ST509_HW3_2024020409

Hwijun Kwon

2024-03-29

1

```
S <- function(z, lambda){
    sign(z) * max(abs(z) - lambda, 0)
}


S(z = 3,lambda = 10)
```

```
## [1] 0
```

```
cd.lasso <- function(x, y, lambda){
  # centering y and standardizing X
  z <- scale(x)
  m <- attr(z, "scaled:center") # save original mean X
  s <- attr(z, "scaled:scale") # save original scale of Y
  u <- (y-mean(y))

  #initialization
  beta <- coef(lm(u ~ z - 1))  # -1 : exclude intercept
  r <- u - z %*% beta

  for (iter in 1:100) {
    new.beta <- beta
    # coordinate 1 to p
    for (j in 1:p){
      # Update coefficients
      temp <- beta[j] + crossprod(z[, j], r)/n
      new.beta[j] <- S(temp, lambda/s[j])
      # Update residuals
      r <- r - (new.beta[j] - beta[j]) * z[, j]
    }

    delta <- max(abs(new.beta - beta))
    if (delta < 1.0e-3) break
    beta <- new.beta
  }

  # Transform back to the original scale
  beta <- new.beta/s
  beta0 <- mean(y) - crossprod(beta, m)
  c(beta0, beta)
```

```r
}
```

```r
set.seed(1) ; n <- 100 ; p <- 5
x <- matrix(rnorm(n*p, 1, 1), n, p) # X : 100 * 5 matrix
e <- rnorm(n, 0, 0.5) # noise

true.beta <- rep(0, p+1) ;
true.beta[1] <- 1 # intercept
true.beta[2:(p+1)] <- c(rep(1, 3), rep(0, p-3))
true.beta
```

```
## [1] 1 1 1 1 0 0
```

```r
y <- true.beta[1] + x %*% true.beta[-1] + e
```

```r
library(glmnet)
```

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

```
## Loaded glmnet 4.1-8
```

```r
est0 <- coef(lm(y ~ x))
est1 <- cd.lasso(x, y, lambda = 0.1)
est2 <- coef(glmnet(x, y, lambda = 0.1, standardize = F))

result <- cbind(true.beta, est0, est1, est2)
colnames(result) <- c("true", "lm", "ours", "glmnet")
rownames(result) <- 0:p
print(round(result, 4))
```

```
## 6 x 4 sparse Matrix of class "dgCMatrix"
##   true      lm   ours glmnet
## 0    1  1.1402 1.4218 1.4218
## 1    1  0.9440 0.8180 0.8180
## 2    1  0.9909 0.8733 0.8733
## 3    1  0.9870 0.8831 0.8831
## 4    . -0.0480 .      .
## 5    . -0.0288 .      .
```

(A)

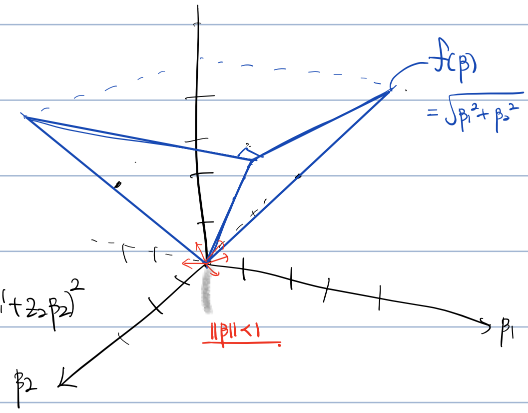
```r
#pd.elastic <- function()
```

(by glmnet)

```r
train <- matrix(scan("train.txt"), 500, 51)
test  <- matrix(scan("test.txt"), 500, 51)
x <- train[,-51] ; y <- train[,51]
x.test <- test[,-51]
y.test <- test[,51]
```

```r
library(glmnet)
alpha = 0.5 # alpha for elastic net
```

#2



(a) $f(\beta)=\sqrt{\beta_1^2+\beta_2^2}$ $(\beta=(\beta_1,\beta_2))$

vector z is subgradient of at 0 when $f(\beta')\geq f(0)+z^T(\beta'-0)$

$\Leftrightarrow f(\beta')\geq z^T\beta'$ for every β'

$\sqrt{\beta_1'^2+\beta_2'^2}\geq z_1\beta_1'+z_2\beta_2' \Leftrightarrow \beta_1'^2+\beta_2'^2\geq(z_1\beta_1'+z_2\beta_2')^2$

by Cauchy-Schwarz $(z_1^2+z_2^2)(\beta_1^2+\beta_2^2)\geq(z_1\beta_1+z_2\beta_2)^2$

When $z_1^2+z_2^2\leq 1$, then $z=(z_1,z_2)^T$ is subgradient

(b) Under the orthogonality condition $Z_j Z_j - 1$, Show that $-Z_j^T(r_j-Z_j\hat{\beta}_j)+\lambda\hat{s}_j=0$, where $\hat{\beta}_j=(1-\lambda/\|Z_j^Tr_j\|_2)_+Z_j^Tr_j$

Group lasso solves : $\min_\beta \frac{1}{2}\|y-\sum_{j=1}^{J}Z_j\beta_j\|^2+\lambda\sum_{j=1}^{J}\|\beta_j\|_2$, where $\|\beta\|_2=\sqrt{\beta_1^2+\cdots+\beta_p^2}$

Subgradient equation : $-Z_j^T(y-\sum_{j=1}^{J}Z_j\hat{\beta}_j)+\lambda\hat{s}_j=0$, for $j=1,\cdots,J$, where $\hat{s}_j=\begin{cases}\frac{\hat{\beta}_j}{\|\hat{\beta}_j\|} & \text{when } \hat{\beta}_j\neq 0\\ \text{any vector with } \|\hat{s}_j\|\leq 1 & \text{when } \hat{\beta}_j=0\end{cases}$

$-$ $= -Z_j^T(y-\sum_{k\neq j}Z_k\hat{\beta}_k-Z_j\hat{\beta}_j)+\lambda\hat{s}_j$

$= -Z_j^T(r_j-Z_j\hat{\beta}_j)+\lambda\hat{s}_j$

$= -Z_j^Tr_j+\hat{\beta}_j+\lambda\hat{s}_j$

i) $\hat{\beta}_j\neq 0$ $(\Leftrightarrow 1-\lambda/\|Z_j^Tr_j\|>0$

① $-Z_j^Tr_j+\hat{\beta}_j+\lambda\hat{s}_j = -Z_j^Tr_j+\hat{\beta}_j+\lambda\frac{\hat{\beta}_j}{\|\hat{\beta}_j\|}$

$\|\hat{\beta}_j\|_2=\|(1-\frac{\lambda}{\|Z_j^Tr_j\|_2})Z_j^Tr_j\| = (1-\frac{\lambda}{\|Z_j^Tr_j\|_2})\|Z_j^Tr_j\| = \|Z_j^Tr_j\|-\lambda$ o|RZ

① $= -Z_j^Tr_j+(1+\lambda/(\|Z_j^Tr_j\|-\lambda))(1-\lambda/\|Z_j^Tr_j\|_2)Z_j^Tr_j$

$= -Z_j^Tr_j+\left(\frac{\|Z_j^Tr_j\|}{\|Z_j^Tr_j\|-\lambda}\cdot\frac{\|Z_j^Tr_j\|-\lambda}{\|Z_j^Tr_j\|_2}\right)\cdot Z_j^Tr_j$

$= -Z_j^Tr_j+Z_j^Tr_j$

$= 0.$

ii) $(\hat{\beta}_j=0)\Leftrightarrow\left(1-\frac{\lambda}{\|Z_j^Tr_j\|}<0\right)\Leftrightarrow\frac{\|Z_j^Tr_j\|}{\lambda}<1$