Machine Learning Forecast Project – Forecasting S&P 500

Group Members: Benedikt Graf, Cristian Martinez, Yiran Sun, and Yuxi Ma

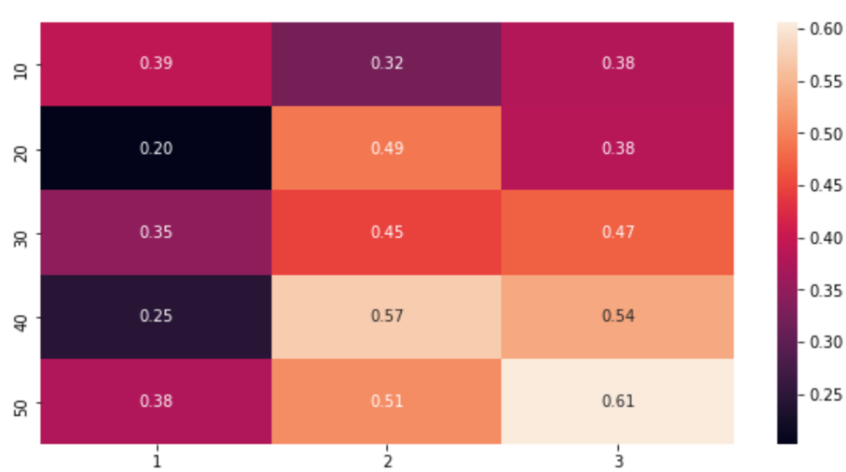
Abstract

In this report, we focus on forecasting the return of *SPY (S&P 500 ETF Trust)* using four predictors *Dollar Index (Dixie), VIX (Cboe Volatility Index), Breakeven Inflation Rate* and *Federal Funds Rate*. As SPY is commonly assumed to reflect the overall market condition, to better measure the change of the market, we take into consideration the international trading, investors sentiment about equity market (namely market risk or behavioral factor), investors sentiment about bond market or the inflation expectation and the risk-free rate.

Since we adopt a panel data with period unit of monthly, we choose to fit a *Recurrent Neural Network (RNN)* model which incorporates time-wise context information and set the dependent variable as the *Sharpe Ratio* of monthly SPY returns. In the following part, we will elaborate on the hyper-parameters we set up and the process of selecting the optimal pair of hyper-parameters.

Hyper-parameters Selection

As the window size and the number of hidden layers are the two vital hyper-parameters in RNN, in the report, we set 5 initial candidate value for window size and 3 for the number of hidden layers, which in total form 15 different pairs of hyper-parameters. In this step, we first split data into three equal sets and use the first section of data as training sample. After fitting models with different pairs of hyper-parameters, here is the heatmap we get.

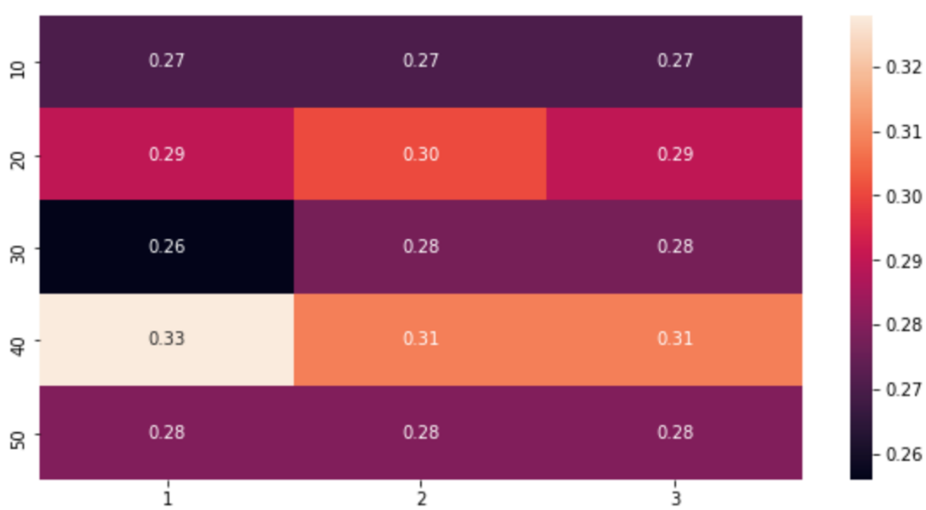


On the heatmap above, the numbers inside is the Sharpe Ratio under different sets of hyper-parameters indicated by the two axis. As Sharpe Ratio measures the amount of returns investors can obtain for each unit of risk taken when holding one asset or portfolio, note that the risk here stands for the total risk, including both market and idiosyncratic. By definition, the greater a portfolio's Sharpe ratio, the better its risk-adjusted-performance. In our case, we take the monthly Sharpe Ratio by taking the ratio of SPY's excess return with respect to monthly risk-free rate (which we set as 1%) and its standard deviation.

Recall the heatmap above, here are 2 observations, 1) Overall, the model seems to perform better when both window size and the number of hidden layers increase, as in the graph, bottom right corner looks the lightest. 2) Targeting the largest Sharpe Ratio, we can see that WINDOWS=50 and LAYERS=3 performs the best, and this is the pair of hyper-parameters we will use for the following testing and evaluation.

Model Evaluation

To evaluate our model, we take the remaining two sections of data as testing sample, and apply the best pair of hyper-parameters. However, the out-of-sample Sharpe Ratio forecast does not seem to perform well, which is approximately -0.3329, implying an even poorer performance than risk-free rate. The heatmap below gives the hyper-parameter selection result by modelling these remaining two sections.



Comparing with the heatmap using the first section of data, this time the heatmap does not give any pattern. Overall, window sizes of 20 and 40 seems to perform better, however, for the number of hidden layers, it seems to be random. Another thing to note here, for these two data sections, the forecast value is much lower than the one using the first section, which is nearly twice in value. In this part, the best pair of hyper-parameters is WINDOWS=40 and LAYERS=1.

Conclusion

In this report, our group fit an RNN model to forecast the monthly Sharpe Ratio for SPY. During the hyper-parameter selection process, we try different sets of training samples, which result quite different results. We suspect there are three possible reasons. First, our data covers periods from 2003 to 2020, which means that in the first section, the market functions well and stably. However, in the remaining two sections, it covers events like 2008 Financial Crisis and recent covid-19 crisis, which suffer from extremely high volatility. This may be the reason of the abnormal differences in Sharpe Ratio forecast. Second, there may be some over-fitting problem in the first section. As compared with the last two sections, the first one is in nature more stable and explainable. Finally, the market itself may be more efficient than we expected.

In the future, we may try to separate the market risk from the idiosyncratic one, namely using the Treynor Ratio instead of Sharpe Ratio to reduce the credit crisis effect incurred by historical events mentioned previously. Besides, we may also try to consider various factors other than the four predictors in this report to better capture the change in SPY.