ERP Report

Creative Category

TEAM PESTO

Feb 2020

1 Prediction

Based on LSTM, we use Bayesian Optimization to tune the parameters and finalize our model by validations.

1.1 Data Selection

use data after 2008 with all features included more illustration...

1.2 Long short-term memory (LSTM)

Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. Comparing all the models we have with different numbers of hidden layers, with/without dropout layer, stacked/unstacked models, we have determined ... to be the optimized model with the highest R-squared in the period from ...to... More illustrations.....

1.2.1 Bayes Optimization

To optimize all the models we have and support for future model selections. We use Bayesian Optimization to tune the parameters in each model. Mathematical illustrations will be added soon!

1.2.2 Gaussian Process

Mathematical illustrations will be added soon!

1.3 Visualization

Use D3 to create an interactive visualization of predicted result. You will be able to interact with it on our website! more illustration...

1.4 Future Improvements

1.4.1 Purging Cross Validation

2 Trading Strategy

2.1 Darvas Boxes

Darvas boxes are an indicator created by drawing a line along lows and highs to make the box. [1]

2.2 Trading Logic

We believe in a short period the stock price fluctuate within the "box". Inside a box, when the current price hits the lower bound and the predicted future price has an increasing trend, the trading decision is BUY. Similarly, when the current price hits the upper bound and the prediction future price has a decreasing trend, the trading decision is SELL. Otherwise, HOLD.

$$\begin{cases} SELL, if Predict_{current} \leq Bound_{lower} \ \ \& \ \ Predict_{future} < Predict_{current} | loss \geq \sigma \\ BUY, if Predict_{current} \geq Bound_{upper} \ \ \& \ \ Predict_{future} > Predict_{current} \\ HOLD, Otherwise \end{cases} \tag{1}$$

where σ is the stop-loss order which needs to be determined by more research on the past performance of $SP500$.
In addition, the upper bound and the lower bound is determined by the average of lowest 3 and average of highest 3 price in the past n trading days since previous starting day(the first day of a box), where n needs to be tested on.

2.3 Visualization

Same as the visualization section in prediction!

2.4 Future Improvements

Investigate on more indicators and criteria that limit our trading decision including MACD, risk indicator, volatility coefficients and so on.

3 Summary

Different from some professionals who insist on variable transformations, feature engineering and traditional regressions. We trust in neural network, which automatically select valuable variables and is robust to missing values. Also, with accurate enough data, the amount of

[1] https://www.investopedia.com/terms/d/darvasboxtheory.asp

data which is needed for a decent model is not overwhelming which allow us to use data after 2008. More illustrations...