Retail Sales Forecasting

Steven Miller

Abstract

This paper will look at the challenge of forecasting retail sales on a weekly basis. Sales are broken down by departments within a store, and numerous factors internally and externally are examined. External factors include economic indicators, such as the price of fuel, consumer price index, and unemployment rate in the area. Internal factors include the department, the store, and the size of the store the department is located in. Also included are the presence or absence of promotional events intended to drive sales.

Based on the data available, several research questions are presented. Does a relationship exist between CPI and weekly sales? Does a relationship exist between the unemployment rate and weekly sales? Does a relationship exist between fuel prices and weekly sales? How much of an impact do promotional events have on weekly sales?

Because most of the potential drivers of sales are external factors, the ability for a company to leverage these factors to drive sales is severely limited. Additionally, if promotional events are found to be significantly correlated with an increase in sales, further internal studies would be necessary to in order to determine the actual profitability of markdowns as an increase in revenue comes at the expense of profit margins. The largest benefit from this study is likely to be an ability to better forecast future revenues and plan operations accordingly, rather than the ability to identify drivers of revenue to be leveraged.

The Data

A dataset of historical sales for 45 stores was acquired from Kaggle. Three files were included, one containing data on each store, one containing data on the weekly sales, and a dataset of features for each store and week.

The features dataset includes 8,190 rows and 12 features:

1. Store – an integer representing the store that the other features correspond to.
2. Date – the date representing the week the other features correspond to. With store, these combine to make a unique index.
3. Temperature – the average temperature for the week in degrees Fahrenheit in the region the store is located
4. Fuel Price – the price of fuel for the week in dollars per gallon in the region the store is located
5. MarkDown1 – anonymized data related to the presence, or absence, of promotional events at the store for the week
6. MarkDown2 – anonymized data related to the presence, or absence, of promotional events at the store for the week
7. MarkDown3 – anonymized data related to the presence, or absence, of promotional events at the store for the week
8. MarkDown4 – anonymized data related to the presence, or absence, of promotional events at the store for the week
9. MarkDown5 – anonymized data related to the presence, or absence, of promotional events at the store for the week
10. CPI – the consumer price index for the week for the region the store is located
11. Unemployment Rate – the unemployment rate for the week for the region the store is located
12. IsHoliday – a Boolean value representing whether a holiday was occurring during the week. For this data, four weeks are considered holiday weeks. These are the weeks of the Super Bowl, Labor Day, Thanksgiving, and Christmas.

The stores dataset includes 45 rows and 3 features

1. Store – an integer representing the store the other features correspond to.
2. Type – a categorical variable denoting A, B, or C indicating the type of store. Because this data has been anonymized, what the store type means is not immediately apparent.
3. Size – the size of the store in square feet.

While the store types are not defined anywhere, an analysis shows that stores of type A are typically the largest, with a square footage on average of 177,248 square feet. Stores of type B are in between, with an average of 101,191 square feet. Stores of type C are the smallest, averaging 40,541 square feet. While it's impossible to determine from the data what stores this data was derived from, or what the type of store indicates it's interesting to note that the square footage of a standard Target store ranges from 80,000 to 125,000 square feet, a Super Target is approximately 174,000 square feet, and an urban, small-format Target store is approximately 50,000 square feet. That's not to say that the data for these stores is necessarily Target data, but it does likely represent a company of similar size and variety of stores.

The sales dataset includes 421,570 rows and 5 features.

1. Store – an integer representing the store the other features correspond to.
2. Department – an integer representing the department the other features correspond to.
3. Date – the date representing the week the other features correspond to. With Store and Department, these combine to make a unique index.
4. Weekly_Sales – a floating point value representing the sales in dollars for the department for the week
5. IsHoliday - a Boolean value representing whether a holiday was occurring during the week. For this data, four weeks are considered holiday weeks. These are the weeks of the Super Bowl, Labor Day, Thanksgiving, and Christmas.

Some initial data cleaning was the next step. Many MarkDown values in the features dataset were listed as null. I considered this to be evidence that no promotional event had taken place, and filled all null values for these fields with zeroes.

Several weekly_sales values were less than, or equal to zero. This may have been an error, or it could indicate a slow week where returns were greater than the actual sales. Either way, this composed less than one-third of one percent of the overall data and I was comfortable with discarding these entries from any further analysis.

With the goal being to predict the weekly sales value based on the remaining features, weekly sales is a good place to begin exploring. A histogram shows that the distribution is heavily right-skewed, with a few large values and many smaller ones.

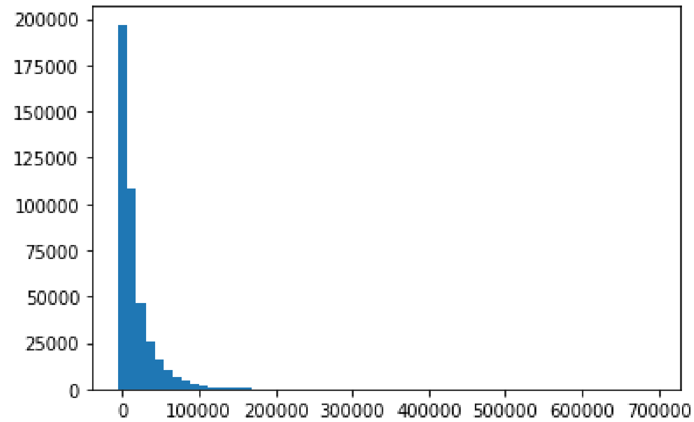

Figure 1: Distribution of Weekly Sales across Departments

Performing a log-transformation to the weekly sales resulted in a distribution that retains some skew but is much closer to a normal distribution than previously. This will be beneficial when attempting to predict weekly sales.

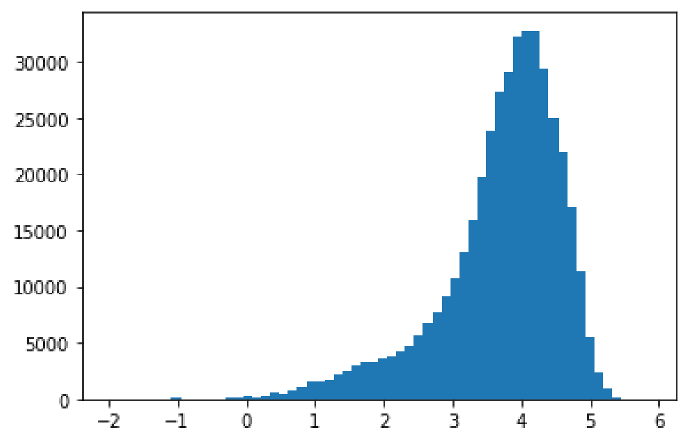

Figure 2: Distribution of log-transformed weekly sales across departments

Moving on to the explanatory variables, one early concern was that the economic indicators, CPI, unemployment, and fuel prices, would be too closely correlated to be useful as independent explanatory variables. After performing a correlation analysis, these concerns were unfounded. A correlation of -0.19 was found between fuel prices and CPI. A correlation of -0.03 between fuel prices and unemployment, and a correlation of -.30 between unemployment and CPI. Overall, these were all weak enough to not require any feature reduction at this stage.

Initially, many of the correlations observed between the explanatory variables and weekly sales were quite weak, but present. By performing a permutation test, it was possible to determine that the small correlations that were present were not by chance, as all tests had a p-value below 0.05.

Modeling

While several models were tested, it was important to start with a simple method, an OLS regression in this case. This allowed for greater interoperability and understanding of the results of the regression that may not be as easily accessible in more complex regression methods.

The first model, an OLS regression used the log-normalized weekly sales values as the output variable, and CPI, Temperature, Fuel Price, Unemployment, the 5 MarkDown features, and whether or not it was a holiday as explanatory variables. The result was less-than-stellar, with an adjusted r^2 value of 0.007. All variables involved were found to be statistically significant, except whether it was a holiday.

The next step was to continue with the same model but add some additional features. This was accomplished by joining the store data to the rest of the dataset, adding the type of store and the square footage as features to be used by the model. An additional permutation test showed that correlations for these new features were all statistically significant. The result of adding these features to the model was an improvement to an adjusted r^2 of 0.112, and a mean squared error of 0.703. This was better, but it still indicates that the model only explains about 11% of the variance in weekly sales and this is of very little use in a production environment.

Next, additional models were tested, starting with a Random Forest Regressor, which performed significantly worse, with an adjusted r^2 of 0.01. From there, a boosted tree algorithm was used via XGBoost. The results were better than the Random Forest Regressor, but still lagged behind the initial OLS regression, even after testing various hyperparameters.

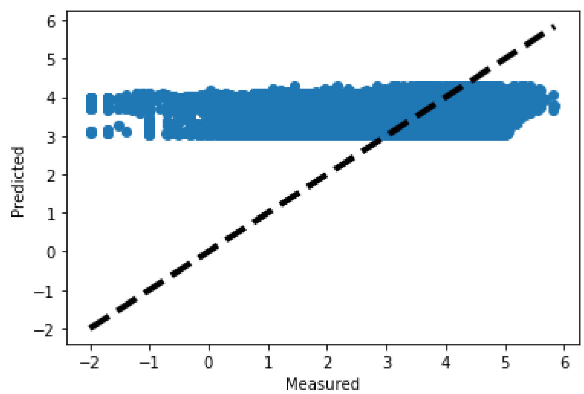

Figure 3: Plot of predicted values compared to actual values of OLS Regression.

As evidence in Figure 3 above, it appears that the algorithm at this point is almost exclusively predicting values between 3 and 4, or $100 to $1,000 in weekly sales. The variance being captured by the features at this point do not appear to contain enough information to explain weekly sales that fall outside of this range.

To improve results, additional features were added at this point. The categorical variables of store numbers and department numbers were included as dummy variables in the model. This took the number of features in use from 8 to 134, which did increase the complexity of the model. The result was an initial OLS regression with an r^2 value of 0.679. Obviously, this is a significant improvement over earlier results of 0.112, and the model now captures the variance in performance on a per-store basis.

While satisfied with the improved results of the OLS regression after adding new features, it was still worth exploring other models to see how the results compared. A Random Forest regressor with 40 estimators resulted in an adjusted r^2 value of 0.77, about 9% higher than the OLS regression.

Using default parameters with XGBoost, a model was trained in 100 rounds with a RMSE of 0.504. Further testing using a grid search two parameters at a time reduced the RMSE down to just 0.278. This final model had an adjusted r^2 value of 0.735. This figure was slightly below the random forest value of

0.77. Next, a Python library called Hyperopt was used to tune parameters in a more intelligent manner than a random grid search. Using this method, I was able to reduce RMSE even further, down to 0.176. This resulted in an adjusted r^2 of 0.795, an improvement over the base Random Forest regression.

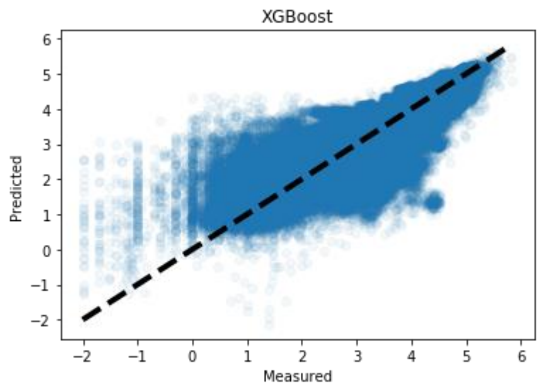

Figure 4: Predicted and Actual Values Using XGBoost When All Features Were Included

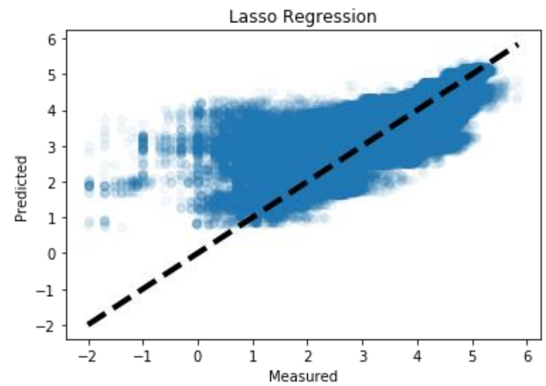

Figure 5: Predicted and Actual Values Using OLS Regression When All Features Were Included

This leads to a question of whether the extra time needed to train a more complex, parameter-tuned model is worth the increase in accuracy. In this case, I believe it is. As new data is only being recorded on a weekly basis, retraining of the model would not need to be a frequent occurrence. Once the model is saved, much of the heavy lifting has been done. The model can be saved and used to make predictions relatively quickly.

Conclusion

While there still seems to be some struggles with predicting some of the outlier values, those ranging below weekly sales of $100, I'm relatively satisfied with the accuracy of this model. The accuracy gained over OLS by moving to XGBoost was worth the sacrifice in simplicity. Any forecasting of this nature would be relatively lightweight in nature, even for a large corporation with thousands of stores. It may take some time to train the model but creating predictions from the XGBoost model on a weekly basis would be a quick process, and I don't believe the complexity of the model would be an issue.

Finally, I want to look at the results of the OLS regression by feature and circle this back to the original research questions.

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.8072	0.053	34.278	0	1.704	1.911
IsHoliday_sales	0.0108	0.003	3.285	0.001	0.004	0.017
Temperature	-0.0006	5.24E-05	-11.294	0	-0.001	0
Fuel_Price	-0.0228	0.003	-8.009	0	-0.028	-0.017
MarkDown1	3.47E-07	2.70E-07	1.282	0.2	-1.83E-07	8.77E-07
MarkDown2	-8.05E-07	1.64E-07	-4.901	0	-1.13E-06	-4.83E-07
MarkDown3	1.90E-06	1.47E-07	12.92	0	1.61E-06	2.19E-06
MarkDown4	-8.02E-07	3.78E-07	-2.122	0.034	-1.54E-06	-6.13E-08
MarkDown5	1.40E-06	2.25E-07	6.212	0	9.57E-07	1.84E-06
CPI	-0.001	0	-2.347	0.019	-0.002	0
Unemployment	0.0059	0.002	3.458	0.001	0.003	0.009
Size	8.95E-06	4.58E-08	195.518	0	8.86E-06	9.04E-06

Figure 6: A Table of some of the features included in OLS regression. Categorical features for store and department IDs omitted for size.

While omitted from this table, the store and department identifications had the largest coefficients, indicating the largest impact on the forecasting of sales. My original questions had focused more on the external factors.

- Does a relationship exist between CPI and weekly sales?
    - A statistically significant relationship exists between CPI and sales however it is rather weak with a 1-point increase in CPI resulting in a 0.001 decrease in the log-transformed value of weekly sales.
- Does a relationship exist between the unemployment rate and weekly sales?
    - A statistically significant relationship exists between unemployment and sales, though it is small and unexpected. A 1% increase in the rate of unemployment corresponds with a 0.0059 increase in the log-transformed value of weekly sales.
- Does a relationship exist between fuel prices and weekly sales?
    - A statistically significant relationship exists between fuel prices and weekly sales. A $1.00 increase in the price of fuel corresponds with a 0.0228 decrease in the log-transformed value of weekly sales.
- How much of an impact do promotional events have on weekly sales?
    - The impact of promotional events is statistically significant, but also so minor that it’s difficult to even measure based on the MarkDown features. One possible explanation is the impact that the presence of a holiday week has on sales, which is an increase of 0.0108 in the log-transformed value. As many promotional events would occur around holidays, it is possible that the holiday feature is covering up some of the impact that promotional events are having.

The one drawback to this project, I would say, is that the dataset has been anonymized. As a result, the features regarding promotions are essentially just numbers that have been devoid of any meaning. They work fine for training a machine learning model, but there’s no way to derive actionable information from these columns without knowing what they mean. Additionally, the fact that the model relies most heavily upon store and department identifications to forecast sales means that it would take some time before it would be able to accurately forecast sales for a new store opening, as the necessary data wouldn’t be available for training.

References

Bergstra, J., Yamins, D., Cox, D. D. (2013) Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. To appear in Proc. of the 30th International Conference on Machine Learning (ICML 2013).

Chu, C.-W., & Zhang, G. P. (2003). A comparative study of linear and nonlinear models for aggregate retail sales forecasting. International Journal of Production Economics, 86(3), 217–231. https://doi.org/10.1016/s0925-5273(03)00068-9

Curtis, A. B., Lundholm, R. J., & Mcvay, S. E. (2014). Forecasting Sales: A Model and Some Evidence from the Retail Industry. Contemporary Accounting Research, 31(2), 581–608. https://doi.org/10.1111/1911-3846.12040

Duff, K. (2012). The Advantages of Sales Forecasting. Retrieved December 4, 2019, from Azcentral.com website: https://yourbusiness.azcentral.com/advantages-sales-forecasting-10308.html

Furseth, P., & Stumbaugh, J. (2017). The Importance of Sales Forecasting Why is Sales Forecasting Important? Retrieved from https://www.orm-tech.com/app/uploads/2017/08/ORM-Technologies-The-Importance-of-Sales-Forecasting.pdf

Geurts, M. D., & Patrick Kelly, J. (1986). Forecasting retail sales using alternative models. International Journal of Forecasting, 2(3), 261–272. https://doi.org/10.1016/0169-2070(86)90046-4

Guo, Z. X., Wong, W. K., & Li, M. (2013). A multivariate intelligent decision-making model for retail sales forecasting. Decision Support Systems, 55(1), 247–255. https://doi.org/10.1016/j.dss.2013.01.026

Harris, E. (2014, January 16). Target Tests Small Store for Urban Shoppers as Young People Pick Cities Over Suburbs. The New York Times. Retrieved from https://www.nytimes.com/2014/01/17/business/target-is-testing-a-small-store-for-city-shoppers.html

hyperopt. (2019, December 18). hyperopt/hyperopt. Retrieved December 23, 2019, from GitHub website: https://github.com/hyperopt/hyperopt

Lam, S., Vandenbosch, M., & Pearce, M. (1998). Retail sales force scheduling based on store traffic forecasting. Journal of Retailing, 74(1), 61–88. https://doi.org/10.1016/s0022-4359(99)80088-8

Martin, P. (2019). Retail store sales forecasting | Machine learning blog. Retrieved December 4, 2019, from Neuraldesigner.com website: https://www.neuraldesigner.com/blog/retail-store-sales-forecasting

Ni, Y., & Fan, F. (2011). A two-stage dynamic sales forecasting model for the fashion retail. Expert Systems with Applications, 38(3), 1529–1536. https://doi.org/10.1016/j.eswa.2010.07.065

Spotio. (2019, May 8). Sales Forecasting Technique: Regression Analysis. Retrieved December 4, 2019, from Spotio website: https://spotio.com/blog/regression-analysis/

Tuttle, B. (2016, October 5). Why Target Is Focusing on Millennials and Cities Instead of Suburban Moms. Retrieved December 15, 2019, from Money website: https://money.com/target-small-stores-millennials-cities-college-towns/