Pearson

Final report

Milosz Gray

1. The Idea

I came across multiple different ideas. Some of them were moving average, linear regression, a bit of AI etc. After playing with the dataset I realized that some methods were much to easy and some were overkill for this task. After a couple of days of thinking about this problem, I thought that one option to solve the task is a combination of two solutions. The first solution looks at a much longer period of time. I decided to go with a Kalman filter which can be adjusted to the dataset and may have a long inertion. The second solution neede to be much easier and adjust to short term changes. Here I decided to use a simple average value of two past months. The unknown was how to adjust the weight of the information that comes from the Kalman filter and the average value.

2. Data wrangling

In the submitted program I have focused only on monthly sales and a good predictor. That's why I rejected information about stores or product categories. The first operation that I do is products data import for which I only extract the product ID. The second operation is sales history import. Here I have created a class that holds every line from the the train set. In further steps, it is used to aggregate data by month. This information could be transformed into monthly aggregates but at the moment of writing this part of the program, I was still struggling with myself if I should use more data or not.

3. Model performance evaluation

For this dataset and time that I was left with after coming up with this solution, I didn't have much time to divide the entire set into training and test datasets. In most cases when I work with the model fitting I use a rule of 80/20 train to test sets. In the case here I don't have a sufficient amount of training data I do multiple data splitting and model fitting to find the best result or I try to multiply the data by transforming it into different views (in case of image processing).

4. Results

Please find it in "results.csv" file. I hope that my understanding of this task was correct and I have prepared results for the upcoming month that was not present in the set.

5. Things that I’m most proud of.

I think that the one thing that I’m proud of is the simplicity of this design combined with fairly good results. I think that it’s easy to overthink this problem and spend much more time than is required to solve this task.