Enhancing Media Campaign Strategy through Optimizing Cost and Maximizing Revenue

Ben Goldberg

Northwestern University

CIS435 Practical Data Science Using Machine Learning

Assignment 1: Supervised Learning with Regression

January 23, 2022

**Abstract**

Although the company is unknown, review and optimization of a marketing strategy spans across industries. Spending money to bring in potential customers in hopes that they turn into sales dollars is a strategy implemented by almost all companies in the world today. In this example, I are looking at a sample marketing strategy where 6 media channels are used to generate sales. The 6 media channels are: TV, Radio, Newspaper, Google, Facebook, and Linkedin. The goal of this assignment is to reduce the amount media channels that are not influencing sales as much as other media channels. I will use regression algorithms in machine learning (utilizing Python) to understand the performance of each media channel and its ultimate impact on sales.

**Marketing strategy based on Machine Learning**

The first step in understanding which channels to remove is to match the goal with an appropriate machine learning strategy. Over the last decade, businesses have begun adopting data driven strategies to maximize profits and drive down costs. The usage of machine learning strategies has enabled businesses to harness the power of complex statistical processes with increased speed and accuracy. Machine learning can be used in a broad spectrum of analysis and ensuring I pick the correct one is paramount. When looking at marketing tasks, I are trying to understand influence on sales today and predict what could impact sales in the future. A simple machine learning model that can fit this situation is linear regression. I will use linear regression to pick highly performant media channels to optimize our marketing strategy.

**Machine Learning Algorithms**

Now that I have ventured into the world of linear regression, our next step is to understand possible algorithms that best fit our goal. In this assignment, I will use OSL, Ridge, and Lasso algorithms and compare their outcomes. Using the algorithm’s fit to our data will determine which I should use as a basis for our analysis.

**Data Preprocessing**

However, before I can apply the 3 algorithms listed above, I first need to ensure our data is accurate and complete. The dataset provided contains 200 markets with its corresponding cost of marketing and sales in thousands of dollars. The below table demonstrates the structure of our data.Table

Description automatically generated

**Missing Data/Data Quality Review**

When reviewing the data, I determined there were no blank values for any of the media channels or sales. However, I found there were negative numbers listed only within Facebook’s marketing channel. Without further knowledge of rebates/cost of Facebook’s marketing, a representative from those markets would need to confirm if those values are correct. For this assignment, I will assume the negative amounts are accurate and complete. The last check conducted is to ensure no market is represented multiple times. Without the name of the market or its segmentation, I used the combination of values from the media channels and sales to determine its uniqueness. As a result of that analysis, no rows of data were duplicated.

**Statistical Dataset Review**

Using a package in Python called Sweetviz, I can grab quick statistical figures regarding our data. In the attached file (found in the appendix), I can see the relationship of each media channel, and which is ultimately impacting sales. When looking at the Sweetviz, it is suggesting that Google has the highest correlation to sales while newspaper and Linkedin are the two lowest media channels. Additionally, I can see that both newspaper and Linkedin costs are skewed to the right. This observation suggests that when presented with both cheap and expensive marketing options, both newspaper and Linkedin used a cheaper option.

**Applying Machine Learning**

OLS, Ridge, and Lasso all use linear regression to best predict future outcomes. With any linear regression algorithm, all of them follow the same simple formula, y = mx + b. To adapt the function to our use case of 6 media segments, I need to expand our function to accept 6 different variables. The following function helps us accomplish this:

Y = x0 + wx1 + wx2 + wx3 + wx4 + wx5 + wx6

In the above function, x0 is our y-intercept and each media segment is represented by x*n*. Each *n* value from 1 to 6 is a unique x value for its corresponding media segment. Lastly, the w value is the coefficient. Depending on the coefficient, it determines the amount of impact our x*n* values have on the overall y value.

In an Ordinary Least Squares (OLS) approach, it uses the sum of least squares to minimize the cost between known data points and predicted data points. Put simply, the algorithm is trying to place a line between the data points that minimizes the distance between each point and the line itself. For Ridge regression, it is a step further than an OLS approach. In Ridge regression, I use an alpha character to control the amount of shrinkage. (SKLEARN, 2022) In the below image, you can see the impact of alpha on the least squares function.Chart

Description automatically generated

In our last linear regression algorithm, it’s a step in a different direction than Ridge. With Lasso regression, the goal is to emphasize coefficients that highly impact the validity of the model while pushing the others to 0. In the following table, you can see how Lasso regression pushed media channels to 0 due to their minimal impact.

Table

Description automatically generated

To simplify our analysis, I used a default alpha value of 1 in both Ridge and Lasso algorithms.

**Interpreting Results**

Now that I understand our data and the machine learning algorithms I will be applying to the dataset, I need to review our results. When comparing machine learning algorithms, it is important to review error scores. I calculated 4 different error score calculations but will rely on 2 for our basis of determining which algorithm to use. The first score I use is the R2 error score. R2 error score is a value between 1.0 and lower where a 1.0 score represents a perfect fitting model. As the score decreases, so does the fit of the model. For the second score I used mean squared error (MSE). The MSE is a value that is 0.0 or greater where a 0.0 score represents a perfect fitting model. As the score increases, the fit of the model decreases. Based on the table below, I can see that the Lasso regression had both the highest R2 score and lowest MSE. Having both the highest R2 and lowest MSE score encouraged me to use Lasso regression for the assignment, but I believe Ridge regression will give us the best results. Due to Lasso pushing values to 0 if they had minimal impact, I believe it could cause us to generalize too much when comparing media channels. If 4 channels were pushed to 0, it does not mean they all performed the same, but rather the model didn’t use it because it wasn’t contributing as much as another media channel. As a result, I will be using Ridge regression for the remainder of the analysis. In the table below, we can see the error scores listed.

**Table

Description automatically generated**

Table

Description automatically generatedWith our algorithm selected, we next need to look at the coefficients of each media channel. In the below graphics, I gathered the coefficients from all algorithms, but will focus on Ridge results.

Chart

Description automatically generated

A coefficient value in the above table represents the amount of sales created per x amount of marketing cost. For example, if I put $1 into radio advertising, the algorithm predicts $87.80 in sales because of that dollar. In the Ridge coefficients, I can clearly see that Radio and Google media channels are the main drivers of sales. Comparing to Ridge to Lasso and OLS, I can see that both Radio and Google are still the top 2 media channels.

The next point to take away from our Ridge regression is that Facebook and Linkedin may not be driving as many sales compared to Radio and Google, they are at least driving positive sales. Looking at TV and Newspaper, I see a negative sales dollar created per dollar spent. With this analysis at hand, I believe I have enough to recommend next steps for optimizing our media strategy.

**Next Recommended Steps**

Up to this point, I understand the accurate and completeness of our data, the machine learning algorithms used to help predict future values, and now interpretation of those results. At the end of our analysis, I determined that Radio and Google were two top performing media channels while TV and Newspaper were driving negative sales. For each dollar spent in Newspaper and TV combined, I am predicting a loss of sales of almost $67. To maximize our sales and optimize our marketing spend, I recommend removing both Newspaper and TV media channels from the overall marketing strategy. With removal of those two media channels, I am predicting a $67 increase in sales along with a 33% decrease in marketing cost. To calculate the decrease in marketing cost, I am reducing the 2 of 6 total marketing channels for a result of 33%.

**References**

G. L. T. (2022, January 12). *What is ridge regression?* GreatLearning Blog: Free Resources what Matters to shape your Career! Retrieved January 23, 2022, from https://www.mygreatlearning.com/blog/what-is-ridge-regression/

*1.1. Linear Models*. scikit. (n.d.). Retrieved January 23, 2022, from https://scikit-learn.org/stable/modules/linear\_model.html

Alto, V. (2019, August 17). *Understanding the OLS method for simple linear regression*. Medium. Retrieved January 23, 2022, from https://towardsdatascience.com/understanding-the-ols-method-for-simple-linear-regression-e0a4e8f692cc

*SKLEARN.METRICS.MEAN\_SQUARED\_ERROR*. scikit. (n.d.). Retrieved January 23, 2022, from https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean\_squared\_error.html?highlight=mean+squared+error#sklearn.metrics.mean\_squared\_error

*Sklearn.metrics.r2\_score*. scikit. (n.d.). Retrieved January 23, 2022, from https://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2\_score.html

Wikimedia Foundation. (2022, January 10). *Mean squared error*. Wikipedia. Retrieved January 23, 2022, from https://en.wikipedia.org/wiki/Mean\_squared\_error

**Appendix**

Sweetviz HTML file (attached in canvas)