Project 2 Report

Dreycen Foiles [dfoiles2]
Olek Yardas [yardsol]

November 8, 2022

1 Introduction

In this assignment, we were tasked with predicting the weekly sales of different departments at several Walmart stores using data from 45 different stores in different regions.

2 Data processing

The data categories in the raw dataset are as follows: Store, Dept, Date, Weekly_Sales, IsHoliday. To generate the training data, test data, and extra training folds, we first remove Date and replace it with two new categories: Week and Year. We assigned a number from 1 to 52 for each week, and a number from 1 to 3 for each year. These categories are much easier for a statistical model to interpret.

3 Methods

To do the regression, we used a very simply model that we know was not going to perform as well as we hoped, but we did not have enough time to improve on. We simply assumed that there was no variation year-over-year. The sales at a given store at a given department would be the same as the same date 52 weeks ago. This method worked surprisingly well and, in our view, gave a ratio of accuracy to evaluation speed.

4 Results

Fold	Runtime (s.)	WAE
1	0.171	2262.42
2	0.169	1787.08
3	0.200	1779.05
4	0.217	1716.12
5	0.283	2400.40
6	0.182	1696.90
7	0.163	2086.96
8	0.267	1750.28
9	0.164	1719.89
10	0.165	1680.96
Avg	0.203	1888.89

5 Conclusion

We struggled to get more robust models to successfully predict weekly sales. Unfortunately, our best model was not good enough to beat all the benchmarks. We attempted to fit a linear using weeks as features for each department in each store, but that was producing predictions that were very off. In this end, by far our best model was simply predicting that weekly sales would not change year-over-year.