Capstone Project Proposal

Martin Piják

Bratislava, Slovakia

Abstract

This is proposal for using Machine Learning methods in futures trading.

Keywords: Machine Learning, Trading, Udacity, Nanodegree

1. Domain

In this project we will investigate futures trading on Chicago Mercantile Exchange (CME) markets with Machine Learning methods. The goal is to find the trading strategy mostly based on the price, Commitment of Traders report (COT) and seasonality pattern. We will compare this trading strategy to commonly used investing approaches as returns of Dow Jones Industrial Average or Bonds.

Example of Machine Learning used in futures algorithmic trading.

- Algorithmic Trading of Futures via Machine Learning

    - In this article you can find details about feature engineering, ML-algorithm selection and training

    - Input vector is 2 years of price and volume data.

- A Machine Learning framework for Algorithmic trading on Energy markets

    - This article deals with general pipeline setup for trading

2. Datasets and Inputs

I will use free data from Quandl.

Data range will be from 1st of Januray 2000. Until current trading day. On average trading year has 252 trading days. This gives us about about 4500 ($252 * 18 = 4536$) data points over past approximately 18 years.

I will retrieve data through Quandl API and then save resulting dataframe as CSV file. This way it will be possible to run project without the need to call Quandl APIs.

CME traded commodities below will be used.

- Gold

- Corn

- Coffee

I will use COT data joined with trading data (OHLC and volume) daily data.

I might look into other free data sources depending on the results I will get.

3. **Data Analysis Tools**

   I plan to use following libraries/frameworks.

   - Scikit

   - Keras

   - TensorFlow

4. **Solution Statement**

   Main goal is to create trading strategy.
   Input will have approximately following shape:

   $$2 years data 2*252 trading days = 504504*6(number of features in a day) = 3024 features$$

   Output of model will be classification of input. (two years, 2*252 days = 504 days * number of values vector) as trade (long/short position) or do not trade.

5. **Benchmark Model**

   I will use three benchmarks for my model

   - fixed percentage (bank/bond deposit comparison)

   - Dow Jones Industrial Average performance

   - mean reversal trading on given commodity markets

6. **Evaluation Metrics**

   Trained models will be evaluated compared to Benchmark Models.

7. Project Design

I will look into commodities that's why I want to capture seasonality. First idea I want to explore is to use two years sliding window. I will train different classifiers on the different futures market and based on that make trading decision for the next day. I plan to use daily data, because of that my trading plan is to hold the position (either long/short) for 1–3 days or until stop loss was hit.

I will investigate different features that have impact on trading.

- feature engineering

    - COT
    - price
    - volume
    - date (modify to find seasonality)

- training

    - Random Forest
    - Neural Network
    - . . .

How will be stop-loss selected?

- fixed

    - try different fixed stop-loss values

- floating

    - based on the volatility of previous days