
    lasso_beta[i,1]=lasso$beta[1]
    lasso_beta[i,2]=lasso$beta[2]
    lasso_beta[i,3]=lasso$beta[6]

    #SCAD
    scad=ncvreg(Z, Y[,1],penalty="SCAD",gamma=3.7)
    scad_cv=cv.ncvreg(Z, Y[,1],penalty="SCAD")
    scad_lambda=scad_cv$lambda.min
    scad.beta=coef(scad, lambda=scad_lambda)[-1]
    scad_beta[i,1]=scad.beta[1]
    scad_beta[i,2]=scad.beta[2]
    scad_beta[i,3]=scad.beta[6]

    #adapt lasso
    ols_coef=lm(Y[,1]~Z+0)$coef
    adaptlasso_cv=cv.glmnet(Z,Y,alpha=1,penalty.factor=abs(ols_coef)^(-1),nfolds=5,family="cox")
    adaptlasso_lambda=adaptlasso_cv$lambda.min
```

```
    adalasso=glmnet(Z,Y,alpha=1,penalty.factor=abs(ols_coef)^(-1),lambda=adaptlasso_lambda,family=
    adalasso_beta[i,1]=adalasso$beta[1]
    adalasso_beta[i,2]=adalasso$beta[2]
    adalasso_beta[i,3]=adalasso$beta[6]
    adalasso_ase[i,1]=abs(adalasso$beta[1]-beta[[m]][1])
    adalasso_ase[i,2]=abs(adalasso$beta[2]-beta[[m]][2])
    adalasso_ase[i,3]=abs(adalasso$beta[6]-beta[[m]][6])
  }
  lasso_ase[1]=sqrt(mean((lasso_beta[,1]-beta[[m]][1])^2))
  lasso_ase[2]=sqrt(mean((lasso_beta[,2]-beta[[m]][2])^2))
  lasso_ase[3]=sqrt(mean((lasso_beta[,3]-beta[[m]][6])^2))
  lasso_mad_1=mad(lasso_beta[,1])/1.4826/0.6745
  lasso_mad_2=mad(lasso_beta[,2])/1.4826/0.6745
  lasso_mad_3=mad(lasso_beta[,3])/1.4826/0.6745
  scad_ase[1]=sqrt(mean((scad_beta[,1]-beta[[m]][1])^2))
  scad_ase[2]=sqrt(mean((scad_beta[,2]-beta[[m]][2])^2))
  scad_ase[3]=sqrt(mean((scad_beta[,3]-beta[[m]][6])^2))
  scad_mad_1=mad(scad_beta[,1])/1.4826/0.6745
  scad_mad_2=mad(scad_beta[,2])/1.4826/0.6745
  scad_mad_3=mad(scad_beta[,3])/1.4826/0.6745
  adalasso_mad_1=mad(adalasso_beta[,1])/1.4826/0.6745
  adalasso_mad_2=mad(adalasso_beta[,2])/1.4826/0.6745
  adalasso_mad_3=mad(adalasso_beta[,3])/1.4826/0.6745
  result=list(N=n,
              LASSO=c(lasso_mad_1,lasso_ase[1],
                      lasso_mad_2,lasso_ase[2],
                      lasso_mad_3,lasso_ase[3]),
              SCAD=c(scad_mad_1,scad_ase[1],
                      scad_mad_2,scad_ase[2],
                      scad_mad_3,scad_ase[3]),
              ALASSO=c(adalasso_mad_1,apply(adalasso_ase,2,mean)[1],
                       adalasso_mad_2,apply(adalasso_ase,2,mean)[2],
                       adalasso_mad_3,apply(adalasso_ase,2,mean)[3]))
}
```

```
#table3 result
times=100
m=1
list(table3(100,0.25,m,times),
```

```
    table3(200,0.25,m,times),
    table3(300,0.25,m,times))
```

```
## [[1]]
## [[1]]$N
## [1] 100
##
## [[1]]$LASSO
## [1] 0.1598919 0.1889406 0.1562115 0.1849151 0.1544657 0.2144567
##
## [[1]]$SCAD
## [1] 0.3612965 1.2879541 0.4263913 1.2729028 0.4479873 1.3246434
##
## [[1]]$ALASSO
## [1] 0.2085276 0.1706096 0.2013739 0.1896768 0.1895452 0.1776023
##
##
## [[2]]
## [[2]]$N
## [1] 200
##
## [[2]]$LASSO
## [1] 0.12875388 0.12832869 0.11369227 0.13709926 0.09160785 0.14508600
##
## [[2]]$SCAD
## [1] 0.2259000 1.3523296 0.2878167 1.3659731 0.2340661 1.3540166
##
## [[2]]$ALASSO
## [1] 0.10649110 0.10118649 0.13289940 0.10815069 0.11131454 0.09577207
##
##
## [[3]]
## [[3]]$N
## [1] 300
##
## [[3]]$LASSO
## [1] 0.07634139 0.10659110 0.09785275 0.10261654 0.08683963 0.12078139
##
## [[3]]$SCAD
```

```
## [1] 0.2286211 1.3865765 0.2525005 1.3701606 0.1884524 1.3502352
##
## [[3]]$ALASSO
## [1] 0.08049327 0.07684410 0.10406714 0.07755377 0.07303353 0.08800625
```

```
#表4
#table4 result
times=100
m=2
list(table1(200,0.25,m,times),
     table1(300,0.25,m,times),
     table1(200,0.4,m,times),
     table1(300,0.4,m,times))
```

```
## [[1]]
## [[1]]$N
## [1] 200
##
## [[1]]$rate
## [1] 0.25
##
## [[1]]$LASSO
## [1] 3.97000000 0.14000000 0.04581189 0.02860607
##
## [[1]]$SCAD
## [1] 3.8700000 0.5500000 1.2523621 0.3817622
##
## [[1]]$ALASSO
## [1] 4.4500000 0.4600000 0.0575580 0.0354363
##
##
## [[2]]
## [[2]]$N
## [1] 300
##
## [[2]]$rate
## [1] 0.25
##
## [[2]]$LASSO
```

```
## [1] 3.72000000 0.08000000 0.03236527 0.01847241
##
## [[2]]$SCAD
## [1] 3.9100000 0.2800000 1.3006336 0.3061811
##
## [[2]]$ALASSO
## [1] 4.26000000 0.26000000 0.03735732 0.02671019
##
##
## [[3]]
## [[3]]$N
## [1] 200
##
## [[3]]$rate
## [1] 0.4
##
## [[3]]$LASSO
## [1] 4.16000000 0.31000000 0.06233871 0.04532618
##
## [[3]]$SCAD
## [1] 4.1800000 0.6900000 0.6557120 0.1002886
##
## [[3]]$ALASSO
## [1] 4.30000000 0.64000000 0.07218104 0.04697357
##
##
## [[4]]
## [[4]]$N
## [1] 300
##
## [[4]]$rate
## [1] 0.4
##
## [[4]]$LASSO
## [1] 3.66000000 0.10000000 0.03957489 0.02167880
##
## [[4]]$SCAD
## [1] 3.66000000 0.40000000 0.67245062 0.08354008
##
```

```
## [[4]]$ALASSO
## [1] 3.97000000 0.34000000 0.04760092 0.03071185
```