model_gross.R

atchirc

Sat May 20 20:23:26 2017

```
library(MASS)
library(car)
library(DataCombine)   # Pair wise correlation
library(stargazer)
library(dplyr)         # Data aggregation
library(glmnet)
source('./code/atchircUtils.R')


data    <- read.csv('./intrim/eleckart.csv')

data_week <- data %>% group_by(week) %>%
  summarise(gmv=sum(gmv),
            product_mrp=mean(product_mrp),
            discount=mean(discount),
            sla=mean(sla),
            procurement_sla=mean(procurement_sla),
            n_saledays=mean(n_saledays),
            TV=mean(TV),
            Digital=mean(Digital),
            Sponsorship=mean(Sponsorship),
            ContentMarketing=mean(ContentMarketing),
            OnlineMarketing=mean(OnlineMarketing),
            Affiliates=mean(Affiliates),SEM=mean(SEM),
            Radio=mean(Radio),
            Other=mean(Other),
            TotalInvestment=mean(TotalInvestment),
            NPS=mean(NPS),
            list_mrp=mean(list_mrp),
            units=sum(units),
            COD=sum(COD),
            Prepaid=sum(Prepaid))

data_week[,c(8:17)] <- data_week[,c(8:17)]*10000000

# Prune, Insignificant variables
# week, sla, procurement_sla, content.marketing, Total Investment,
# units, radio, digital, product_mrp, prepaid, n_saledays

model_data <- data_week[,-c(1,3,5,6,9,11,13,15,17,20,21,22)]
```

*

*PROCs:** Set Class definitions

```
setOldClass('elnet')
setClass(Class = 'atcglmnet',
         representation (
           R2 = 'numeric',
           mdl = 'elnet',
           pred = 'matrix'
         )
)
```

```
setOldClass('lm')
setClass(Class = 'atclm',
         representation (
           R2 = 'numeric',
           mdl = 'lm',
           pred = 'matrix'
         )
)
```

Finding min lambda from 1000 iterations

```
findMinLambda <- function(x,y,alpha,folds) {
  lambda_list <- list()
  for (i in 1:1000) {
    cv.out <- cv.glmnet(as.matrix(x), as.vector(y), alpha=alpha,
                        nfolds=folds)
    lambda_list <- append(lambda_list, cv.out$lambda.min)
  }
  return(min(unlist(lambda_list)))
}
```

Linear Model with Regularization

```
atcLmReg <- function(x,y,l1l2,folds) {
  # l1l2 = 0 for L1,  1 for L2

  if (l1l2) { # Lasso/L2
    min_lambda <- findMinLambda(x,y,1,folds)
  } else { # Ridge/L1
    min_lambda <- findMinLambda(x,y,0,folds)
  }
  mdl        <- glmnet(x,y,alpha=l1l2,lambda = min_lambda)
  pred       <- predict(mdl,s= min_lambda,newx=x)

  # MSE
  mean((pred-y)^2)
  R2 <- 1 - (sum((y-pred )^2)/sum((y-mean(pred))^2))
  return(new('atcglmnet', R2 = R2, mdl=mdl, pred=pred))
}
```

*

*Linear Model:**

```
mdl      <- lm(gmv~., data=model_data)
step_mdl <- stepAIC(mdl,direction = 'both',trace = FALSE)

stargazer(mdl,step_mdl, align = TRUE, type = 'text',
          title='Linear Regression Results', single.row=TRUE)
```

```
##
## Linear Regression Results
## ================================================================================================
##                                              Dependent variable:
##                      ----------------------------------------------------------------------
##                                                      gmv
##                                  (1)                                  (2)
## ------------------------------------------------------------------------------------------------
## discount             3,631,796.000** (1,713,870.000)   3,671,988.000** (1,551,946.000)
## n_saledays            6,711,679.000* (3,559,829.000)   6,459,005.000* (3,383,312.000)
## TV                           0.216 (1.342)
## Sponsorship                  0.248 (0.197)                    0.188* (0.108)
## OnlineMarketing              0.100 (0.653)
## SEM                         -0.305 (0.487)
## Other                       -0.038 (0.367)
## NPS                   -4,430,321.000 (3,180,796.000)  -4,774,056.000*** (1,544,062.000)
## list_mrp                  12,558.840 (7,931.825)           13,617.750* (7,487.708)
## Constant             80,267,093.000 (186,393,888.000) 96,324,620.000 (119,655,858.000)
## ------------------------------------------------------------------------------------------------
## Observations                      53                                   53
## R2                               0.544                                0.538
## Adjusted R2                      0.448                                0.489
## Residual Std. Error     34,936,963.000 (df = 43)         33,619,841.000 (df = 47)
## F Statistic              5.689*** (df = 9; 43)            10.945*** (df = 5; 47)
## ================================================================================================
## Note:                                                      *p<0.1; **p<0.05; ***p<0.01
```

```
knitr::kable(viewModelSummaryVIF(step_mdl))
```

var	Estimate	Std.Error	t-value	Pr(>|t|)	Significance	vif
discount	3.672e+06	1.552e+06	2.366	0.02215	*	1.813414
list_mrp	1.362e+04	7.488e+03	1.819	0.07534	.	1.828340
n_saledays	6.459e+06	3.383e+06	1.909	0.06237	.	1.311196
NPS	-4.774e+06	1.544e+06	-3.092	0.00334	**	1.931501
Sponsorship	1.882e-01	1.082e-01	1.739	0.08850	.	1.773203

```
pred_lm <- predict(step_mdl, model_data)
```

Regularized Linear Model:

```
x = as.matrix(model_data[,-1])
y = as.vector(data_week$gmv)

ridge_out <- atcLmReg(x,y,0,3)  # x, y, alpha, nfolds
lasso_out <- atcLmReg(x,y,1,3)  # x, y, alpha, nfolds

ridge_out@R2
```

```
## [1] 0.5222714
```

```
lasso_out@R2
```

```
## [1] 0.5393413
```

*

*Plot Model prediction and base sales:**

```
plot(model_data$gmv)
lines(model_data$gmv)
lines(pred_lm,col='red',lwd=2)
lines(ridge_out@pred,col='green',lwd=2)
lines(lasso_out@pred,col='blue',lwd=2)
abline(h=step_mdl$coefficients['(Intercept)'],col='red',lwd=2,lty=2)
abline(h=ridge_out@mdl$a0,col='green',lwd=2,lty=2)
abline(h=lasso_out@mdl$a0,col='blue',lwd=2,lty=2)
legend('topright',inset=0, legend=c('GMV','LM','LM+L1','LM+L2'),horiz = TRUE,
       lwd = 2, col=c(1:4), cex = 0.5)
```