DataMiningHW01

Xue Wu

9/7/2020

The Function of Coefficients Estimation Using successive orthogonalization

```
estimate_beta<- function(Y, X_Design){

  beta_index<- seq(1:ncol(X_Design))
  y_prime<- matrix(0, ncol = ncol(X_Design), nrow = nrow(X_Design))
  xp_prime<- matrix(0, ncol = ncol(X_Design), nrow = nrow(X_Design))
  coef<- NULL
  for (i in 1:ncol(X_Design)){
    l<- lm(Y~X_Design[,beta_index[beta_index!=i]]-1)
    y_prime[,i]<- l$residuals
    l2<- lm(X_Design[,i]~X_Design[,beta_index[beta_index!=i]]-1)
    xp_prime[,i] <- l2$residuals
    l3 <- lm(y_prime[,i]~xp_prime[,i]-1)
    coef[i]<- l3$coefficients

  }
  coef
}
```

Data Simulation

```
library(mvtnorm)
Sigma <- matrix(c(4,2,2,3), ncol=2)
X <- rmvnorm( n = 100, mean = rep(0, nrow(Sigma)), sigma = Sigma)
Beta <- c(2,3,4)
X_Design <- cbind(1, X)
eps <- rnorm(100, 0, 1)
Y <- X_Design %*% Beta + eps
```

Comparison of Results between one-stage regression and two-stage regression

```
estimate_beta(Y,X_Design)
```

```
## [1] 2.046730 2.946633 3.921280
```

```
lm(Y~X[,1]+X[,2])
```

```
##
```

```
## Call:
## lm(formula = Y ~ X[, 1] + X[, 2])
##
## Coefficients:
## (Intercept)       X[, 1]       X[, 2]
##       2.047        2.947        3.921
```

As we can see, the estimated coefficients obtained from the two-stage regression function are the same as that obtained from one-stage regression `lm` function.