Beating the market:

A simple model to predict asset returns

Cory and Xin

Motivation

- Stock market is volatile and unpredictable
  - Current Asset Pricing models fundamentally flawed
  - Assumption driven (distributional and behavioral)
  - Some empirical prediction worse than random walks
- Goal:

  Without assumptions, develop a data driven model to predict stock returns.

Project-specific questions

At any given time, t , what's a simple strategy to maximize return on a portfolio in the following period?

Question road map:

1. Build machine-learning model to predict y_{t+1} (return)
2. What goes in X_t: feature set, all observable at time t
3. Based on the model, what's the strategy and its performance?

Data Sources

- Data Sources (2010-2019):
  - (Daily) stock prices in S&P 500 constituent
  - (Quarterly) balance sheet, income statement, earnings announcement
  - (Fixed) company factsheet: industry, location

- Feature Set:
  - Technical Features (generated from trailing price and volume), e.g. RSI
  - Fundamental Features (generated from trailing quarter), e.g. liabilities
  - Dummy Features

- Summary (aggregated monthly), >500 features, 50K observations

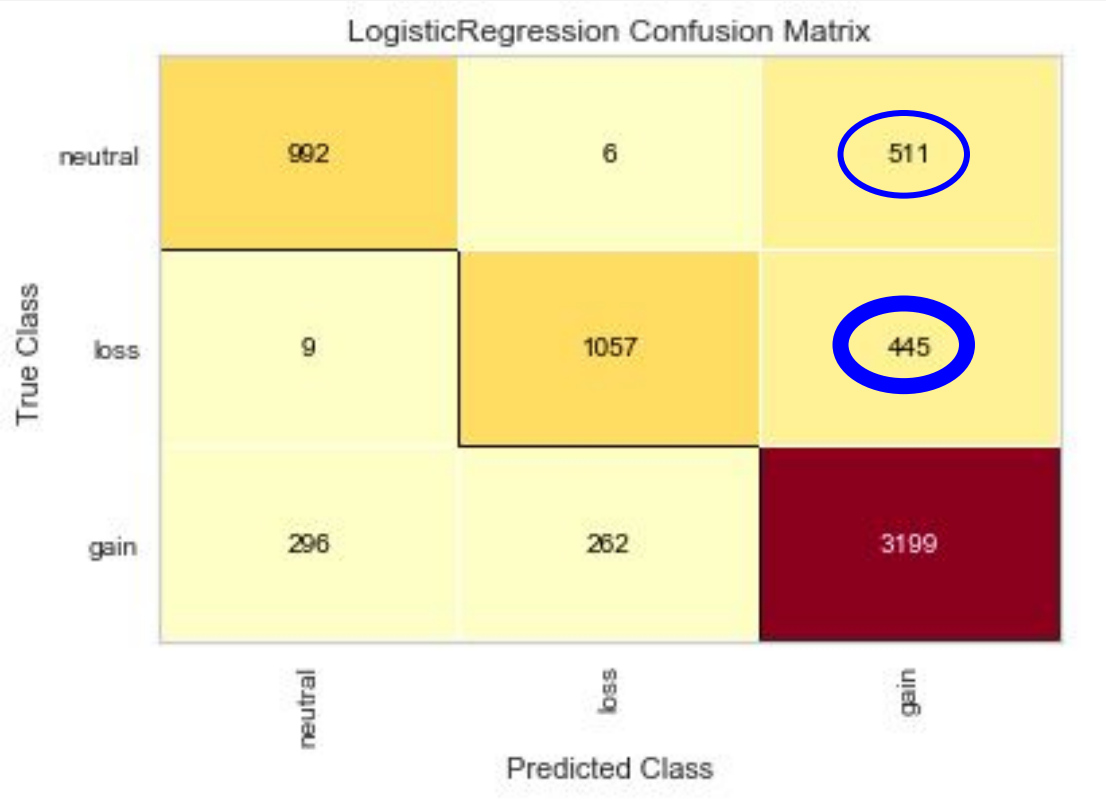
Model and Prediction

- Y: Next Month Return 3 Clusters:
  - Gain (> 5%)
  - Loss (< - 5%)
  - Neutral

- X_t (PCA, Kernel, Polynomial, …)

- Classification Models:
  - Logistic
  - Random Forest
  - SVM
  - Ensembling

- **Average testing accuracy: 76%**
- **Worse Case: 11%**



Model-derived Strategy Vs. Index

Strategy:

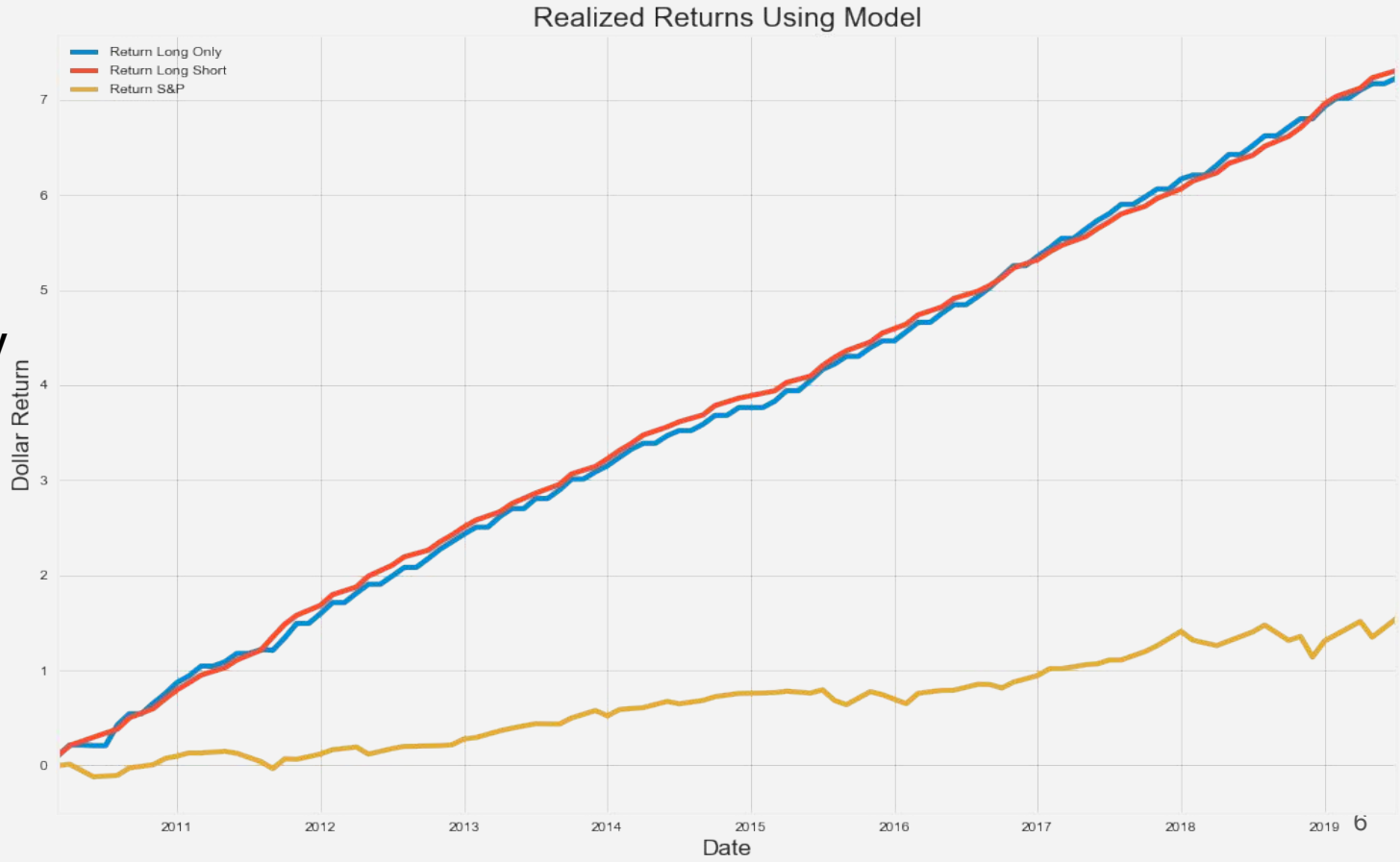
Buy predicted gain

Short predicted loss

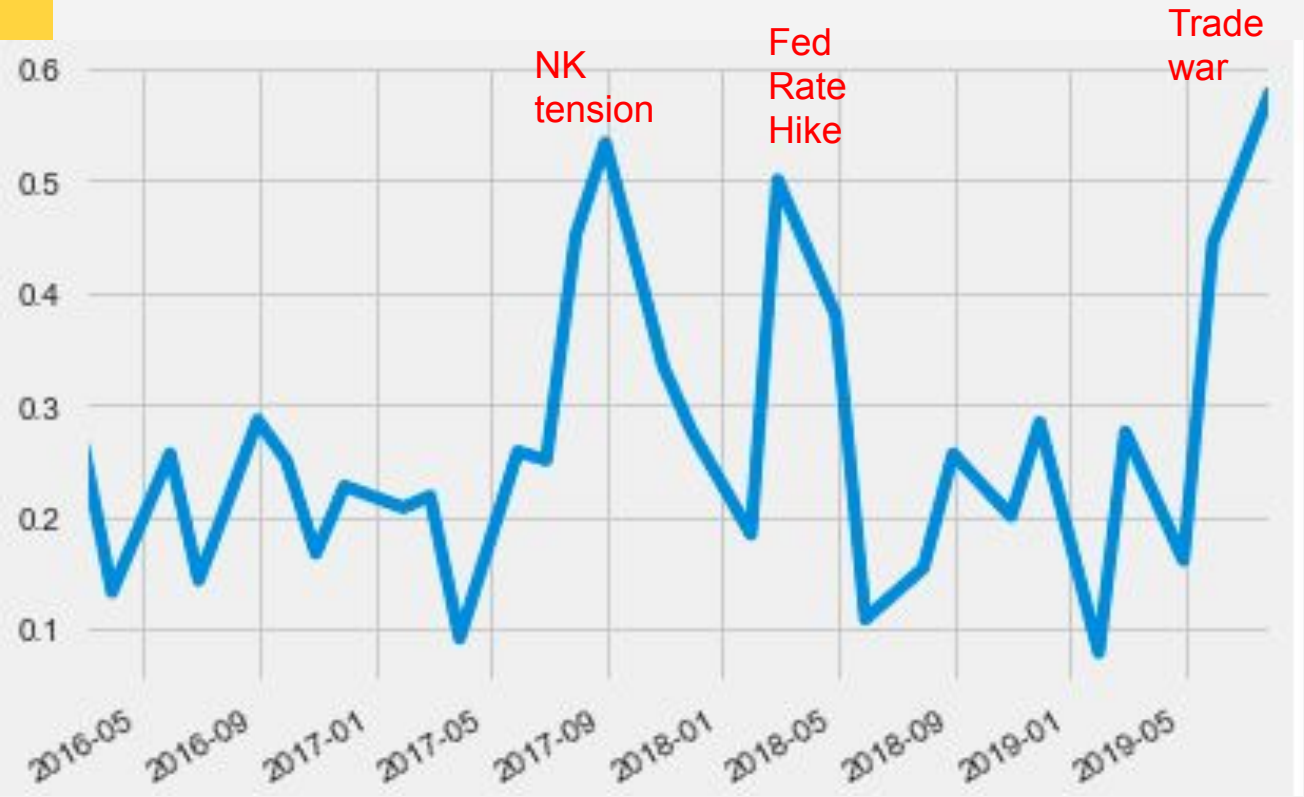
Adjust position every month

Realized Performance:

Strategy: $1 >> $8
S&P: $1>> $2.5



Model Caveat



Predicted Gain, Lost Big

- Volatility in recent years caused our model to fail.
- Propensity of false prediction ratched up to 50%.

Next Steps

Limitation

- Model was trained on bull market

Potential Work:

- Add news sentiment and Macroeconomic features
- Backtest on previous 2008 crisis
- Update data and re-train model parameters

Questions?