**Machine Learning**

Dong Kim

3/5/2018

To frame my main question of capstone project as a machine learning problem, linear regression was used to supervise the main problem. My main question for the machine learning was what prediction can be made with current datasets and what are the relationship between different variables including Date\_1, Date\_2, Outcome\_M*X,* Quantitative variables, and Categorical Variables. I used Date\_1 and Date\_2 which are my advertising campaign variables as my dependent variables. Quantitative and Outcome\_MX were two dependent variables that could be used to predict the sales.

To measure the relationship and for the prediction of the sales, I made training and test set for each linear model using training dataset. There were 117 products that were filtered from 12 months period that had highest outcomes. I made linear models and calculated R-squared to find the relationship between two variables.

When making linear models for Date\_1, Date\_2 , and Outcomes, there were very challenging when reducing P-values of the model. The model1 and model2 had high P-values that tells us that there’s no linear relationship. Outcome values were logged ang log10 was used to lower the p-value. Although, using log10 lowered the P-values but it was not strong enough for the regression. Using log when reducing p-values helped for the fit. Also, ridge regression was used to see the relationship of the variables. Tuning alpha values had no effects on R-square values.

The linear models for quantitative variables were very successful. The models had very strong p-values (lower than 0.05). Also, the R-square values were close to 1. I strongly gained that quantitative variable has strong linear relationship with advertising campaigns.