Python for finance and optimization
Homework II: Machine learning for finance

Download the csv file `trading_data.csv` on the EPI.

Dataset

The CSV contains a table from an artificial bond dealer. This table contains the following columns corresponding to requests from clients:

- `midprice`: the reference price of the bond at the time of the request.[1]
- `id`: identification of the client. There are 4 client ids in the table.
- `buy/sell`: side of the request (+1 for a client willing to buy, −1 for a client willing to sell).
- `answeredprice`: the price answered by the dealer to the client as a response to his/her request.
- `deal`: the first 2000 rows contain 1 if the client accepted the price proposed by the dealer, and 0 otherwise. The last 200 rows contain no information about the occurrence of a trade.

We assume that a client c accepts the price answered by the dealer and indeed trades with a probability that depends on the difference between the answered price S and the reference price S_{ref}. More precisely, we assume that this probability writes

$$\frac{1}{1+e^{\alpha^c+\beta^c(S_{\text{ref}}-S)}} \text{ for a client willing to sell and } \frac{1}{1+e^{\alpha^c+\beta^c(S-S_{\text{ref}})}} \text{ for a client willing to buy.}$$

Question 1:

In this question we ignore the identity of clients and consider that they all have the same behaviour, i.e. α^c and β^c are independent of c : we denote them by α and β.

Write a notebook to estimate α and β using the dataset, and then evaluate the probability of a trade for each of the last 200 requests (rows).

Question 2:

We now consider client ids. Among the 4 clients, there are in fact two types of clients. Propose and code in the notebook a method that classifies the clients into two relevant groups as far as their behaviour is concerned.

Question 3:

Using the identity of clients and the results of Question 2, propose and code in the notebook a new algorithm to evaluate the probability of a trade for each of the last 200 requests (rows).

[1] Nobody can trade at that price but it evaluates the current price based on consensus data.