SWEEP Operator

Nicholus Tint Zaw

The sweep operator is part of the linear vs. non-linear model day one lecture topics, and this rmd file was created for self-learning purposes.

Data Processing & Regression model

```
# load the dataset
df <- read.csv("Chicago_Taxi_Trip.csv")

df <- df %>%
  mutate(Payment_Method_Cash = ifelse(Payment_Method == "Cash", 1, 0),
         Payment_Method_Credit_Card = ifelse(Payment_Method == "Credit Card", 1, 0),
         intercept = 1)

reg <- lm(Trip_Payment ~ Payment_Method_Cash + Payment_Method_Credit_Card +
            Trip_Minutes + Trip_Miles, df)
summary(reg)
```

```
##
## Call:
## lm(formula = Trip_Payment ~ Payment_Method_Cash + Payment_Method_Credit_Card +
##     Trip_Minutes + Trip_Miles, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.72815 -0.45857  0.09088  0.54013  1.60581
##
## Coefficients: (1 not defined because of singularities)
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 12.27128    1.81156   6.774 4.47e-06 ***
## Payment_Method_Cash         -7.85429    0.54291 -14.467 1.31e-10 ***
## Payment_Method_Credit_Card       NA         NA      NA       NA
## Trip_Minutes                 0.05787    0.03130   1.849   0.0831 .
## Trip_Miles                   2.13158    0.09974  21.371 3.43e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.211 on 16 degrees of freedom
## Multiple R-squared:  0.9794, Adjusted R-squared:  0.9756
## F-statistic: 253.7 on 3 and 16 DF,  p-value: 1.067e-13
```

Practice SWEEP Code

```
# Predictors and Outcome combined matrix
z <- df %>%
  dplyr::select(intercept, Payment_Method_Cash, Payment_Method_Credit_Card,
                Trip_Minutes, Trip_Miles, Trip_Payment) %>%
  relocate(intercept, Payment_Method_Cash, Payment_Method_Credit_Card,
           Trip_Minutes, Trip_Miles, Trip_Payment)

zt <- t(z)
Z <- zt %*% as.matrix(z)
```

```
# sweep one row by each time using sweep.operator function from fastmatrix pkg
# Intercept
k = c(1)
z <- sweep.operator(as.matrix(Z), k = k)

# Payment_Method_Cash
k = c(1, 2)
z <- sweep.operator(as.matrix(Z), k = k)

print(0.000000e+00)
```

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

```
# Payment_Method_Credit_Card
k = c(1, 2, 3)
z <- sweep.operator(as.matrix(Z), k = k) # result error because sweep operator row was 0
```

```
## Error in sweep.operator(as.matrix(Z), k = k): symmetric sweep operator gave code 3
```

```
# Trip_Minutes
k = c(1, 2, 4)
z <- sweep.operator(as.matrix(Z), k = k)

# Trip_Miles
k = c(1, 2, 4, 5)
z <- sweep.operator(as.matrix(Z), k = k)
```

```
# customized function
# ref: https://www.stat.cmu.edu/~brian/711/week06/linreg-1.pdf

sweepk <- function(A,k) {

  n <- dim(A)[1]
  if (n!=dim(A)[2]) stop("A not square!")
```

```
  b <- abs(A[k,k])
  if(b==0) stop("Can't sweep on zero!")

  A[k,] <- A[k,]/b

  A[,k] <- A[,k]/b

  others <- (1:n)[-k]

  for (i in others) {
    for (j in others) {
      A[i,j] <- A[i,j] - A[i,k]*A[k,j]*b
      }
  }

  A[k,k] <- -1/b

  return(A)

}
```

```
# step-by-step sweep operation
sweep_mat_1 <- sweepk(as.matrix(Z), 1)
sweep_mat_2 <- sweepk(as.matrix(sweep_mat_1), 2)
sweep_mat_3 <- sweepk(as.matrix(sweep_mat_2), 3) # will get the error message as the operator
```

```
## Error in sweepk(as.matrix(sweep_mat_2), 3): Can't sweep on zero!
```

```
sweep_mat_4 <- sweepk(as.matrix(sweep_mat_2), 4)
sweep_mat_5 <- sweepk(as.matrix(sweep_mat_4), 5)
```

```
# looping for all predictors
p <- nrow(Z) - 1
sweep_mat <- Z

for (x in 1:p){
  print(x) # loop check
  if(sweep_mat[x,x] != 0){
    print(sweep_mat[x,x]) # condition check
    sweep_mat <- sweepk(as.matrix(sweep_mat), x)
  }

}
```

```
## [1] 1
## [1] 20
## [1] 2
```

```
## [1] 5
## [1] 3
## [1] 4
## [1] 1603
## [1] 5
## [1] 147.3365
```

```
#landscape(knitr::kable(head(sweep_mat), "latex"))
# ref: https://stackoverflow.com/questions/32265676/longtable-in-a-knitr-pdf-document-using-
kable(sweep_mat, "latex", longtable = T, booktabs = T) %>%
  kable_styling(latex_options = c("repeat_header"), font_size = 9) %>%
  landscape()
```

	intercept	Payment_Method_Cash	Payment_Method_Credit_Card	Trip_Minutes	Trip_Miles	Trip_Payment
intercept	-2.2389535	0.1440362	1	0.0161846	0.0918457	12.2712763
Payment_Method_Cash	0.1440362	-0.2010957	-1	-0.0006546	-0.0011576	-7.8542943
Payment_Method_Credit_Card	1.0000000	-1.0000000	0	0.0000000	0.0000000	0.0000000
Trip_Minutes	0.0161846	-0.0006546	0	-0.0006686	0.0005512	0.0578697
Trip_Miles	0.0918457	-0.0011576	0	0.0005512	-0.0067872	2.1315752
Trip_Payment	12.2712763	-7.8542943	0	0.0578697	2.1315752	23.4519289