Final Project Proposal

A price elasticity-based LSTM prediction model

Nisim Hurst

Wednesday 25 April 2018

Profesor: Edgar Emmanuel Vallejo Clemente

Team Members: Nisim Hurst, A01012491

Problem description

Densely connected simple neural networks are good when the i.d.d. (independent and identically distributed) assumption holds on the sample data. However, data that is correlated in time or space is a bit trickier to treat.

One example of these kind of task is when you try to predict product sell quantities from price. Right price strategies are difficult to predict because most of the product are subject to behavioral economics variables like the Veblen Effect or changes in demographics that are outside of the scope of the data.

Price elasticity models are usually designed by a human and optimized by a machine, but nowadays these models are falling short behind the velocity of global economic changes. Thus, we propose a machine learning-based method that could cope with those changes.

Namely, we would like to estimate units sold using a learned price elasticity model.

Solution proposal

Hidden Markov models fail to capture long term relationships between the data because of the Markov assumption that a future state is only dependent of the immediately previous state. Also, on Markov chains you have to find a way to compress the whole problem into one variable only.

Thus, we propose will try out a long short-term memory recurrent neural network to be able to capture relationships that go over 1 year long.

Dataset description

The proposed dataset was proposed by Dr. Laura Hervert on and is currently being used on the Research Methods class:

1. **Dimensions.** 78578 samples, 15 features, 1030 unique sku (sequences), 70 categories. About 14 sku ssales sequences per category. The data is spread out evenly over a period of roughly 2.5 years, i.e. missing data over the sequences are trifling.
2. **Variables.** sku, category (subjective), time (week precision), prices, cost, units sold, temperature ranges, rain and whether or not the sku had a promotion.
3. **Description.** Here is a description taken from the original dataset page:
    1. Category- Classification for things regarded as having particular shared characteristics
    2. SKU – Products
    3. Year – Year in which the sale occurred
    4. Week – Week of the year in which the sale occurred
    5. System price – Price in the information system of the company at the time of sale
    6. Competitor's price - Price of the principal competitor's at the time of sale

7. Net cost - Cost of the product without considering the profit margin
8. Units – Sold units during the week
9. Sales amount - Income obtained from the sales during the week
10. Real price - Income obtained from the sale divided by the sold units
11. Min temp - Minimum temperature recorded in the week of sale
12. Avg temp - Average temperature recorded in the week of sale
13. Max temp - Maximun temperature recorded in the week of sale
14. Rain intensity – Scale to define the intensity of the rain during the week of sale
15. Promotion – Binary value 1- for a sale with a promotion involved 0 – for a regular sale.

Questions to answer

1. **Can this dataset be used to train a recurrent neural network that produces over 60% accuracy on the predicted 1 month horizon?** Is the data sufficient? are there any bayes error bounds implicit in the data?
2. **How similar products sales relate to the data?** can we use data of one product category (human labeled) and use it to predict the units sold of another category? can we use a kind of transfer learning from other pre-trained datasets?
3. **Is there any seasonality for some products?** are there any periodic elements that would affect the units sold regardless of the other variables.
4. **How elastic are the products?** how changes in prices and other ancillary variables affect the units sold.

Additional comments

1. Some holyday weekends have no data, we could use a connect-the-dots technique or weighted moving average on those days for completing the information cube.