# Media Mix Modeling

Lunjing Yuan

January 20, 2021

```r
# get the required library
library(tidyverse)
library(magrittr)
library(caret)

cardf = read.csv("sales_ads_use.csv", header=TRUE, sep=",")


# AdStockScale() Returns the scale matrix for creating the adstock variable
from the marketing per-period expenses
# Input:
#    lambda - the discounted value
#    n - the length of the time series

AdStockScale <- function(lambda, n) {

  r <- lambda^(seq_len(n)-1)
  m <- matrix(rep(r,n),nrow=n)

  z <- matrix(0,nrow=n,ncol=n)
  z[lower.tri(z,diag=TRUE)] <- m[row(m) <= (n+1-col(m))]
  z
}
```

To create the AdStock $G_t$ matrix from the current per-period marketing matrix $M_t$, all you need is to do the following. Change the code below, to try different values of $\lambda$

```r
MM <- as.matrix(cardf[,4:7])

AdStockScale2 <- AdStockScale(0.2,nrow(MM))
GG2 <- AdStockScale2 %*% MM

AdStockScale8 <- AdStockScale(0.8,nrow(MM))
GG8 <- AdStockScale8 %*% MM
```

Estimate the regression model with different $\lambda$ values, and choose the best model using holdout sample test. Which one do you like the best? Why?

```r
#using the top 30 data points as training
train_id <- 1:30
```

```
for (i in 1:9) {
# create the ad-stock variables
lambda <- i/10
GG <- AdStockScale(lambda, nrow(MM)) %*% MM
alldata <- data.frame(cardf$sales,GG)
train_data <- alldata[train_id,]
test_data <- alldata[-train_id,]

# estimate the model using the training data
reg <- lm(cardf.sales~., data = train_data)
# get the prediction using testing data
Ypredict <- reg %>% predict(test_data)

# print results
print(paste("lambda=",lambda, " RMSE = ",
RMSE(Ypredict,test_data$cardf.sales))) }
```

```
## [1] "lambda= 0.1  RMSE =  134.573799225004"
## [1] "lambda= 0.2  RMSE =  134.132964397067"
## [1] "lambda= 0.3  RMSE =  135.381995904566"
## [1] "lambda= 0.4  RMSE =  138.021480161363"
## [1] "lambda= 0.5  RMSE =  141.277920085548"
## [1] "lambda= 0.6  RMSE =  143.695993661438"
## [1] "lambda= 0.7  RMSE =  143.337199323422"
## [1] "lambda= 0.8  RMSE =  139.550444446578"
## [1] "lambda= 0.9  RMSE =  137.926502556566"
```

Based on these results, the testing data has the lowest RMSE, when $\lambda = 0.2$.

Note that in this exercise, we restricted for all the marketing variables to have the same value of $\lambda$, in real practice, this can be relaxed.