

Machine Learning Course Design - Quantitative Trading of Hang Seng Index

Introduction

This project represents the culmination of the Machine Learning course, focusing on the development of a quantitative trading strategy for the Hang Seng Index. The goal is to enhance the index performance using machine learning models and algorithmic trading techniques. The project spans data preprocessing, feature engineering, model development, intraday trading signal generation, risk control, and performance evaluation.

Project Instructions

my device Information

- CPU: Intel(R) Xeon(R) Platinum 8352Y
- GPU: RTX3090

If you want to try our project yourself, you can follow the steps below and it is strongly recommended that you run it on a device with a high-performance GPU.

environmental preparation

```
1 | conda create -n quant python=3.9
2 | pip install -r requirements.txt
```

time series model building and forecasting

model selection

We try six models to help us to construct our quant model. They are:

1. Exponential Smoothing
2. Prophet
3. ARIMA
4. TCN
5. D-linear
6. TiDE

project running

In `predict.ipynb` :

In this code, we constructed daily, weekly and monthly models to demonstrate the prediction effect of the daily model we selected, and used MSE to compare the performance of the selected models. On our device, this code took about **20 minutes**.

In `produce_expected_rate_of_return.ipynb` :

Since our course goal is: assuming the initial capital is **1,000,000**, the investment date is from **2020-11-17** to **2023-11-17**, to achieve index enhancement, in this code, we use the daily closing prices from **2016-07-12** to before each investment day to train the six models we mentioned above, let it predict the trend of the next 14 days, and calculate the expected rate of return for the 14 days to help us build the quantitative model later. Finally, each model will output the 14-day expected return rate calculated daily within the investment interval, which is summarized in a separate csv file, for example, `Prophet_rate_of_return.csv` . On our device,

the code took about **5 hours**. If you don't want to run this code, we also provide the generated csv file to help you with the following steps.

PS: In this project, we directly trade HSI as individual stocks.

In `quant.ipynb` :

In this code, we utilized the returns generated by the first six models from **November 17, 2020** to **November 17, 2023**. Combining this with baseline indicators such as MCAD, RSI, KDJ, and others, we established our quantitative investment model. In the end, we obtained the three-year investment returns and maximum drawdown rates for each model's corresponding strategy.

In `without_tax_rate.ipynb` :

Our second goal is to build a model based on the 5-minute level data provided, find the best intraday buying and selling timing signals, and write specific trading strategies based on the daily prediction results and intraday buying and selling signals. We assume that the tax rate is very low and can be regarded as no tax. At this time, high-frequency trading can increase the compound interest on our principal as quickly as possible and achieve index enhancement. At this time, our selling strategy is as long as we have money. If we make a profit, we will sell it immediately and obtain funds immediately for the next round of investment. In order to ensure that there is a greater possibility of rising in the future, we analyzed the KDJ indicator and on this basis, we also used MACD to stabilize our judgment.

In `with_tax_rate.ipynb` :

In fact, in real life, tax rates are not negligible. At this time, unconditional high-frequency trading will lead to excessive losses of principal due to transactions.

This is not conducive to us achieving index enhancement, so at this time, our selling strategy is modified to sell when possible at extremely high profits to reduce principal losses caused by taxes during transactions.