PREDICTING THE STOCK MARKET USING MACHINE LEARNING

Jordan Fritz

PROBLEM STATEMENT

- Predict trading signal based on closing price and calculated technical indicators

OUTLINE

- Loaded 15 years of stock market data for AAPL, AMZN, and MSFT
- Feature Engineering (created technical indicators)
- Baseline measurements
- Piecewise Linear Regression
- Trading Signal
- Feature Reduction and Selection (PCA, KernelPCA, FastICA)
- Model Selection

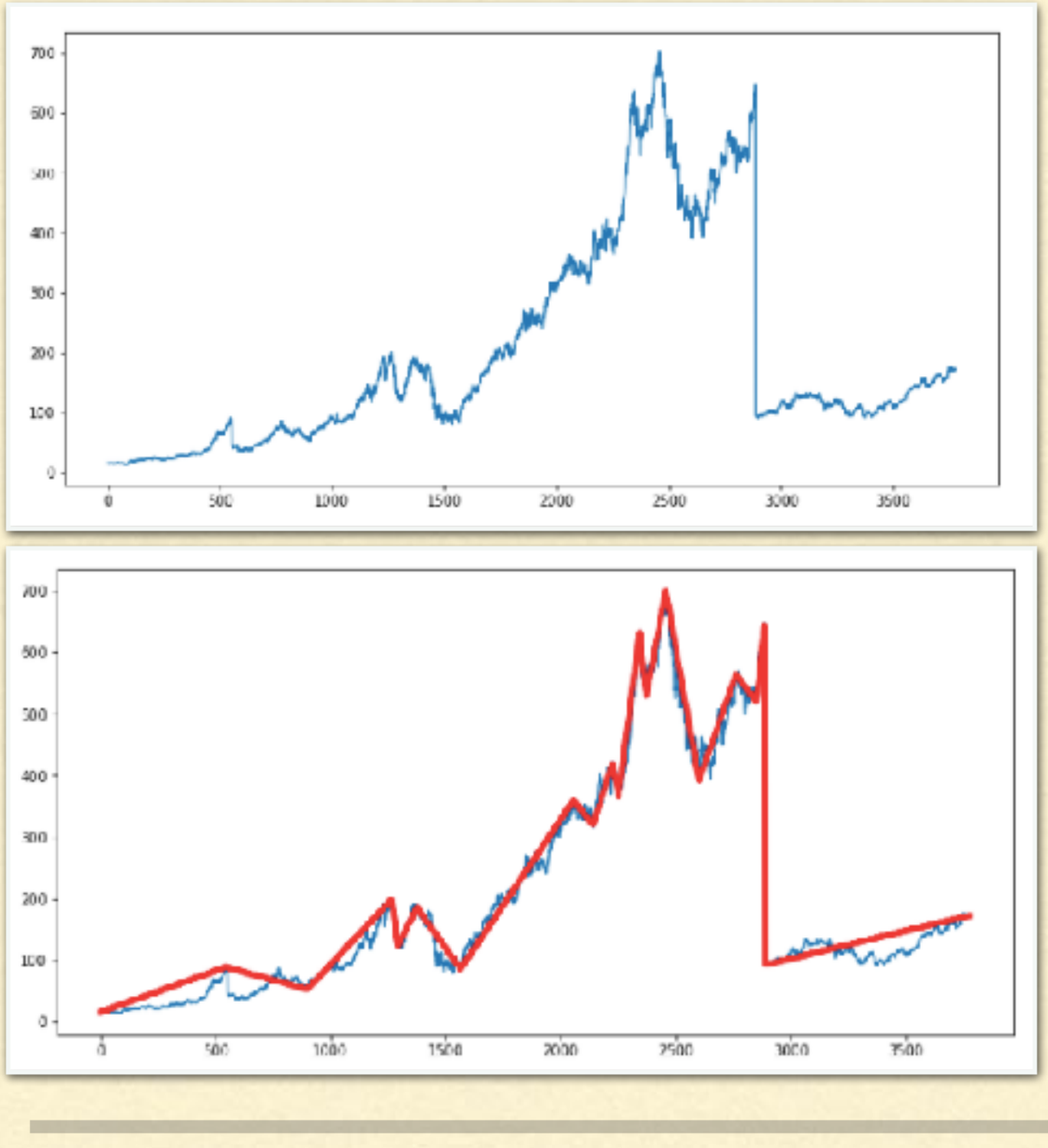
TECHNICAL INDICATORS

```
grid = {"MA": [(1,), (5,), (21,), (63,), (125,), (250,)],
        "STDDEV": [(1,), (5,), (21,), (63,), (125,), (250,)],
        "RSI": [(1,), (5,), (21,), (63,), (125,), (250,)],
        "MACD": [(12, 26)],
        "BBANDS": [(1,), (5,), (21,), (63,), (125,), (250,)],
        "MFI": [(1,), (5,), (21,), (63,), (125,), (250,)],
        "Chaikin": [()],
        "EMA": [(1,), (5,), (21,), (63,), (125,), (250,)],
        "KST": [(10, 10, 10, 15, 10, 15, 20, 30)],
        "TSI": [(25, 13)],
        "TRIX": [(1,), (5,), (21,), (63,), (125,), (250,)],
        "STOK": [()],
        "STO": [(1,), (5,), (21,), (63,), (125,), (250,)],
        "ROC": [(1,), (5,), (21,), (63,), (125,), (250,)],
        "PPSR": [()],
        "OBV": [(1,), (5,), (21,), (63,), (125,), (250,)],
        "MassI": [()],
        "MOM": [(1,), (5,), (21,), (63,), (125,), (250,)],
        "COPP": [(1,), (5,), (21,), (63,), (125,), (250,)],
        "ADX": [(14, 50)],
        "ATR": [(1,), (5,), (21,), (63,), (125,), (250,)],
        "FORCE": [(1,), (5,), (21,), (63,), (125,), (250,)],
        "ACCDIST": [(1,), (5,), (21,), (63,), (125,), (250,)]}
```

BASELINE MEASUREMENTS

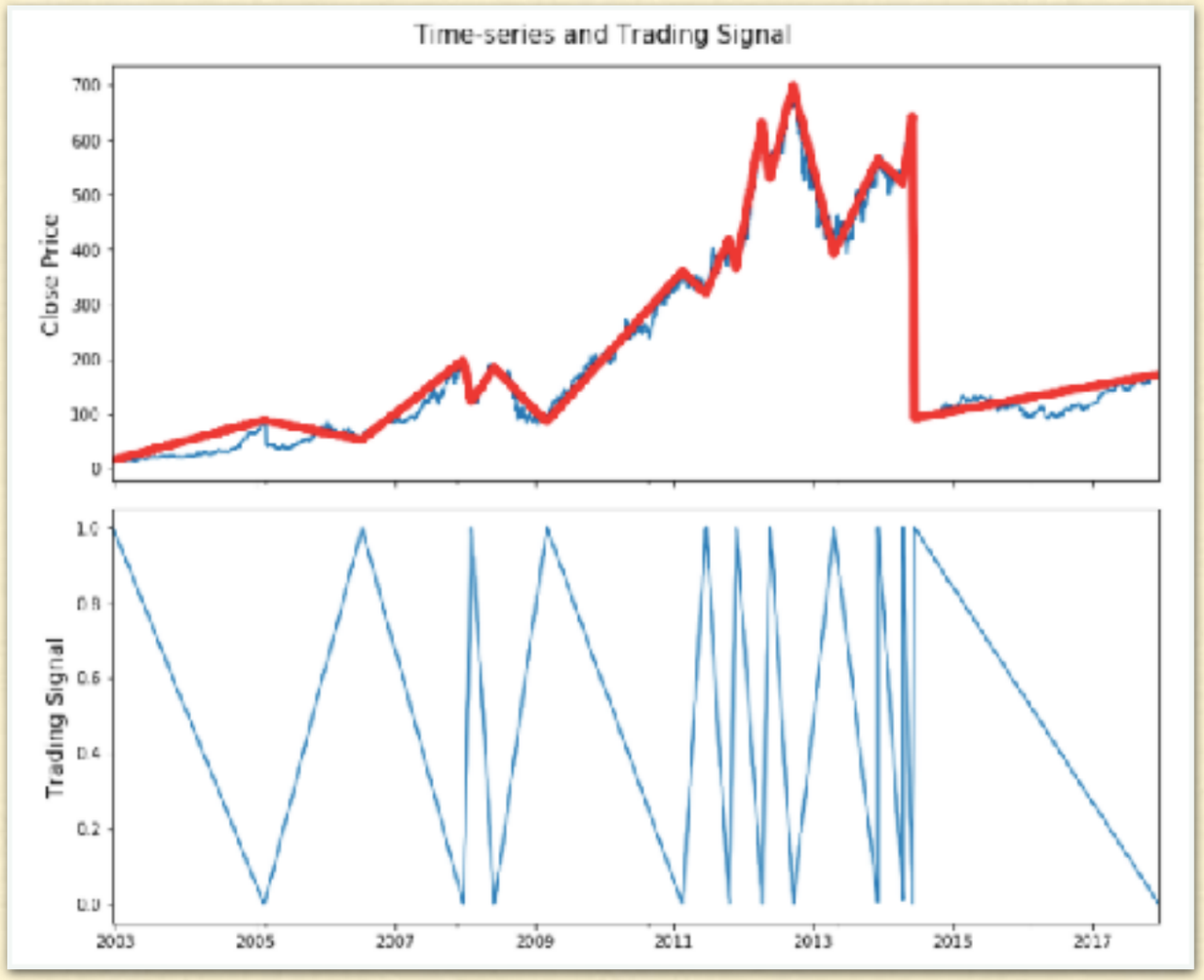
- Used AAPL targets and AAPL indicators
- Predicting tomorrow's close with yesterday's close: 0.99 (R2)
- Predicting tomorrow's close with yesterday's indicators (untuned): 0.98 (R2)
- Predicting tomorrow's close with yesterday's indicators after eliminating as many of the indicators as possible: 0.98 (R2) Lasso, 0.81 (R2) RandomForest

PIECEWISE LINEAR REGRESSION



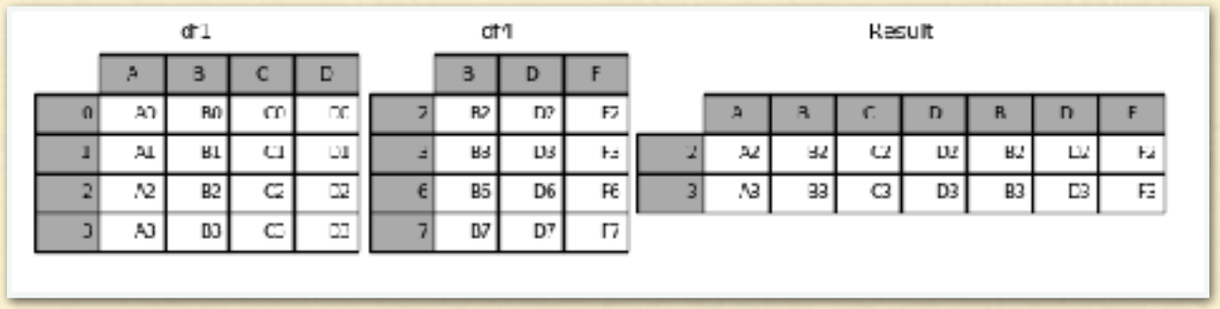
- Created trend-lines to predict when the stock is going up and down using PLR

TRADING SIGNALS

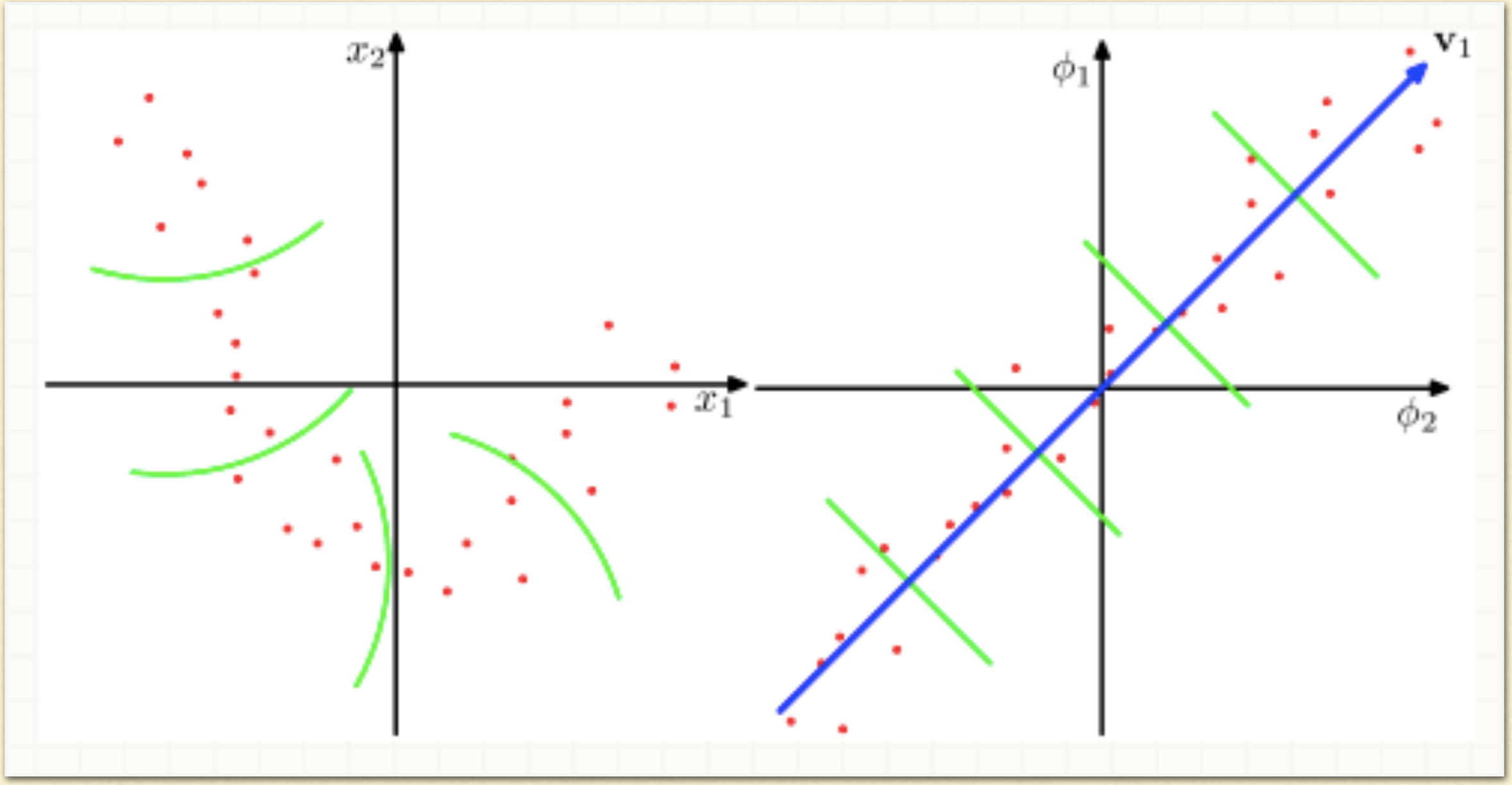


- Using the PLR, created trading signals
- Try to predict trading signal to understand when to buy, hold, and sell

DATA SETUP



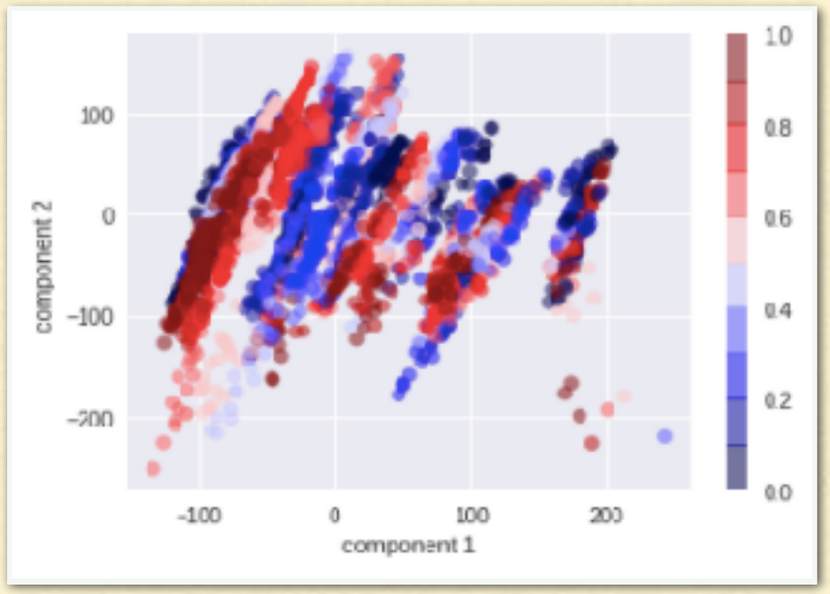
- Used AAPL, MSFT, and AMZN to try to predict AAPL
- Used AAPL target variable (trading signal) and technical indicators for each day for all 3 stocks
- 3525 days (just over 14 years)
- 98 technical indicators for AAPL
- 97 technical indicators for AMZN and MSFT
- 3525 x 291 matrix, 290 TIs and 1 target

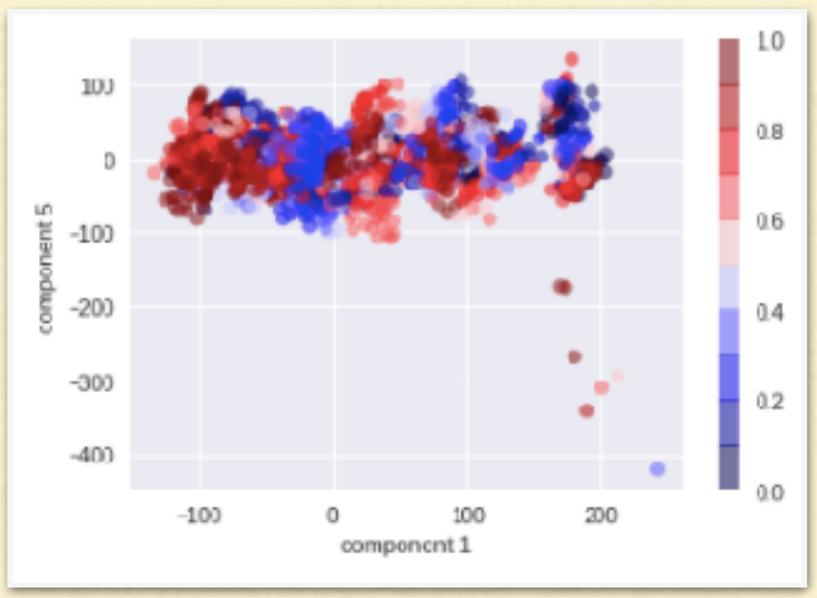


KERNELPCA

- PCA: tries to find a low-dimensional *linear* subspace that the data are confined to
- KPCA: tries to find a low-dimensional *non-linear* subspace that the data are confined to

KERNELPCA

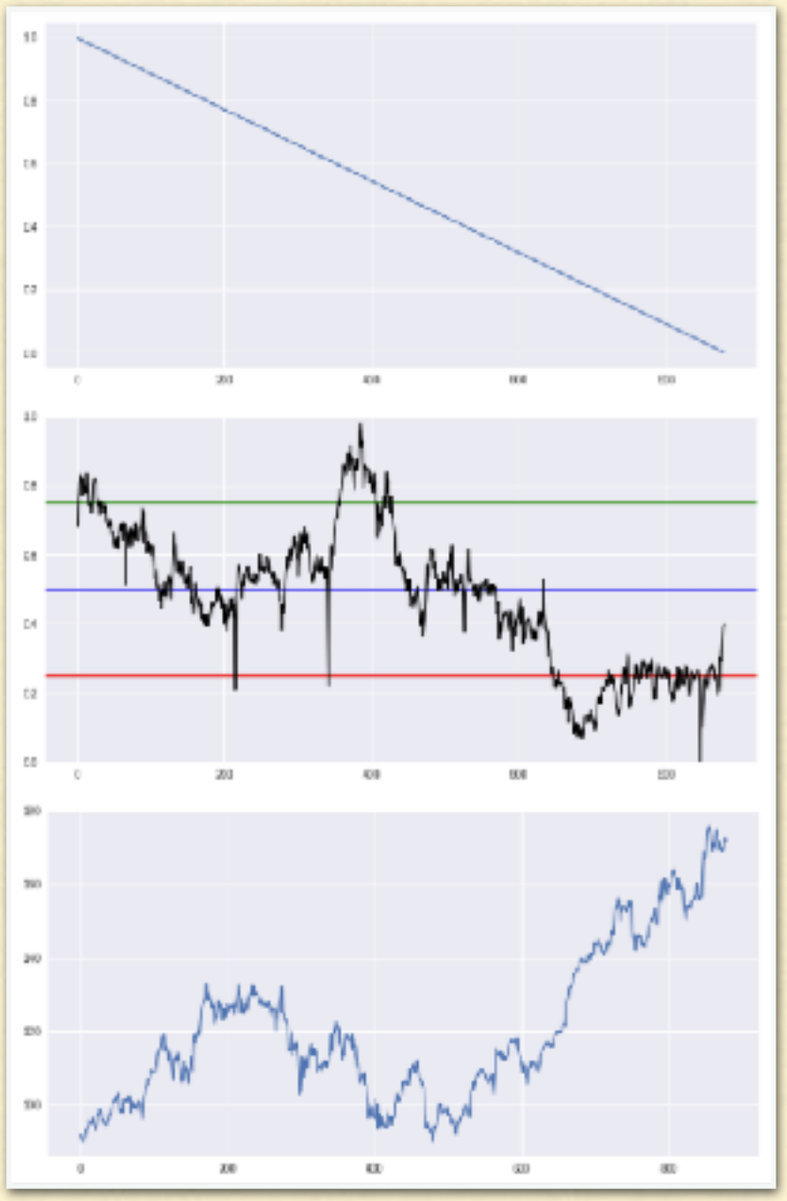




- For feature reduction and selection, tried PCA, KernelPCA, and FastICA
- Reduces number of features from 290 to 31
- Best model

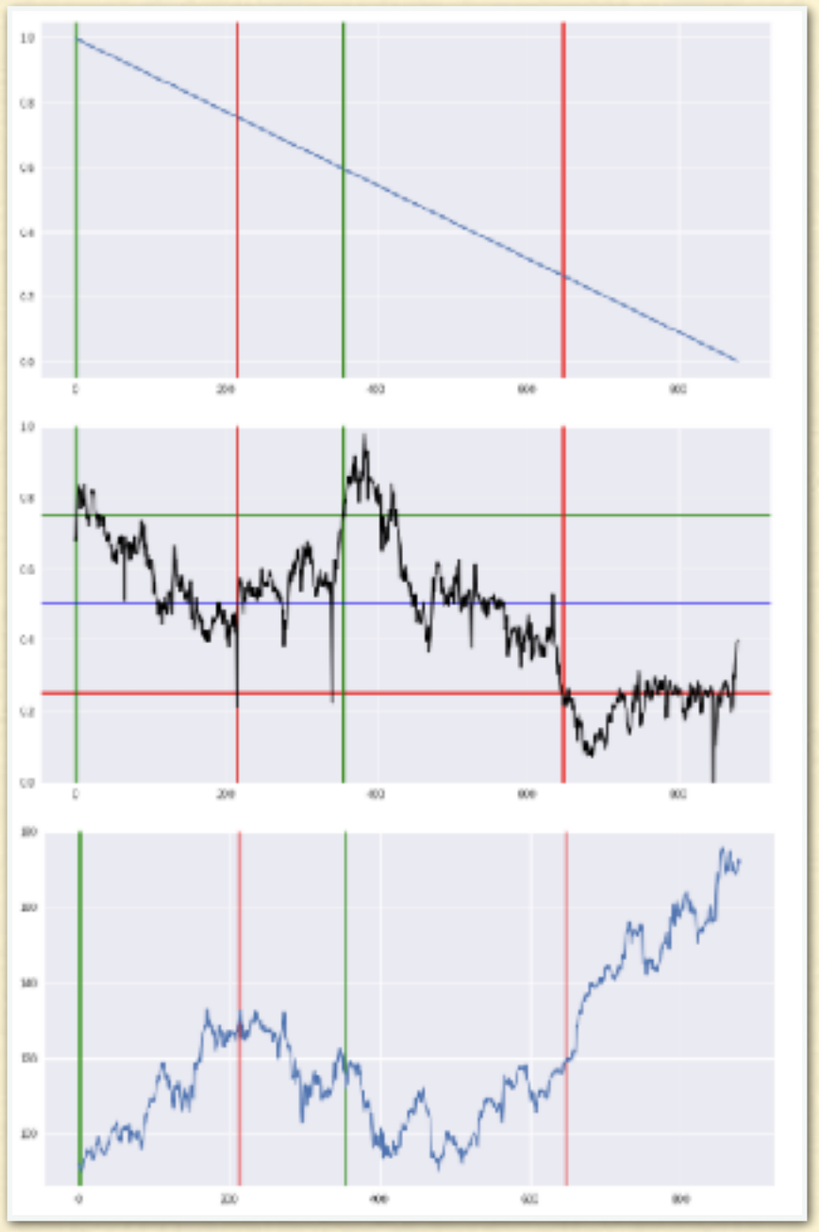
```
kpca = KernelPCA(random_state = 42,
                 n_components = 31,
                 degree = 1,
                 gamma = 100,
                 kernel = 'poly')
```

PREDICTING APPLE TRADING SIGNAL



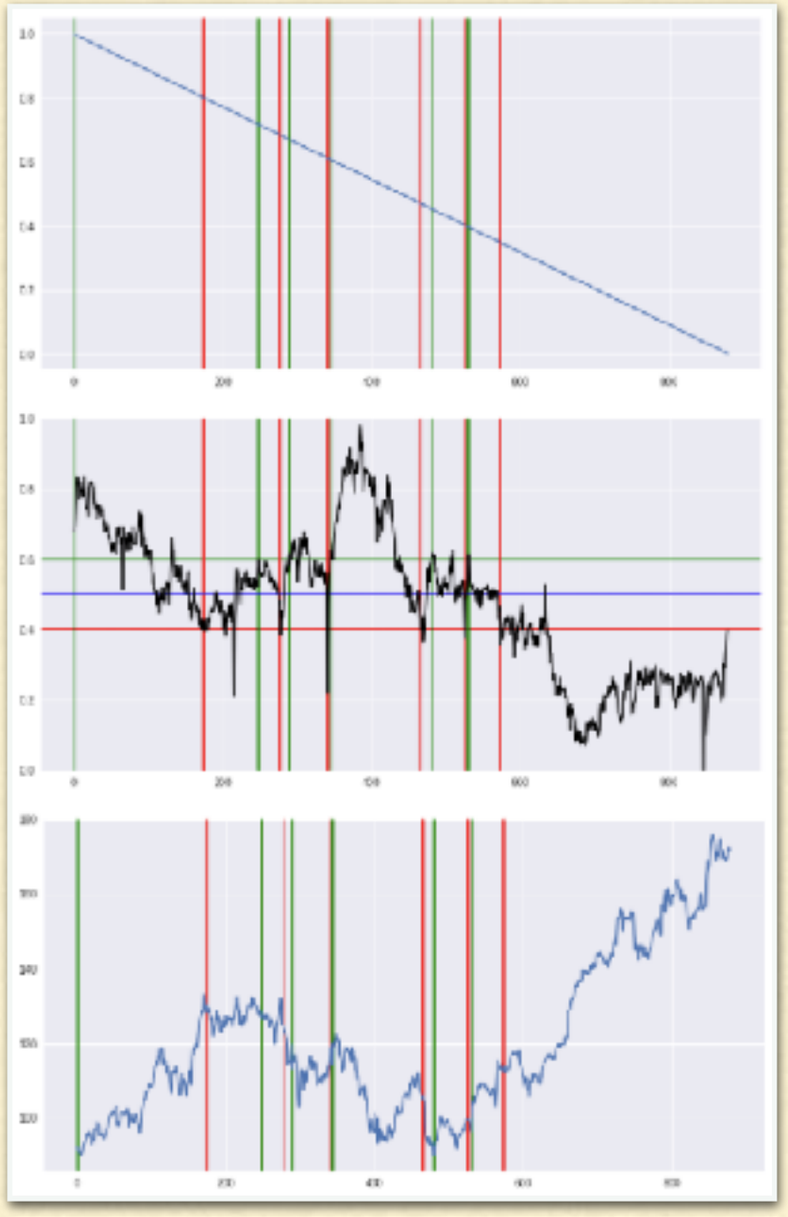
- Fit a Lasso model using the KPCA components found
- Train Score: 0.533 (R2)
- Test Score: 0.530 (R2)
- MAE: 0.168
- MSE: 0.039

PREDICTING APPLE TRADING SIGNAL



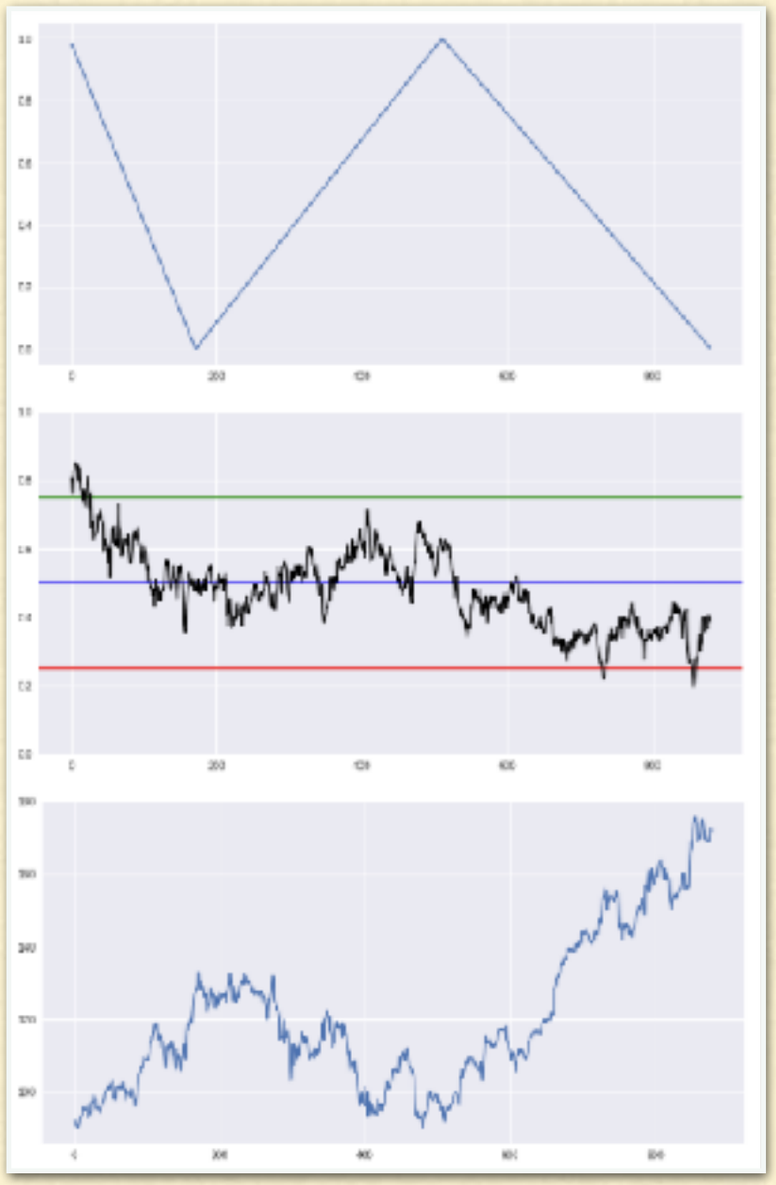
- **AAPL over 880 day test period increased $0.0911/day**
- 0.75, 0.25 threshold for trading
- Total profit of +$46.28
- +$0.092/day where a position was held

PREDICTING APPLE TRADING SIGNAL



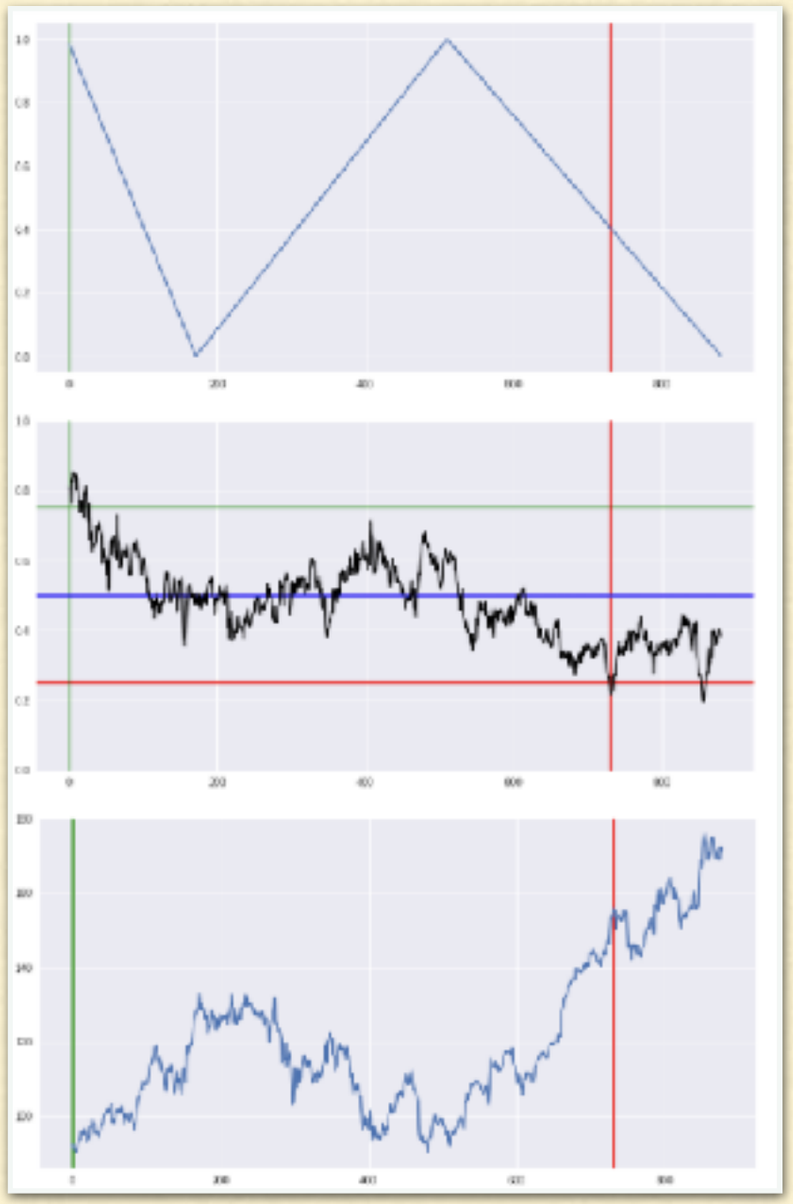
- **AAPL over 880 day test period increased $0.0911/day**
- 0.6, 0.4 threshold for trading
- Total profit of +$45.29
- +$0.097/day where a position was held

PREDICTING APPLE TRADING SIGNAL (LOWER EPSILON)



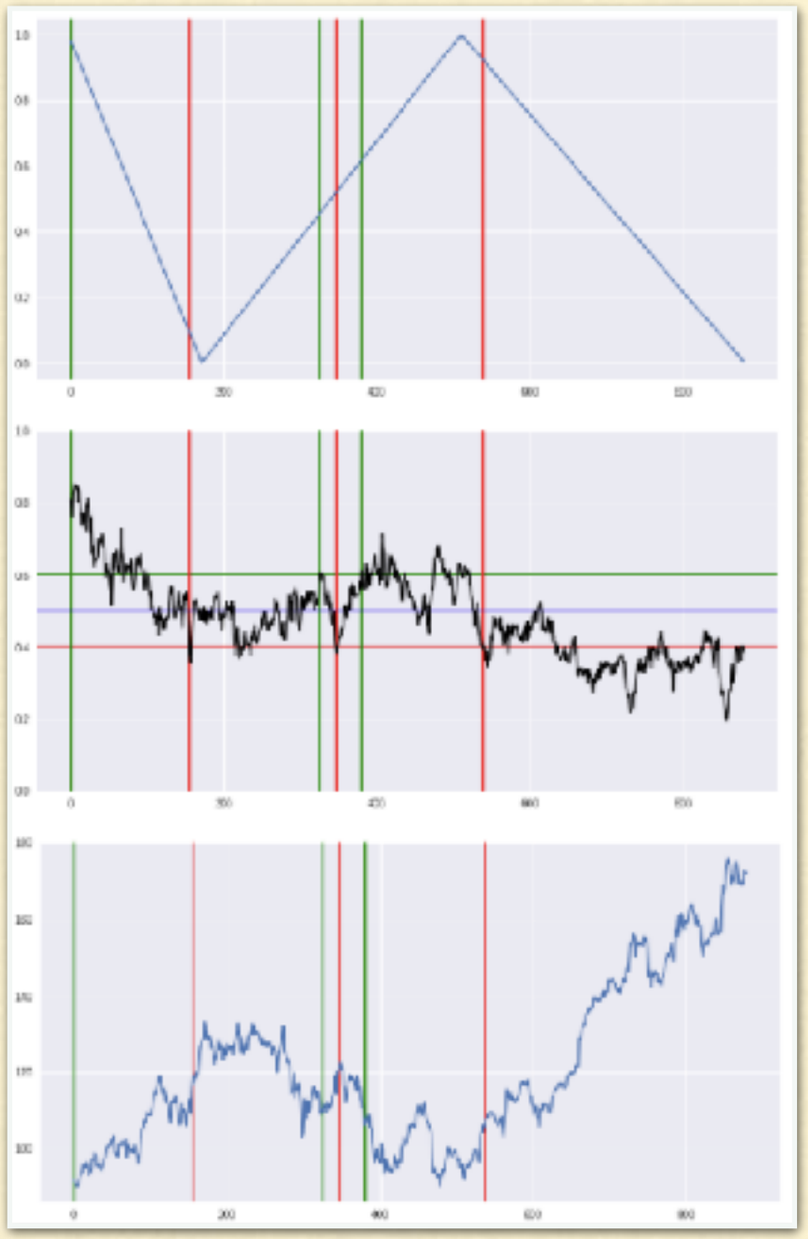
- Lowered the epsilon value to try to include smaller trends
- Fit a Lasso model using the KPCA components found
- Train Score: 0.256 (R2)
- Test Score: 0.207 (R2)
- MAE: 0.219
- MSE: 0.066

PREDICTING APPLE TRADING SIGNAL (LOWER EPSILON)



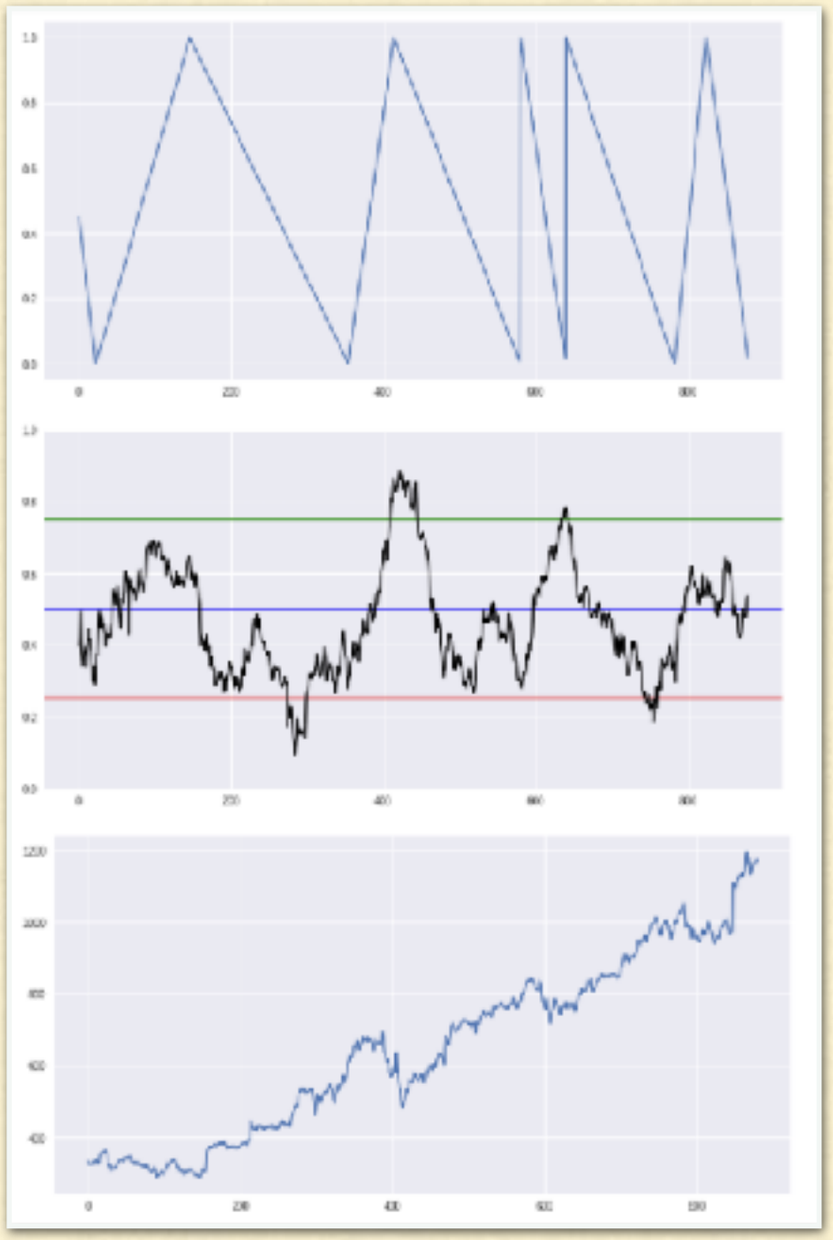
- **AAPL over 880 day test period increased $0.0911/day**
- 0.75, 0.25 threshold for trading
- Total profit of +$63.92
- +$0.087/day where a position was held

PREDICTING APPLE TRADING SIGNAL (LOWER EPSILON)



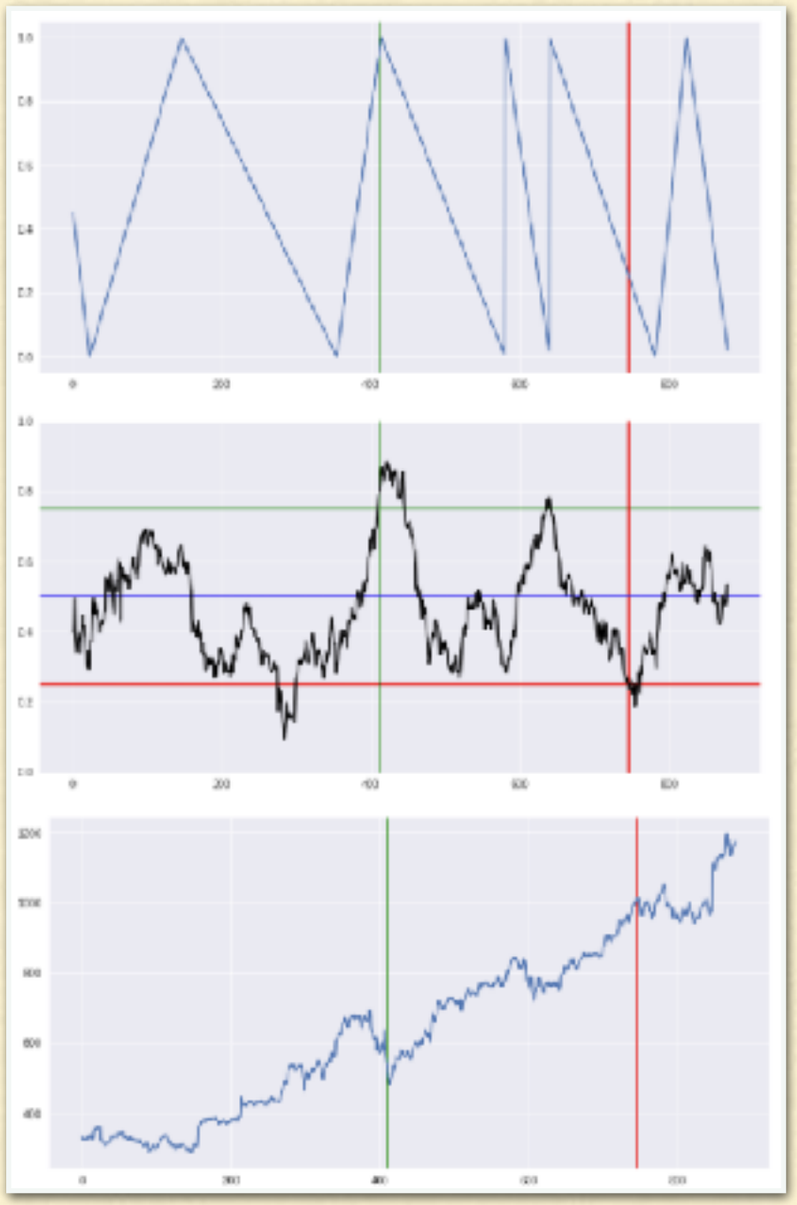
- **AAPL over 880 day test period increased $0.0911/day**
- 0.6, 0.4 threshold for trading
- Total profit of +$38.97
- +$0.117/day where a position was held

PREDICTING AMAZON TRADING SIGNAL



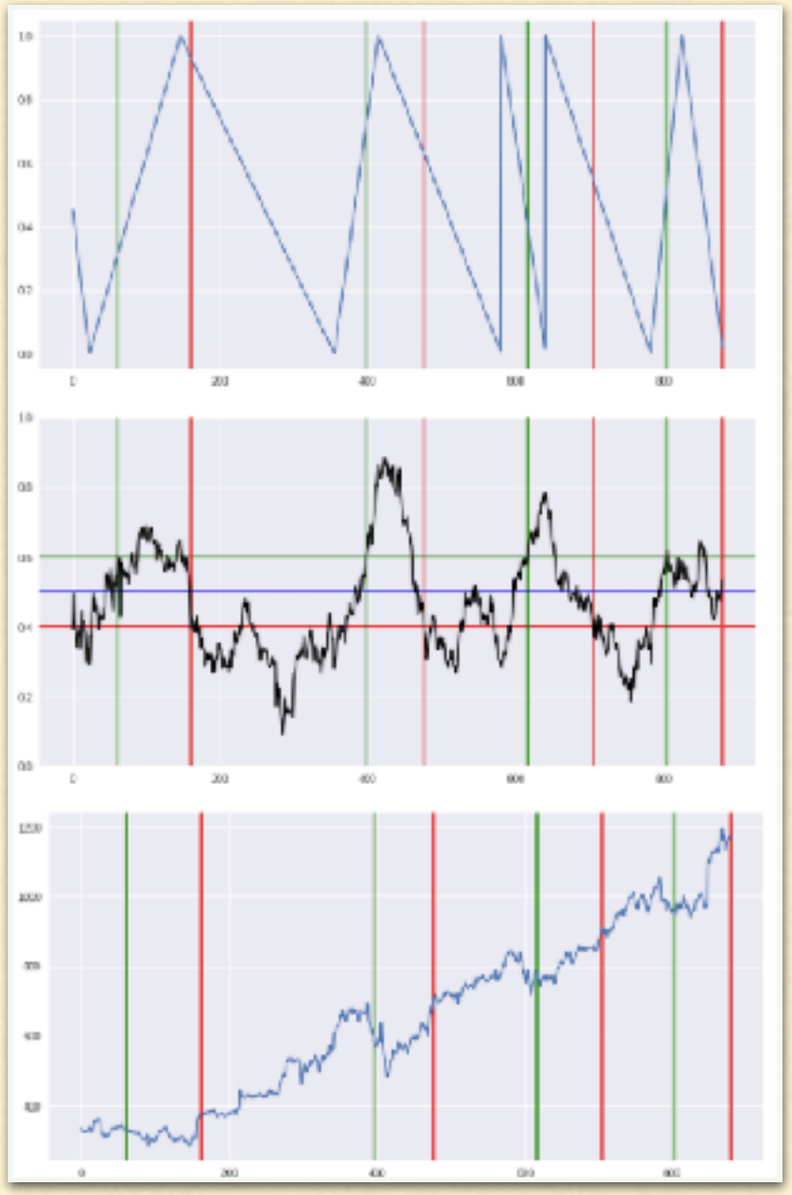
- Fit a Lasso model using the KPCA components found
- Train Score: 0.538 (R2)
- Test Score: 0.193 (R2)
- MAE: 0.217
- MSE: 0.067

PREDICTING AMAZON TRADING SIGNAL



- **AMZN over 880 day test period increased $0.955/day**
- 0.75, 0.25 threshold for trading
- Total profit of +$480.27
- +$0.70/day where a position was held

PREDICTING AMAZON TRADING SIGNAL



- **AMZN over 880 day test period increased $0.955/day**
- 0.6, 0.4 threshold for trading
- Total profit of +$492.60
- +$1.442/day where a position was held

CONCLUSION

- Would like to see what would happen if more data was added and different stock targets were attempted
- Would like a more powerful machine to run a model with technical indicators of all stocks in S&P500 to predict one stock, kernel kept dying