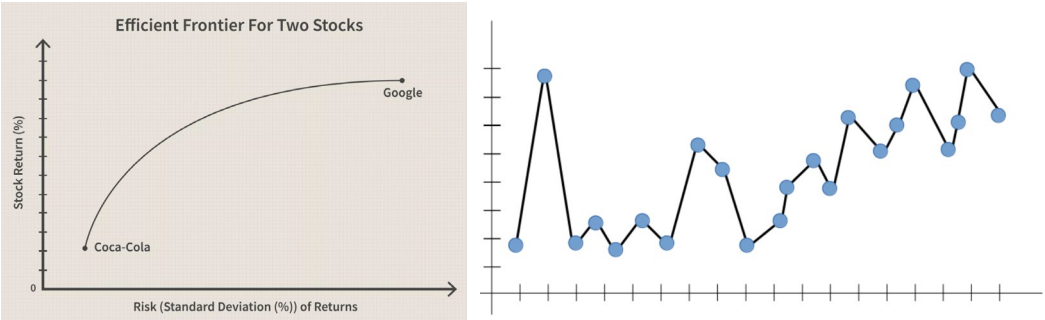
Machine Learning for Portfolio Selection

Other than the traditional portfolio selection through mean-variance analysis, machine learning is an increasingly important but controversial method in portfolio selection. Although machine learning algorithms can uncover subtle, contextual and non-linear relationships, and have proven to be more effective than traditional statistical techniques in many areas outside finance, overfitting poses a major challenge when trying to extract signals from noisy historical data. For this project, we try to use machine learning techniques to forecast the cross-section of stock returns while limiting the risk of overfitting.



Overfitting occurs when a model picks up noise instead of signals. Overfit models have very good in-sample performance, but little predictability on unseen data. Relationships between factors and returns are frequently noisy, and many potential factors exist, increasing the dimensionality of the problem. Because of the low signal-to-noise ratios in forecasting stock returns, it is particularly important to avoid overfitting.

Forecast Combinations is a genius idea of how to make the forecast to be robust and not the result of overfitting. There are already many successful machine learning algorithms which are ensemble algorithms that rely on bagging or boosting. But we can consider to combine forecasts in more dimension (e.g. different training windows, different factor libraries, and different horizons).

Feature engineering is also one of the most effective ways to overcome overfitting because it allows us to increase the signal-to-noise ratio before training the algorithms. In this project, we will consider what we should forecast. For example, instead of predicting expected return, which is close to the white noise, we can predict categories (i.e. outperformer versus underperformer).

In this project, I will present a case study based on the above techniques., and will back-test the performance of this strategy. Here is only the basic idea, and I will enrich it in next steps.