CLASSIFYING DIRECTIONAL STOCK PRICE MOVEMENT

Adeet Patel

Juanjo Capital

September 27, 2019

THE OVERARCHING QUESTION

To what extent can the past history of a common stock's price be used to make meaningful predictions concerning the future price of the stock? (Fama, 1965)

HYPOTHESES

- Stock prices can't be predicted
  - Efficient market hypothesis
    - Prices reflect all available information
  - Random walk theory
    - Maximum prediction accuracy at 50%
- New studies contradict these hypotheses
  - EMH can't explain short spikes over short horizons
  - Behavior driven by news/investor perceptions

BACKGROUND

- Schumaker and Chen (2009): Breaking financial news can be used to predict stock market movements
  - Extract indicators from online sources
- Bollen et al. (2011): Mood states from large Twitter feeds can predict daily up and down changes in the DJIA
- Moat et al. (2013): Frequency of views of Wikipedia's financially-related pages can be an early indicator of stock market moves.

THE PROBLEM

- Two components
  - The information and data sources to gather
  - The machine learning/deep learning algorithms to train on the data

DATA SOURCES / FEATURES

- Using Apple Inc. (AAPL) for project
- Wikipedia and Google News traffic – "collective intelligence"
  - Page hits
  - Disparity
  - Moving averages
- Other technical indicators
  - Relative strength index (overbought/oversold)
  - Rate of change

TARGET VARIABLE

- Machine learning binary classification problem
  - 1 = upward price movement
  - 0 = downward or zero price movement

DATASET & CODE WALKTHROUGH

CURRENT ISSUES

- Feature selection
  - Selecting the appropriate features while also preventing multicollinearity
- Predicting probabilities instead of 1 or 0
  - Some models return erratic predictions
  - Model selection
  - Decision trees, random forest, logistic regression, SVM, KNN, XGBoost, neural network
- Overfitting
  - Example: random forest returning 100% metrics on training data, compared to 50-60% on testing data

WHAT'S NEXT?

- Ensemble various models, and trade only when (almost) all models signal a long
- Find and obtain additional testing data
- Train the models on other stocks
  - Microsoft (MSFT)
  - Google (GOOG)
  - Other small- to mid-cap companies, etc.

COLLABORATORS

- Andrew Holpe (Richmond Quantitative Advisors)