There exist two kind of financial analysis extended nowadays. First, we have the fundamental analysis, a method in which investors look at the value of the stocks and consider several external markers: Macroeconomic data, political changes, … On the other hand, we have the technical analysis. In that technique, investors consider the values of the stocks and studies other technical indicators based only in the market activity: trends, open and close prices, other statistical values based in transformation...

In this article, we are going to work with technical analysis. Since the non-stationarity of the stock prices (\*explain why is no stationary), the most widely used algorithms are based in ANN (Artificial Neural Networks) and SVM (Support Vector Machines).

Is also common the use of technical indicators that usually smoots the prices along the time. Some classic technical indicators include average, XXXXXX

If we search along different papers, we can find that SVM usually have better performance that ANN. Second, that technical indicators could be worse predicting that using raw prices. [PAPER] Also gets the same conclusion about it.

In general, we have found that some studies use different time periods without a formal criterion to choose that pattern. We study the better period for each company.

In this exercise, we are going to validate if we can find a better SVM performance in the study were technical indicators have the greatest performance we have found: India.

To do it, we have developed a predicting frame that could find which is the best SVM core to find a predictable value in time.

Also: we have tested time with profitability and the raw prices to choose the better performance.

We are using the SVM in regressive and classification mode

We have removed the data of the days that the Company don’t open.

The size of the vectors is based on the length of data of one year. Naturally, one year of data have a different length on each company. But trying to get a simplification the definition of ‘one year of data’, we have use the median of the companies’ vectors length. The reason to use the median instead of the average is only to get a value not so vulnerable to extreme values. There are a lot of companies in a year with less duration that even a half year, so the median help us to find easily a center value: 244.

The goal is to get during full year of data, the best SVM method for each company. Using 3 different kernels: "linear", "rbf", "sigmoid".

To do it, we have trained the SVMs with samples of:

-Half year

- 2 years

- 4 years

with a sliding window that can be moved in intervals of half year, 4a whole year and two years.

Schema:

Vector1: X X-1 X-2 … X-W

Vector2: X+1 X X-1 … X-W+1

Vector3: X+2 X+1 X … X-W+2

…

Vector244: X+244 X+243 X+242 … X-W+243

To train the SVM no future values are used, only from past.

It’s used two SVMs:

1. In regression configuration
2. In classification configuration

And each one is trained using only the raw prices and a very basic transformation:

X0=1

Xi’ = Xi/Xi-1

We expected this ratio to work better in this scenario that using the raw prices.

In a first approach, it was tried to normalize the prices using as reference the first value, but it doesn’t take effects on the data by how is SVM built in python.

We used data from Indian Ruppes due to the study that…

BBDD

The first problem is geta ll data from different stock

Resultados y discursión

Referencias