Final data analysis exercise - predicting sales

January 5, 2017

In this project the goal is to predict the sales figures of a set of stores over a period of 6 weeks.

Training and testing models

Examine **fellows-train.csv** and use it to build and test various prediction models. It contains the sales figures for all stores from the beginning of 2013 to the middle of 2015. The fields are either self-explanatory or easy to understand using the data.

Scoring

Once you've trained your model, use it to predict the sales figures of **fellows-test.csv** and submit the result to the server using **submit(predictions_list, team_ID)** in **submit_predictions.py**. You can submit as many times as you want.

The score is the RMSPE (Root Mean Square Percentage Error):

$$\text{score} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{y}_i - y_i}{y_i}\right)^2}$$

where $\hat{y}_i$ is the predicted sales figure for a single store in a specific day and y is the actual sales figure. For your convenience this score is implemented in **score_predictions.py**.

Good luck and may the schwartz be with you!