Forecasting Agricultural Commodity Prices through Supervised Learning

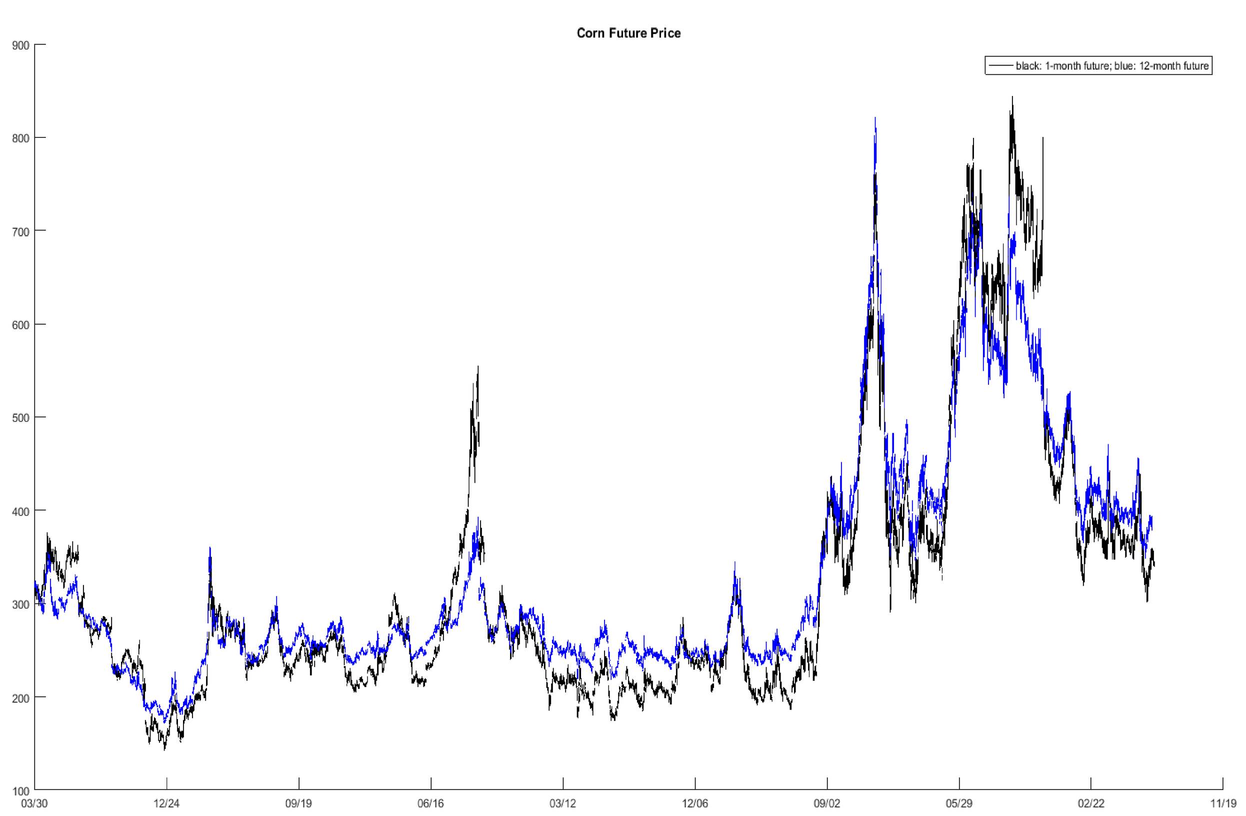
CS229 Project, Fan Wang, wang420@stanford.edu

Introduction

Commodity future is one asset class in financial markets which has historically demonstrated a high degree of volatility. Within the commodity future, agricultural commodities are particularly volatile. The objective of this project is to develop machine learning based model(s) to predict the corn future[1] prices will be up or down for a given number of days in the future.

Data

The daily prices (from 1989-12-21 to 2016-11-11) of 1-month and 12-month futures of three commodities, including corn, crude oil and soybeans, are obtained for analysis.



Features

Four different types of features are explored in this project:

- % price deviation from 5-day, 10-day, 15-day, and 20-day moving average (corn future)
- % price difference for 1-month vs. 12-month contract (corn future)
- % price difference for corn feature vs. soybeans future
- % price change of crude oil future for 5-day, 10-day, 15-day, and 20-day time window

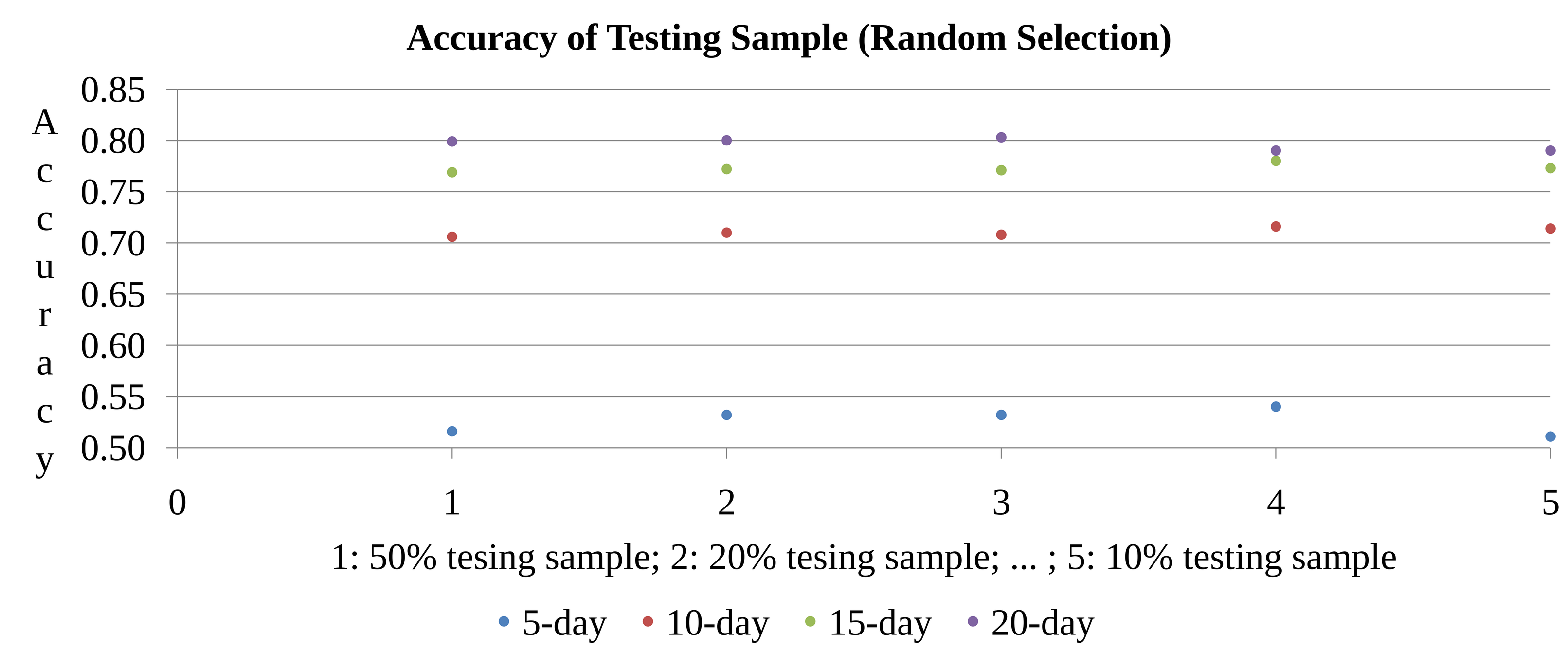
Exploratory Analysis

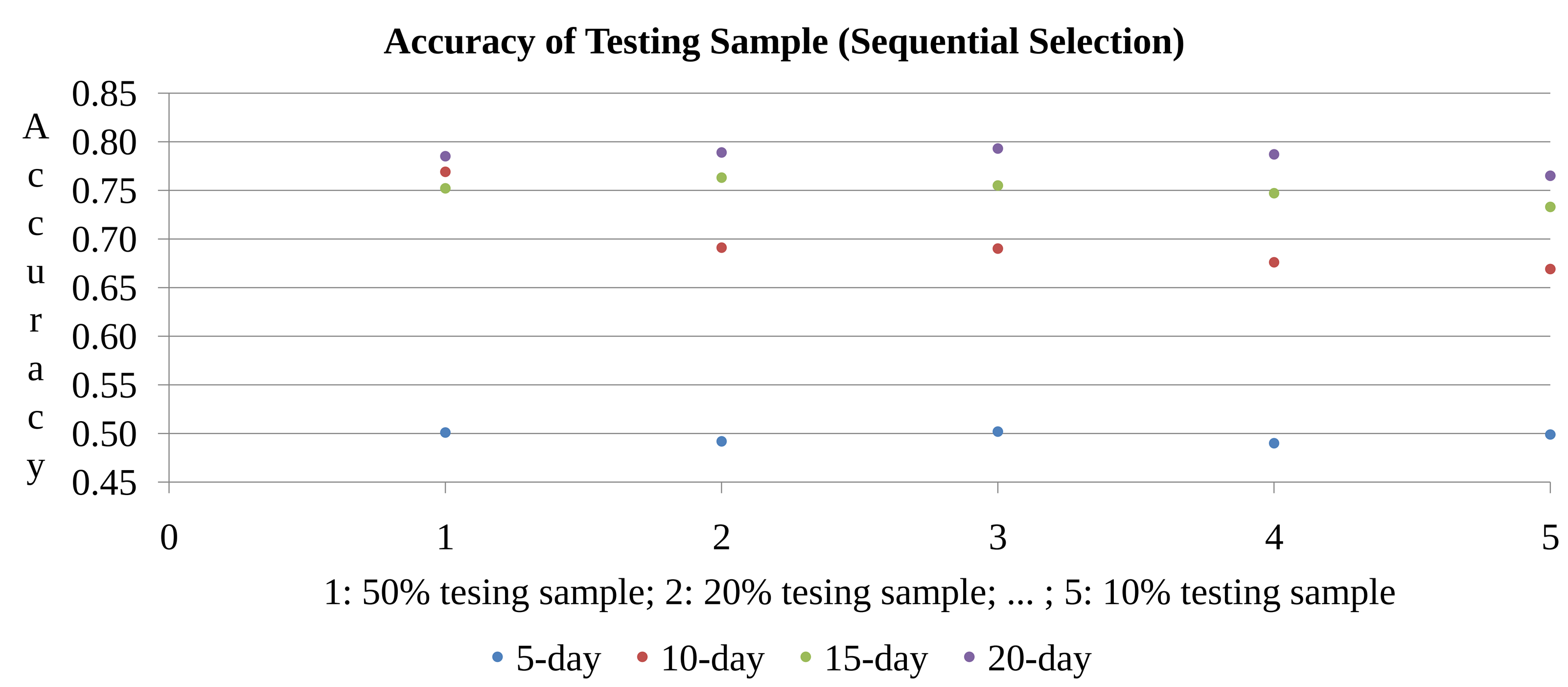
Pearson Correlation Coefficients	1-month Corn	12-month Corn	1-month Crude Oil	12-month Crude Oil	1-month Soybeans	12-month Soybeans
1-month Corn	1.000	0.966	0.773	0.782	0.919	0.936
12-month Corn		1.000	0.855	0.872	0.934	0.970
1-month Crude Oil			1.000	0.989	0.834	0.861
12-month Crude Oil				1.000	0.833	0.873
1-month Soybeans					1.000	0.976
12-month Soybeans						1.000

After performing correlation analysis, we observe the following: 1). 1-month contract and 12-month contract are strongly correlated for the same future; 2). corn is more correlated with soybeans, compared to crude oil; 3). 12-month crude oil contract is slightly more correlated with corn and soybeans, compared to 1-month crude oil contract.

Model 1: Logistic Regression

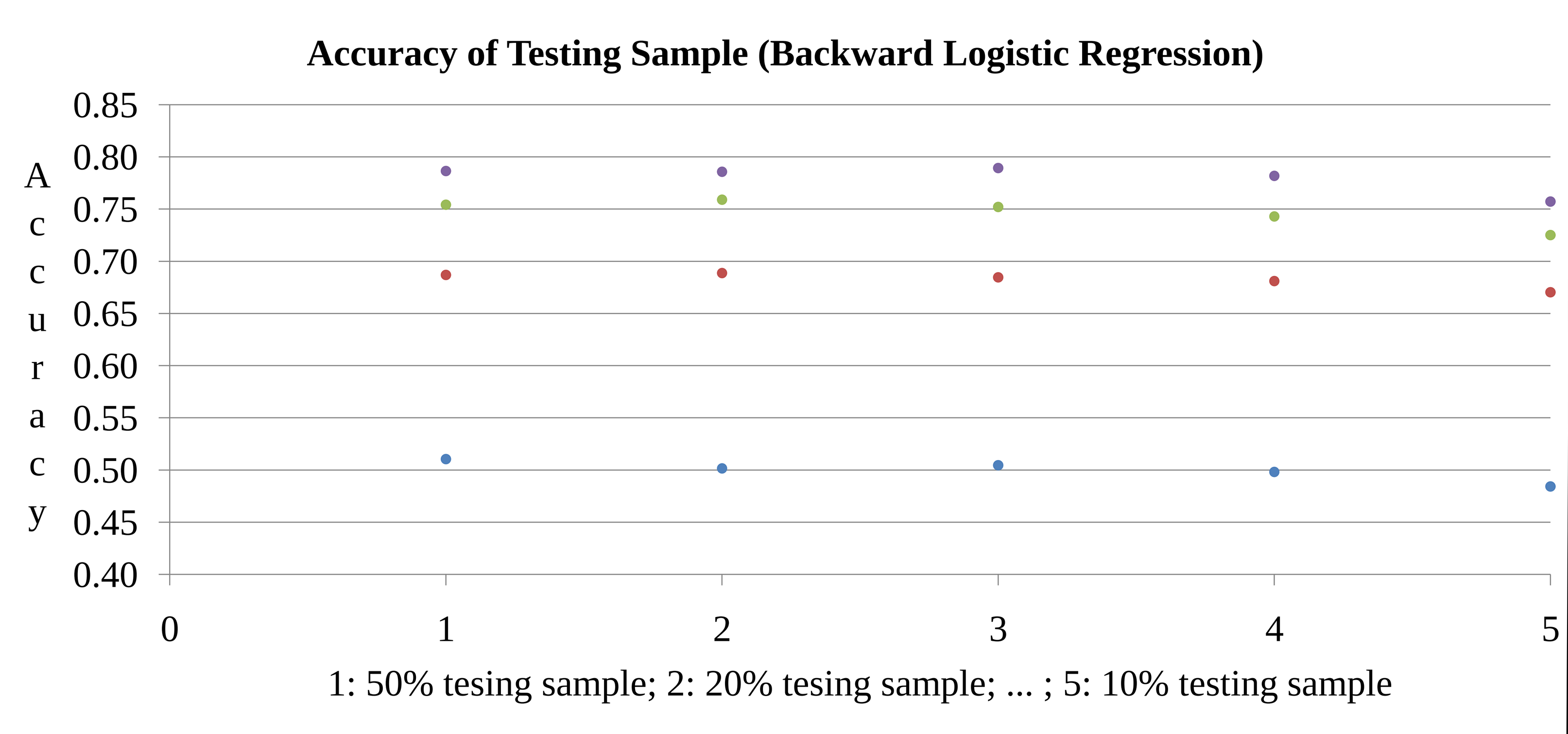
- Target: 5-day, 10-day, 15-day and 20-day return (positive or negative return) of 1-month corn future price
- Features: lag-5 days, lag-10 days, lag-15 days, and lag-20 days of features discussed above
- We train the model on both randomly selected and sequentially selected sample from 50% to 90% of the entire dataset; then use the rest dataset as test sample.





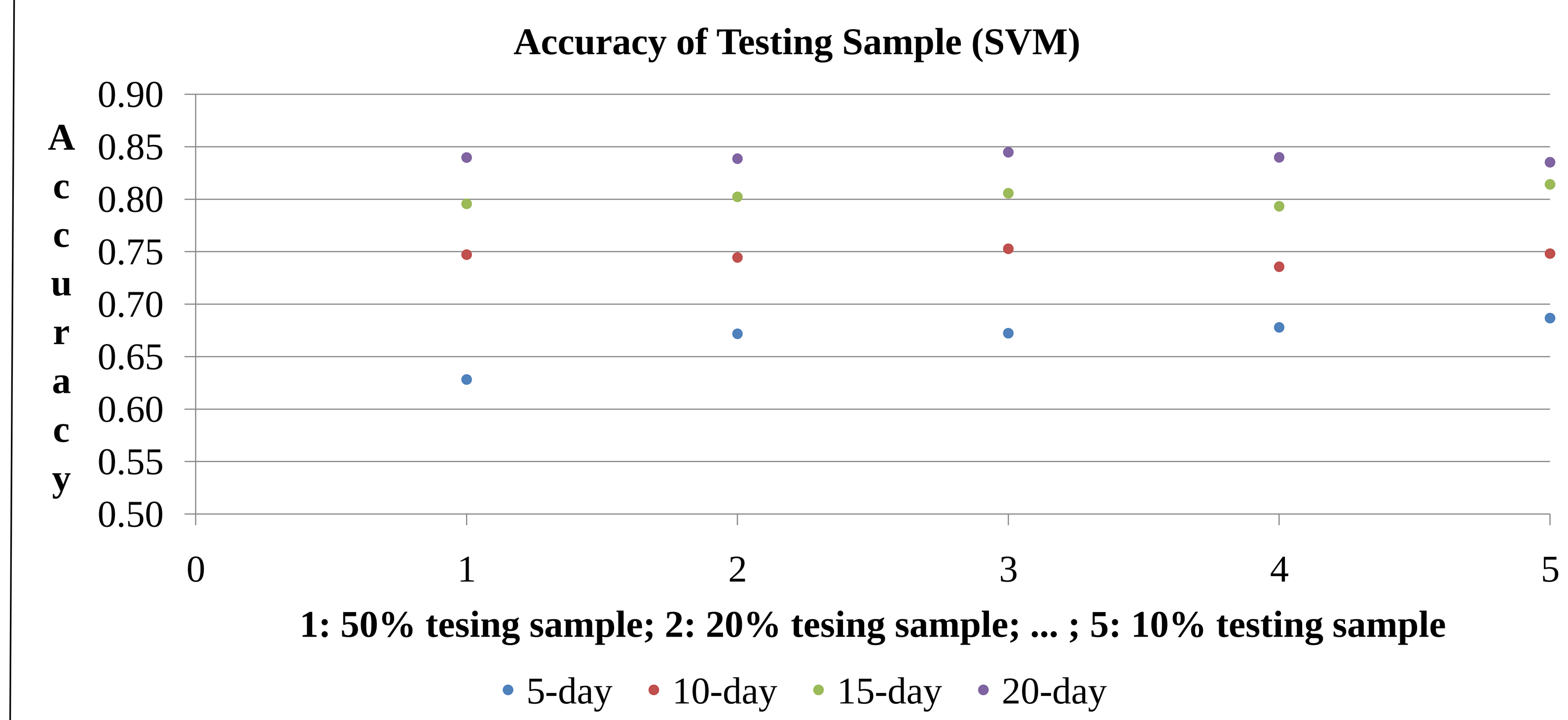
Model 2: Logistic Regression with Backward Selection

- From model 1, we observe that model built on sequentially selected sample is marginal better than the randomly selected sample. Thus, we will use the sequentially selected sample in the next 2 models.
- To avoid overfitting, we apply backward selection algorithm to control the number of selected features.
- The logistic regression model with backward selection is trained on the sample under same train/split logic and use the same features.



Model 3: SVM

- We continue to train SVM model on sequentially selected sample under the same train/test split logic and use the same features.



Conclusion

We find SVM generally outperforms logistic regression in predicting the future direction. The predictions for long term returns tend to be more accurate than short-term returns. Backward selection algorithm performs well on short term return models (i.e., the number of selected features shrinked), but perform poorly on long-term return models (i.e., the number of selected features does not shrink).

Future Work

The economic or financial relationship (i.e., positive or negative relationship) between corn future return and different features should be taken into consideration when building logistic regression model. Furthermore, SVM models with different kernels and ensemble methods should be explored to improve the testing sample accuracy.

Reference

Ticlavilca, Feuz, and McKee, Forecasting Agricultural Commodity Prices Using Multivariate Bayesian Machine Learning Regression, 2010.

Huang, D; Jiang, F; Tu, J; Zhou, G, Mean Reversion, Momentum and Return Predictability, 2013.

Kase, Cynthia A., How Well Do Traditional Momentum Indicators Work?, Kase and Company, Inc., CTA 2006.

R, https://cran.r-project.org/

SAS, http://www.sas.com/en_us/home.html

Scikit Learn, http://scikit-learn.org/

1. US corn has the largest agricultural futures market (by number of contract issued), and thus will be the primary focus.