Homework01

JongCheolLee

2024-09-30

Data

```
library(MASS)
data(Boston)
```

```
head(Boston)
```

```
##      crim zn indus chas   nox    rm  age    dis rad tax ptratio  black lstat
## 1 0.00632 18  2.31    0 0.538 6.575 65.2 4.0900   1 296    15.3 396.90  4.98
## 2 0.02731  0  7.07    0 0.469 6.421 78.9 4.9671   2 242    17.8 396.90  9.14
## 3 0.02729  0  7.07    0 0.469 7.185 61.1 4.9671   2 242    17.8 392.83  4.03
## 4 0.03237  0  2.18    0 0.458 6.998 45.8 6.0622   3 222    18.7 394.63  2.94
## 5 0.06905  0  2.18    0 0.458 7.147 54.2 6.0622   3 222    18.7 396.90  5.33
## 6 0.02985  0  2.18    0 0.458 6.430 58.7 6.0622   3 222    18.7 394.12  5.21
##   medv
## 1 24.0
## 2 21.6
## 3 34.7
## 4 33.4
## 5 36.2
## 6 28.7
```

```
?Boston
```

```
y <- Boston[, 1]
x <- Boston[, -c(1,4,9)]
x <- as.matrix(scale(x))
```

- 506 observations with 14 variables
- **crim**: response variable (11 variables are scaled predictors)

```
dim(x)
```

```
## [1] 506  11
```

```
length(y)
```

```
## [1] 506
```

```
class(x)
```

```
## [1] "matrix" "array"
```

```
colnames(x)
```

```
##  [1] "zn"      "indus"   "nox"     "rm"      "age"     "dis"     "tax"
##  [8] "ptratio" "black"   "lstat"   "medv"
```

Housing Values in Suburbs of Boston

- **crim**: 마을별 1인당 범죄율.
- **zn**: 25,000 평방 피트 이상의 주거용 토지 비율.
- **indus**: 마을별 비소매업 상업 지역의 비율.
- **nox**: 대기 중 산화질소 농도(10백만분율).
- **rm**: 주택당 평균 방 개수.
- **age**: 1940년 이전에 건설된 자가 거주 주택 비율.
- **dis**: 보스턴의 5개 고용 센터까지의 가중 평균 거리.
- **tax**: 10,000달러당 재산세율.
- **ptratio**: 마을별 학생-교사 비율.
- **black**: 1000(Bk⊠0.63)^2 여기서 Bk는 마을별 흑인 인구 비율.
- **lstat**: 저소득층 인구 비율(백분율).
- **medv**: 자가 소유 주택의 중위값(단위: 1,000달러).

Goal Boston의 각 마을의 1인당 범죄율 예측

```
apply(x, 2, function(t) sum(is.na(t)))
```

```
##      zn   indus     nox      rm     age     dis     tax ptratio   black   lstat
##       0       0       0       0       0       0       0       0       0       0
##    medv
##       0
```

```
head(x, 3)
```

```
##           zn      indus        nox        rm        age      dis        tax
## 1  0.2845483 -1.2866362 -0.1440749 0.4132629 -0.1198948 0.140075 -0.6659492
## 2 -0.4872402 -0.5927944 -0.7395304 0.1940824  0.3668034 0.556609 -0.9863534
## 3 -0.4872402 -0.5927944 -0.7395304 1.2814456 -0.2655490 0.556609 -0.9863534
##      ptratio     black      lstat       medv
## 1 -1.4575580 0.4406159 -1.0744990  0.1595278
## 2 -0.3027945 0.4406159 -0.4919525 -0.1014239
## 3 -0.3027945 0.3960351 -1.2075324  1.3229375
```

Question 1.

```r
x0 <- x[1,]
x0
```

```
##         zn      indus        nox         rm        age        dis        tax
##  0.2845483 -1.2866362 -0.1440749  0.4132629 -0.1198948  0.1400750 -0.6659492
##    ptratio      black      lstat       medv
## -1.4575580  0.4406159 -1.0744990  0.1595278
```

```r
train <- x[-1,]
dim(train)
```

```
## [1] 505  11
```

```r
head(train, 3)
```

```
##           zn      indus        nox        rm        age      dis        tax
## 2 -0.4872402 -0.5927944 -0.7395304 0.1940824  0.3668034 0.556609 -0.9863534
## 3 -0.4872402 -0.5927944 -0.7395304 1.2814456 -0.2655490 0.556609 -0.9863534
## 4 -0.4872402 -1.3055857 -0.8344581 1.0152978 -0.8090878 1.076671 -1.1050216
##      ptratio     black      lstat       medv
## 2 -0.3027945 0.4406159 -0.4919525 -0.1014239
## 3 -0.3027945 0.3960351 -1.2075324  1.3229375
## 4  0.1129203 0.4157514 -1.3601708  1.1815886
```

```r
dist_1 <- function(data) {
  test <- data[1,]
  train <- data[-1,]
  diff <- train - matrix(rep(test, nrow(train)), nrow=nrow(train), byrow=T)
  d_505 <- apply(abs(diff), 1, sum)
  d_505 <- as.numeric(d_505)
  return(d_505)
}

d1_vector <- dist_1(x)

close_10 <- order(d1_vector,
                  decreasing = F)[1:10]
close_10_idx <- close_10+1

fhat_1 <- mean(y[close_10_idx])

fhat_1
```

test chunk

```
## [1] 0.115894
```

```
dist_2 <- function(data) {
  test <- data[1,]
  train <- data[-1,]
  diff <- train - matrix(rep(test, nrow(train)), nrow=nrow(train), byrow=T)
  d_505 <- sqrt(apply(diff^2, 1, sum))
  d_505 <- as.numeric(d_505)
  return(d_505)
}

#sum(dist_2(x) != sqrt(dif2))

d2_vector <- dist_2(x)

close2_10 <- order(d2_vector,
                   decreasing = F)[1:10]
close2_10_idx <- close2_10+1

fhat_2 <- mean(y[close2_10_idx])

fhat_2
```

```
## [1] 0.201866
```

```
dist_3 <- function(data) {
  test <- data[1,]
  train <- data[-1,]
  diff <- train - matrix(rep(test, nrow(train)), nrow=nrow(train), byrow=T)
  numer <- abs(diff)
  denom <- abs(train) + abs(matrix(rep(test, nrow(train)), nrow=nrow(train), byrow=T))
  d_505 <- apply((numer/denom), 1, sum)
  d_505 <- as.numeric(d_505)
  return(d_505)
}

d3_vector <- dist_3(x)
close3_10 <- order(d3_vector,
                   decreasing = F)[1:10]
close3_10_idx <- close3_10+1

fhat_3 <- mean(y[close3_10_idx])

fhat_3
```

```
## [1] 0.074659
```