3. The Winton Stock Market Challenge

3.1 Overview

Goal:

- To Find the hidden signal in the terabytes of noisy, non-stationary data via novel statistical modelling and data mining techniques. In this competition the challenge is **to predict the return of a stock, given the history of the past few days.**
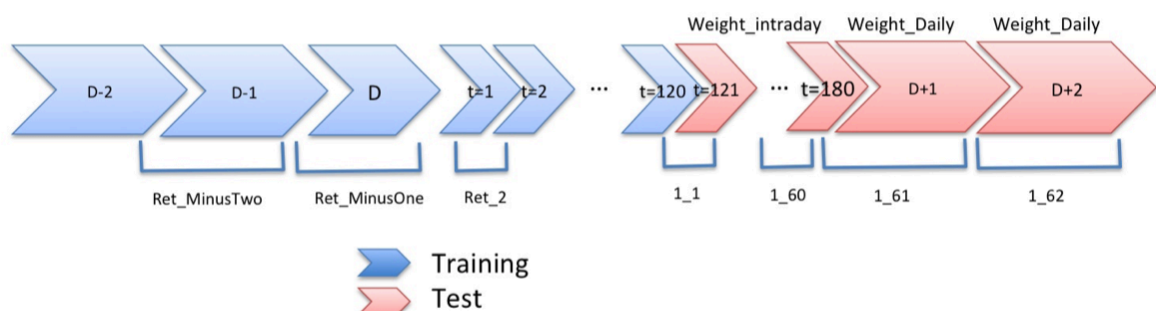
Evaluation Method:

- Provide 5-day windows of time, days D-2, D-1, D, D+1, and D+2. You are given returns in days D-2, D-1, and part of day D, and you are asked to predict the returns in the rest of day D, and in days D+1 and D+2.

- Weighted Mean Absolute Error *Weighted Factors is associated with the return*(similiar with the Benchmark competition):

$$WMAE = \frac{1}{n} \sum_{i=1}^{n} w_i \cdot \left| y_i - \hat{y}_i \right| ,$$

3.2 Data

Basically just train.csv&test.csv, and a csv file for the submission template:

Provide 5-day windows of time, days D-2, D-1, D, D+1, and D+2. You are given returns in days D-2, D-1, and part of day D, and you are asked to predict the returns in the rest of day D, and in days D+1 and D+2.

During day D, there is intraday return data, which are the returns at different points in the day. We provide 180 minutes of data, from t=1 to t=180. In the training set you are given the full 180 minutes, in the test set just the first 120 minutes are provided.

For each 5-day window, we also provide **25 features**, Feature_1 to Feature_25. These may or may not be useful in your prediction.

Each row in the dataset is an arbitrary stock at an arbitrary 5 day time window.

- **train.csv** - the training set, including the columns of:
  - Feature_1 - Feature_25
  - Ret_MinusTwo, Ret_MinusOne
  - Ret_2 - Ret_120
  - Ret_121 - Ret_180: **target variables**
  - Ret_PlusOne, Ret_PlusTwo: **target variables**
  - Weight_Intraday, Weight_Daily

- **test.csv** - the test set, including the columns of:
  - Feature_1 - Feature_25
  - Ret_MinusTwo, Ret_MinusOne
  - Ret_2 - Ret_120

3.3 Selected Solution

Pretty Tricky this one...

3.4 Comment