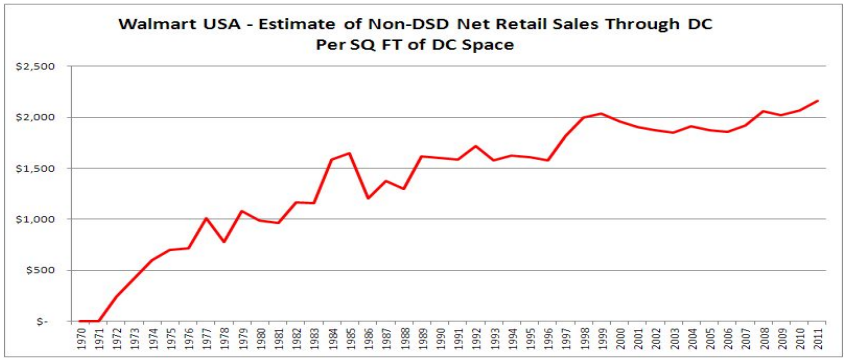
Walmart Sales Prediction



Programming for Data Science Final Project Proposal

Team 10: Nan Su, Yichen Li, Ruichan Ma, Tianyu Wang

Problem Statement

- Data

Sales data from 2010-02-05 to 2012-11-01 with the following fields:

1. Store - the store number
2. Dept - the department number
3. Date - the week
4. IsHoliday - whether the week is a special holiday week
5. Temperature - average temperature in the region
6. Fuel_Price - cost of fuel in the region
7. MarkDown1-5 - anonymized data related to promotional markdowns that Walmart is running.
8. CPI - the consumer price index
9. Unemployment - the unemployment rate
10. Weekly_Sales - sales for the given department in the given store

- Objective

Forecasting the sales for each department in each store(45).

Walmart Sales Prediction

- Data exploration

1. Splits Data Set: In this question, sales during weekdays, weekends and holiday are absolutely different. We may want to split the dataset according to each time period.

2. Looking at feature summaries and making inferences about the data. Imputing missing values in the data and checking for outliers

- Model building and selection

1. Statistical Methods:

   Auto-regressive Integrated Moving Average (ARIMA)

Walmart Sales Prediction

2. Machine Learning Methods:

1) Random Forest
2) Linear Regression
3) K nearest regression

- Evaluation:

Evaluated on the weighted mean absolute error (WMAE):

$$\text{WMAE} = \frac{1}{\sum w_i} \sum_{i=1}^{n} w_i |y_i - \hat{y}_i|$$

where
n is the number of rows
$\hat{y}_i$ is the predicted sales
y_i is the actual sales
w_i are weights. w = 5 if the week is a holiday week, 1 otherwise

Thanks