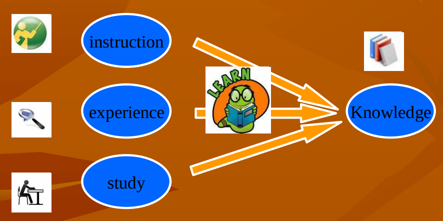
Making Money Automatically? Forecasting The Stock Market!

Han Qin

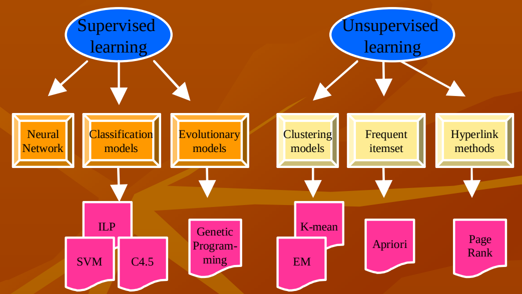
Supervisor: Dejing Dou

How to forecast?

Learning is the process of gaining knowledge from study, instruction or experience. ***Machine Learning*** is the process of programming computers to optimize a performance criterion using example data or past experiences. More broadly, we can say machine learning is the process in which a machine learns something while changing either its structure or its program based on external information so that one can expect improved future performance.



A variety of machine learning models have been developed and most of them fall into two main categories: **supervised learning** and **unsupervised learning**. Supervised learning models learn a mapping from the input to an output whose correct values are provided by a supervisor. On the other hand, unsupervised learning models aim to find the regularities in the input.



Among machine learning models, **Neural network** and **Support Vector Machine** (SVM) are most popular. Neural network models are designed base on the idea of setting thresholds for multiple linear functions to learn the non-linear patterns. Support Vector Machines are learning systems that use a hypothesis space of linear functions in a high dimensional feature space, trained with a learning algorithm from convex optimization theory that implements a learning bias derived from statistical learning theory

Neural Network

- Hidden Layers map to lower dimensional spaces
- Search space has multiple local minima
- Training is expensive
- Requires number of hidden units and layers
- Very good accuracy in typical domains

SVM

- Kernel maps to a very-high dimensional space
- Search space has a unique minimum
- Training is extremely efficient
- Kernel and cost the two parameters to select
- Very good accuracy in typical domains
- Extremely robust

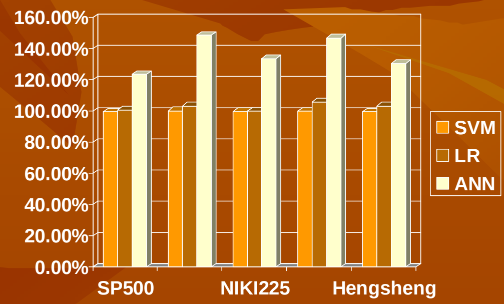
What to forecast?

Financial forecasting is the basis for budgeting activities and estimating future financing needs. The efficient market hypothesis (EMH) was wildly believed and it means that the market price has already reflected the value of itself and the future value is just a ``random walk". Brock et al. argued with the efficient market hypothesis based on the results of experiments tests.

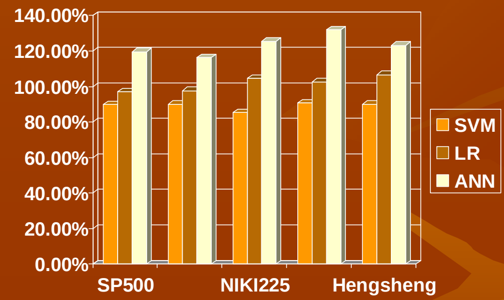
We apply linear regression, neural networks and SVM to several international stock indexes for 1000 shifting time periods and compare the average error rate with the benchmark - random walk model. The measure we use is the relative absolute error (RAE). Both return and volatility are forecasted.

Return: $R = \lg \frac{y}{y-1}$

Volatility: $V = R^2$



Return Forecasting

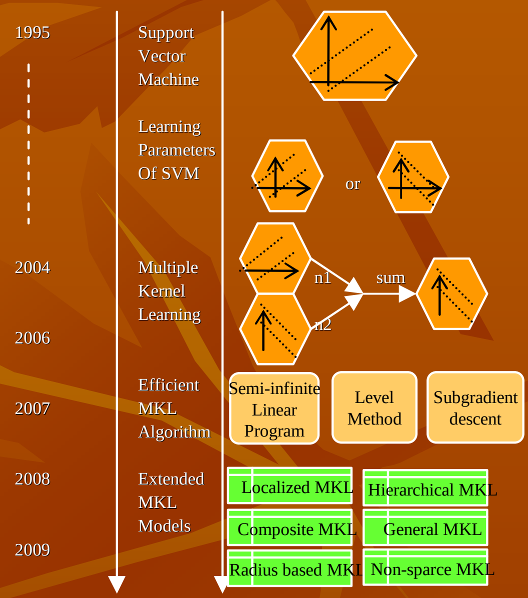


Volatility Forecasting

Conclusion: The experiment results of return forecasting show that with single data source, on average neither neural networks or SVM models perform better than random walk models (i.e., not better than guessing). Nevertheless, the experiment results of volatility forecasting show that both the neural networks and SVM models beats random walk. Furthermore the SVM models with certain kinds of kernels have better accuracy than other models.

Is support vector machine good enough?

Although the SVM methods achieve a global optima with only a few parameters, the performance still replies on the selected kernel. Therefore one question still remains: How to choose the proper kernel or a better kernel for a specific application. The intuitive idea is to search a kernel from the data, which leads to the new machine learning method multiple kernel learning (MKL). MKL is first proposed by Lanckriet in 2004. The initial motivation of MKL is improving the learning accuracy from single data source with different kernels. Another advantage of MKL is to understand the relationships among different data sources. The following figure gives the timeline of the MKL evolution.



Future work:

The tensor combination of multiple kernel matrices is promising since tensor products generate higher dimensional feature spaces so that the searching spaces will be more general. This is related to the new technique **tensor learning**, which has advantages when applying to a smaller training data set. This will also address the problem of non-IID assumption since smaller training data set means the distribution differences between training and testing data set are also smaller.

The implementations of current MKL algorithm depend on the single kernel evaluation in all iterations. Proper approximations of the global optimal will reduce the number of evaluation times compared with starting from random points in the searching space.

Inspired by localized multiple kernel learning, the order of the training data points could be considered in the MKL models as a multiplier function to the kernel matrix. Therefore the function will give higher weight to the data with more similar distributions to the test data. The modified model will address the non-IID assumption problem in a better way than current MKL models.