Time Series Sales Forecasting

James Pao, Danielle Sullivan
Stanford University

Motivation

Today, as markets are global, optimizing an organization's operational efficiency is of premium importance. If a company can match the demand of a product with just the right amount of supply, then there will be no lost sales due to a lack of inventory as well as no costs from overstocking. Sales forecasting uses patterns gleaned from historical data to predict future sales, allowing for informed courses-of-action such as allocating or diverting existing inventory, or increasing or decreasing future production.

Problem Statement

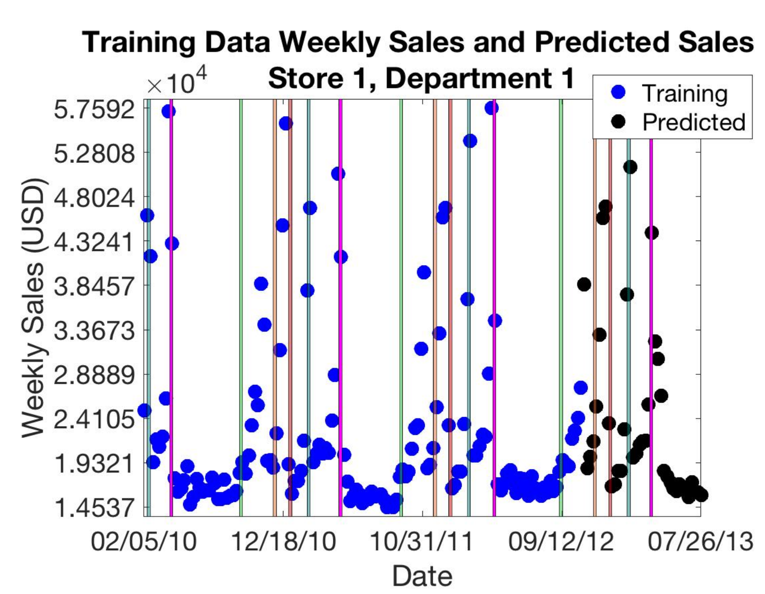
Our project is a prediction problem from a 2014 Kaggle challenge. The dataset contains at most 143 weeks of historical departmental weekly sales data from 45 Walmart stores. The training set has 421,570 samples. Each sample has the departmental weekly sales along with associated store number, store type, date, and a flag indicating if the week contains a major holiday. The test set has 115,064 entries with the same features. Our goal is to predict the weekly sales within a department of a store. We score our models using Kaggle's submission platform online, which scores based on weighted mean absolute error. We also test our models locally using hold-out sets generated from the training samples.

Results

The STL + ARIMA algorithm was implemented in R and MATLAB. It generated 3645 models (a model per department (81 depts.) – store (45 stores) pairing) to make 115,064 predictions (~39, predictions per store dept. pairing). These predictions relied on historical sales data, department id, and store id. Submitting to Kaggle's online platform, we achieved a score within 500 points of the winning submission.

Algorithm	Kaggle Score
STL + ARIMA	2875.6
Decision Tree	4384.4

While our STL + ARIMA algorithm performed well on most holiday data, it did not perform as well as hoped on holidays that do not occur on the same week each year (moving holidays). This was especially noticeable on Easter (magenta line). Tweaking our algorithm to accommodate for moving retail holidays like Easter could improve our results. A decision tree was used as a baseline. Our NANN achieved a mean absolute error of 9582.6 on a hold-out set. Further work is needed, as detailed in the "Next Steps" section.



References

[1] Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees.* Boca Raton, FL: Chapman & Hall, 1984.

[2] Shumway, Robert H., and David S. Stoffer. *Time Series Analysis and Its Applications: With R Examples*. Springer New York, 2011.

[3] Chatfield, Christopher. *Time-Series Forecasting*. Chapman&Hall/CRC, 2001. Chatfield, Chris

[4] Dalinina, Ruslana. "Introduction to Forecasting with ARIMA in R." *DataScience.com*, www.datascience.com/blog/introduction-to-forecasting-with-arima-in-r-learn-data-science-tutorials.

[5] Zhang, G.P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing,* 50, 159-175.

[6] STL: A Seasonal-Trend Decomposition Procedure Based on Loess *Journal of Official Statistics*, Vol. 6, No. 1. (1990), pp. 3-73 by Robert B. Cleveland, William S. Cleveland, Jean E. Mcrae, Irma Terpenning

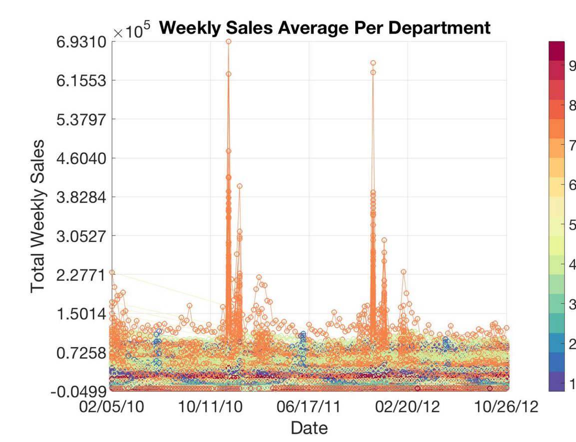
[7] P.W.D. Thahler David, Walmart_competition_code, (2014), GitHub repository, https://github.com/davidthaler/Walmart_competition_code

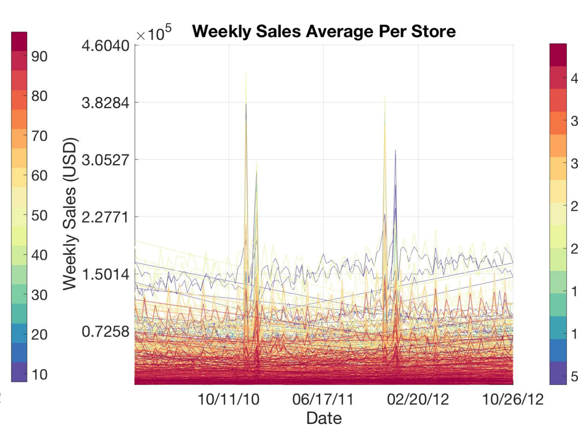
Acknowledgments

Many thanks to Professors Boneh and Ng, all of the TA's ,and our fellow classmates who were so helpful on Piazza

Implementations

We implemented several models, namely decision trees, NANN, and STL + ARIMA, in attempts at finding the best one. There is a strong relationship between weekly sales and the features week number, store id, department number and the holiday flag. Thus our models focus on leveraging these features.

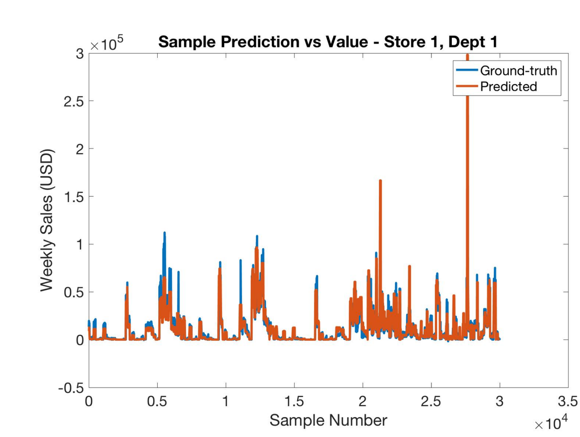




Decision Tree

Select week number, store, department, holiday flag, and the store size as attributes to predict weekly sales

Construct Decision Tree via CART Algorithm



Next Steps

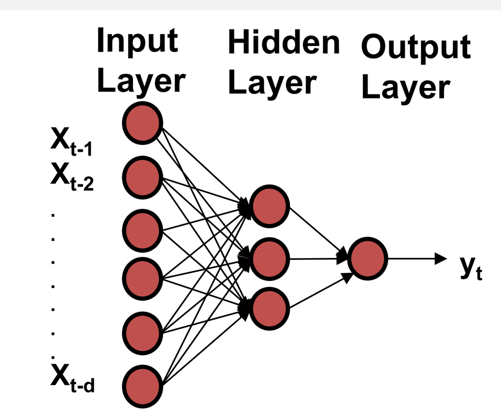
We will continue to improve our neural network for direct time series prediction as well investigating a neural network to model the residuals from our ARIMA model for a hybrid regressor. We will also investigate combining our predictors into an ensemble model.

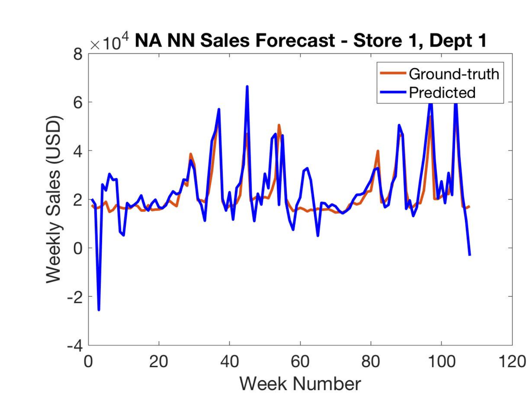
Nonlinear Autoregressive Neural Network

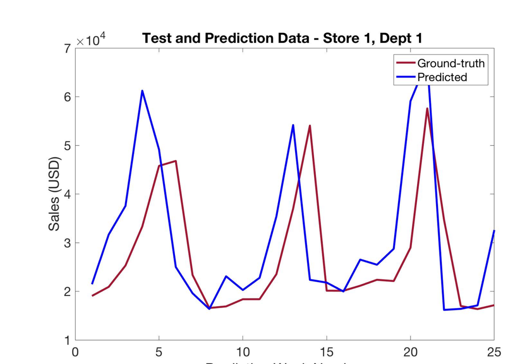
Separate Data by Department and Store Id Pairings

Train the neural network using the last *d* values of x as the features

Generate a Model per Department - Store Pair





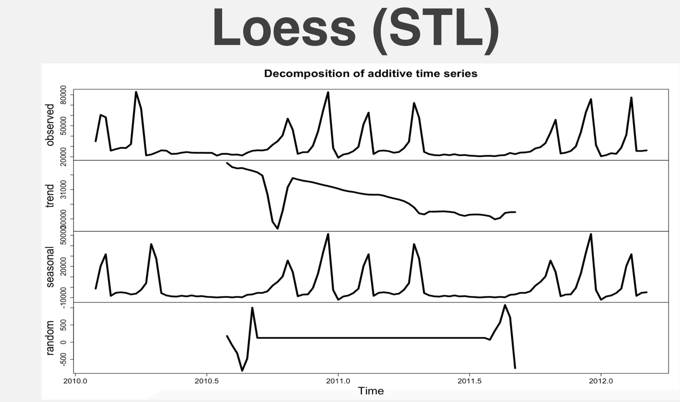


Our NANN uses the last 10 weekly sales values as inputs to the neural network to predict the sales at time *t*. Due to the dependence on past sales trends, this model was unable to pickup on sudden jumps in sales.

STL + ARIMA

Separate Data by Department and Store Id Pairings

Decompose Additive Time Series Data into Seasonal, Trend, and Random using Seasonal Trend Decomposition Using Loess (STL)



ARIMA Modeling for Non-seasonal Component Forecasting

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

Add Seasonal and Non-Seasonal Component Predictions

$$y_t = T_t + S_t + N_t$$

Generate a Model per Department, Store Pair