Price Prediction

May 6, 2021

1 The Problem

The files `train.csv` and `test.csv` both contain data on standardized price movements (column `y`) of different financial assets for particular industry indicators (columns `x0` through `x123`). Your problem is divided into three parts:

1. Develop a model that, to the best of your ability, predicts a price movement given a set of indicator variables. That is, for a vector of indicator variables $\mathbf{x}_i$, you will develop a model f such that $\hat{y}_i = f(\mathbf{x}_i)$ is your best prediction of y_i. Use the data in `train.csv` to develop your model, and use the data in `test.csv` to assess your model's performance. The concept of "best" is for you to decide, but should take into account the task in part 2 below.

2. Now that you have your model you will use it to play a game using the test data. You will apply your model to predict the price movements on the test set, all in one go (predictions are made independent of eachother). You will also provide a single margin of error $M_e > 0$ for each of your predictions. The rules of the game are: if the true value of your prediction lies within the margin of error, i.e $|y_i - \hat{y}_i| \le M_e$ then you win £1, and if not you win nothing. However there is a catch: there is a cost to setting a particular margin of error, and the cost is equal to M_e. For example, for a particular sample k, your prediction is $\hat{y}_k = 1.5$ and the true value is $y_k = 1$. You have set $M_e = 0.6$ and since the true value is within your prediction's margin of error, you will win £1 at the cost of £0.60, and so your profit is £0.40. What is the value of M_e that maximizes your profit on the test data, and what is the average profit per prediction?

3. You are tasked to come up with improvements that would lead to higher profits on an identical game in the future. Provide recommendations with supporting analysis on what you would do to achieve this objective.

Submit your solution in the form of a jupyter notebook running `python 3.X.X` with file name `<first_name>_<last_name>.ipynb`. You may use any python modules that are publicly available.

2 Guidance

There is no single correct solution to this problem. A good solution displays a well organized analysis of the problem, showcasing your thought process and how you approach problem solving. Some general guidance:

1. Use plenty of markdown text in between code cells and comments in the code where appropriate

2. Use multiple visualizations in the form of graphs/charts and make sure they are presentable

3. There is no need to reinvent the wheel
4. Quality over quantity