- Project Goal:
  (1) Try to use the machine learning method to boost the single factor into a multi-factor model.
  (2) Try to use the machine learning method to decrease the turnover rate of the stratedgy.

- Machine Learning Methods:
  - **Algorithm 1 : Naïve Boosting**

    (1) Initialize ωi=1/N, i = 1, 2, … N

    (2) For m = 1 to M:

    Fit a classifier G(x) to training data using weights ωi

    (3) Output G(x) = sign[ωi* Gm(x)]

  - **Algorithm 2 : Standard Boosting (Adaboosting)**

    (1) Initialize ωi=1/N, i = 1, 2, … N

    (2) For m = 1 to M:

    Fit a classifier G(x) to training data using weights ωi

    Compute:

$$err_m = \frac{\sum_{i=1}^{N} w_m I(y_i \neq G_m(x))}{\sum_{i=1}^{N} w_i}$$

Compute $\alpha_m = \log(\frac{1-err_m}{err_m})$ update $w_i = w_i \exp(\alpha_m I(y_i \neq G_m(x_i)))$

(3) Output $G(x) = sign[\sum_{m=1}^{M} \alpha_m G_m(x)]$

  - **Algorithm 3 : Revised Boosting**

Note: algorithm 2 is free of order of weak classifier $G_m(x)$ *, because of the penalty coefficient* α_m *, here I want to revise the boosting method to be order dependence ,i.e. exclude the penalty coefficient* α_m *, the order of the weak classifier* $G_m(x)$ *is determined by the* err_m *, the less* err_m *the higher priority, but one thing should be noted, some 0-1 factor* $Y_m(x)$ *are not suitable for the situation , I still let them free of order and has the penalty coefficient* α_m *.*

(1) Initialize ωi=1/N, i = 1, 2, … N

(2) For m = 1 to M:

Fit a classifier G(x) to training data using weights ωi

Compute:

$$err_m = \frac{\sum_{i=1}^{N} w_m I(y_i \neq G_m(x))}{\sum_{i=1}^{N} w_i}$$

Compute $\alpha_m = \log(\frac{1-err_m}{err_m})$ update $w_i = w_i \exp(\alpha_m I(y_i \neq G_m(x_i)))$

(3) K = argmin(err_m)

(4) $G(x) = sign[\sum_{k=1}^{K1} G_k(x) + \sum_{m=1}^{K2} \alpha_m Y_m(x)]$, K1 is the normal factor number and K2 is the 0-1 factor number

- Data Source & data description
  (1) The basic data , i.e. stock day return , last price , high price, low price , volume etc. is downloaded from Wind database.
  (2) The single factor , which is named as 'alphaxx' , these single factors are generated by some un-public formulas based on basic data.

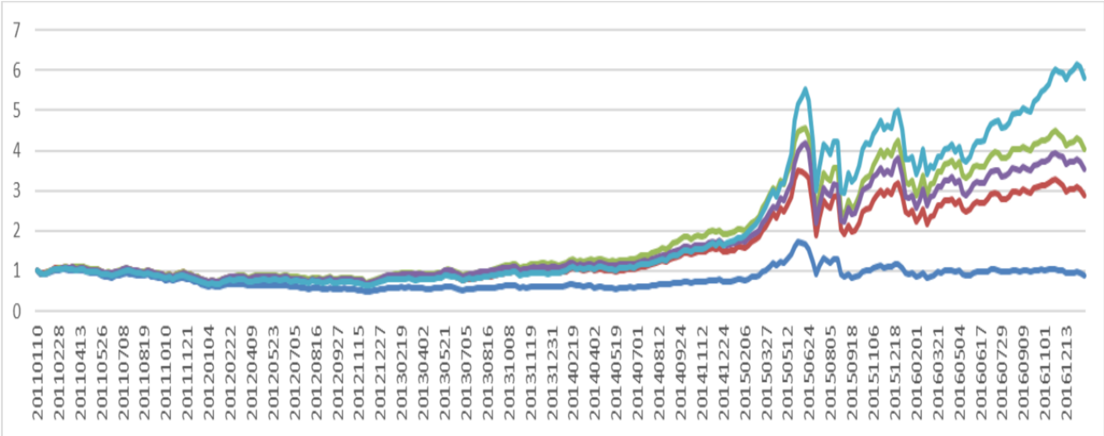
- Further work

Use some filtering methods such as kernel or spline smoothing, wavelet or Kalman filtering to reduce the noise.

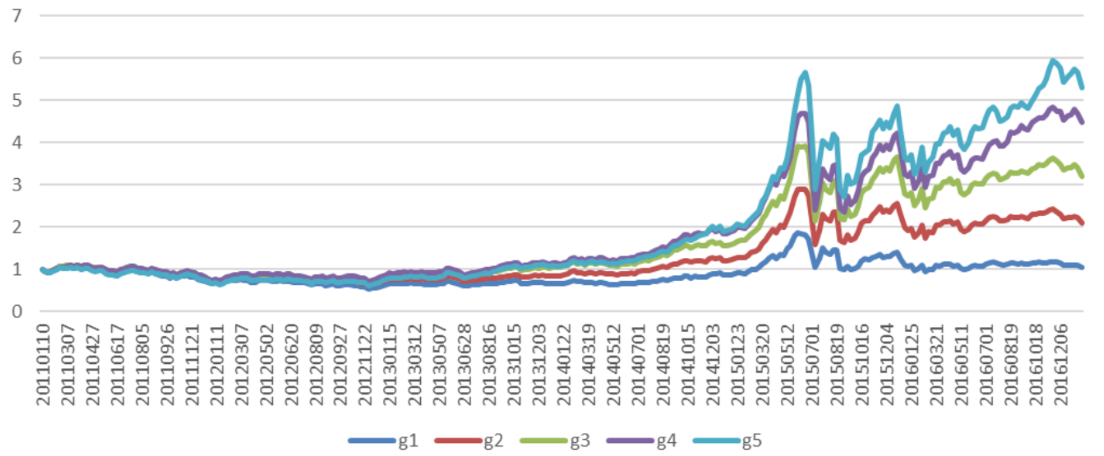
- APPENDIX:

Performance of Some Single Factors

Factor1:



Factor2:



Factor3: