ROBYN UNDER THE HOOD

With Google pulling the plug on third party cookies in 2023 and data privacy regulations in place, most brands have been looking at other methods to measure their marketing effectiveness.

One such method which can aid marketers in creating effective strategies is Marketing Mix Modeling (MMM).

Market Mix Modeling (MMM) is a technique which helps in quantifying the impact of several marketing inputs on the KPI (e.g Sales, Market Share, CTR etc). The purpose of using MMM is to understand how much each marketing input contributes to the KPI, and how much to spend on each marketing input.

We have been requested by many to give our take on open-source libraries like Robyn.

We have been using our proprietary techniques + Robyn for some of our MMM projects.

We are starting a series “Robyn under the hood” to educate and inform MMM users across the globe on Robyn.

Through this series, we dig deep into the methods/ techniques used and try to answer the ‘How’ & ‘Why’ behind these techniques like:

- Why Robyn uses Ridge Regularization?
- Does Ridge reduce overfitting?
- What loss function is used?
- Why is Weibull distribution used for Adstock? Why not any other distribution?

- How Multi-objective optimization works for tuning hyper-parameters?
- What is Pareto optimality?

To kick-off, we will first cover Ridge Regularization in Robyn.

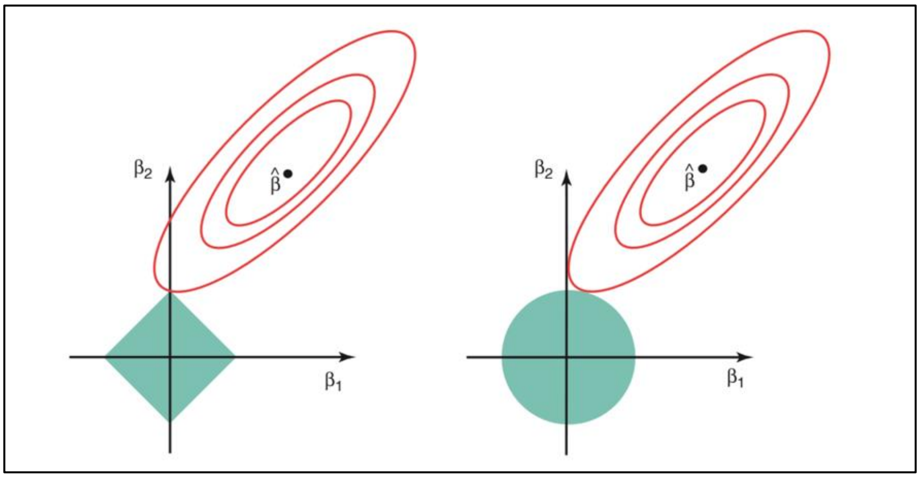
Robyn Under the Hood

"Robyn is an experimental, semi-automated and open-sourced Marketing Mix Modeling (MMM) package from Meta Marketing Science. It uses various machine learning techniques (Ridge regression, multi-objective evolutionary algorithm for hyperparameter optimization, time-series decomposition for trend & season, gradient-based optimization for budget allocation etc.) to define media channel efficiency and effectivity, explore adstock rates and saturation curves."[1]

Under the hood, Robyn uses Ridge regression to regularize multi-collinearity and automate hyper-parameter optimization using evolutionary algorithms from Facebook AI's library Nevergrad. It also makes use of Facebook's prophet library to decompose the time series into trend, seasonality and holidays.

So, first let's do a drill down on Ridge Regression.

1. **Ridge Regression:**



Source: An Introduction to Statistical Learning

Ridge Regression (depicted on the right hand size in the above image) is a regularization technique which helps combat overfitting in the model. For more detailed explanation on why Ridge Regression does not reduce the coefficients to zero, check out the resources here and here.

There are some models which perform well on the training set but fail to perform on the test set. In other words, the model has poor predictive power when it comes to unseen data and is overfitted on the train set. To help combat this, regularization technique is applied which reduces the variance at the cost of adding a bias.

One such method is Ridge Regression.

2. Why is Ridge regression used in MMM?

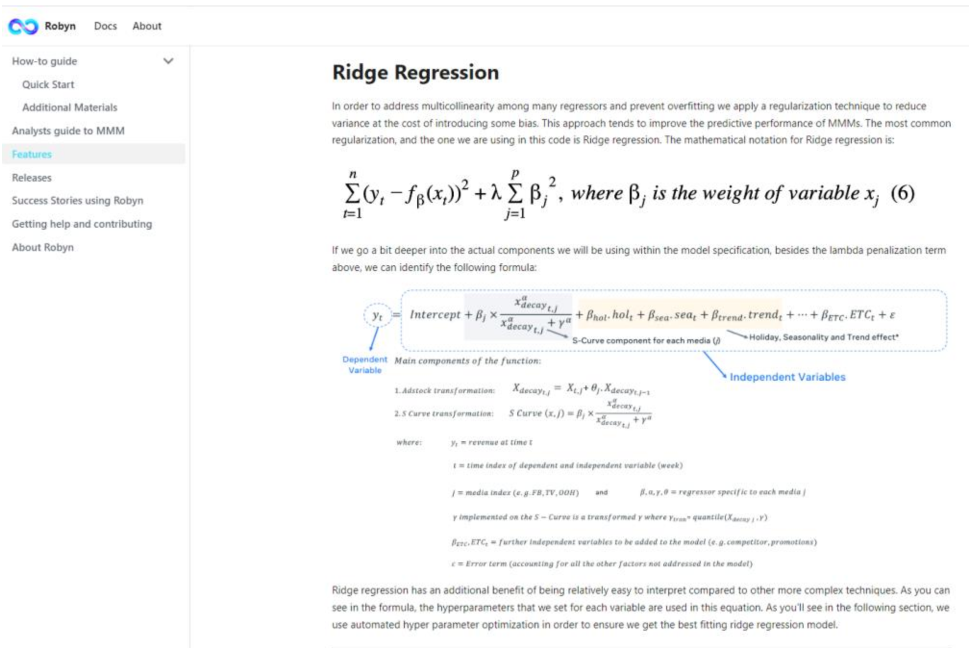
It is very common to observe multicollinearity among different independent variables in MMM. This leads to model being overfitted. Through Ridge regression, a penalty term is introduced to the model cost function. The penalty term is the sum of squares of all the model coefficients multiplied by Lambda (λ). Refer to the equation below:

$$\sum_{t=1}^{n}(y_t - f_\beta(x_t))^2 + \underbrace{\boxed{\lambda \sum_{j=1}^{p} \beta_j^{\,2}}}_{\text{Penalty Term}}, \; \textit{where } \beta_j \textit{ is the weight of variable } x_j \quad (6)$$

The penalty term penalises the regression model and shrinks the coefficients towards zero. Here, lambda controls the severity of the penalty. Higher the value of λ, higher is the impact of shrinkage.

So, if there are too many variables in the model, the impact of some variables will be closer to zero. Ridge Regression does not remove predictors from the model, it just shrinks them towards zero (but does not make it zero). The coefficients are not exactly 0 and we can still see the impact of all the predictors in the model.

3. Ridge Regression in Robyn



Ridge Regression

In order to address multicollinearity among many regressors and prevent overfitting we apply a regularization technique to reduce variance at the cost of introducing some bias. This approach tends to improve the predictive performance of MMMs. The most common regularization, and the one we are using in this code is Ridge regression. The mathematical notation for Ridge regression is:

$$\sum_{t=1}^{n}(y_t - f_\beta(x_t))^2 + \lambda \sum_{j=1}^{p} \beta_j^{\,2}, \; \textit{where } \beta_j \textit{ is the weight of variable } x_j \quad (6)$$

If we go a bit deeper into the actual components we will be using within the model specification, besides the lambda penalization term above, we can identify the following formula:

$$\underbrace{y_t}_{\text{Dependent Variable}} = \underbrace{Intercept + \underbrace{\beta_j \times \frac{x^{\alpha}_{decay_{t,j}}}{x^{\alpha}_{decay_{t,j}} + \gamma^{\alpha}}}_{\text{S-Curve component for each media }(j)} + \underbrace{\beta_{hol}.hol_t + \beta_{sea}.sea_t + \beta_{trend}.trend_t}_{\text{Holiday, Seasonality and Trend effect*}} + \cdots + \beta_{ETC}.ETC_t + \varepsilon}_{\text{Independent Variables}}$$

Main components of the function:

1. Adstock transformation: $X_{decay_{t,j}} = X_{t,j} + \theta_j . X_{decay_{t,j-1}}$

2. S Curve transformation: $S\ Curve\ (x, j) = \beta_j \times \frac{x^{\alpha}_{decay_{t,j}}}{x^{\alpha}_{decay_{t,j}} + \gamma^{\alpha}}$

where: y_t = revenue at time t

t = time index of dependent and independent variable (week)

j = media index (e.g. FB, TV, OOH) and $\beta, \alpha, \gamma, \theta$ = regressor specific to each media j

γ implemented on the S − Curve is a transformed γ where γ_{tran} = quantile($X_{decay\ j}$, γ)

$\beta_{ETC}.ETC_t$ = further independent variables to be added to the model (e.g. competitor, promotions)

ε = Error term (accounting for all the other factors not addressed in the model)

Ridge regression has an additional benefit of being relatively easy to interpret compared to other more complex techniques. As you can see in the formula, the hyperparameters that we set for each variable are used in this equation. As you'll see in the following section, we use automated hyper parameter optimization in order to ensure we get the best fitting ridge regression model.

In the next post, we will try to cover the functions responsible for implementing Ridge regression in Robyn.

Resources:

1. https://facebookexperimental.github.io/Robyn/docs/about
2. https://facebookexperimental.github.io/Robyn/docs/analysts-guide-to-MMM