

Project 2

- Author: Yanan ZHOU
- Contact: adreambottle@outlook.com

Background

Generally, mature quantitative companies will have a very rich **factor library** or **feature pool** internally, and these factors often have some predictive capabilities. The job of some researchers is to combine different factors, that is, to build a model $y = f(x_1, x_2, x_3, \dots, x_N)$

x_i is the factor we can use, f is the model we trained, and y is the predicted value of stock return

All researchers want to improve the performance of their strategies. If everyone uses the same factors, the same labels, and models with similar structures, it will lead to serious homogeneity of strategies.

Problem

Goal: Use market data to find some new features z , that can strengthen or differ the information of the original factors x .

Data: all market data on in Sample, the company's existing factor library in Sample

Evaluation criteria: excess return and correlation evaluation on Out Sample

Method: that is, the process of $f(x) \rightarrow f'(x|z)$

- Use a simple machine learning model $f(\cdot)$ as the evaluation criterion
- $x = x_1, x_2, x_3, \dots, x_N$ is the company's existing factor library
- Find new features z in market data
- Fixed a way to combine z and x to get $x' = x|z$,
 - $x' = x \cdot z$,
 - or $x' = \text{neutralize}(x, z)$
 - or $x' = \begin{cases} ax, z \in A \\ bx, z \in B \end{cases}$
 - or $x' = g(x, z)$ $g(\cdot)$ is a machine learning model
 - or something else
- Get a new prediction of stock return $f'(x|z)$

Difficulties:

- How to find suitable new features z , which can be meaningful in behavioral finance or found through search
- How to effectively combine new features z with original factors x

Simple and feasible method:

- Directly use the genetic algorithm to set the target as $obj = r_m - \delta \cdot corr$
- $r_m = f'(x|z) - f(x)$, marginal return compared to source strategy
- $corr = corr(f'(x|z), f(x))$ The similarity between the new strategy and the original strategy
- Search on market data and find suitable sets of solutions under the condition of limiting the depth of the search tree.

The financial meaning behind it:

- Because many factors which are predictive alphas will gradually decay, there are many reasons for alpha decay. We can make some corrections to the factors according to different market styles.
- If the combination method of simple multiplication is used, it is equivalent to projecting the vector space where the factors are located on the subspace and extracting the information on the subspace.
- If the neutralize method is used to find a new potential market style, the alpha will be corrected.
- If you use some non-linear model, it is equivalent to re-distorting the distribution of factors.