Stock Market Machine Learning

Overview

Goal

To classify S&P 500 stocks as leading or lagging based on financial attributes

Solution

Convex and non-convex classification models optimized via stochastic gradient descent

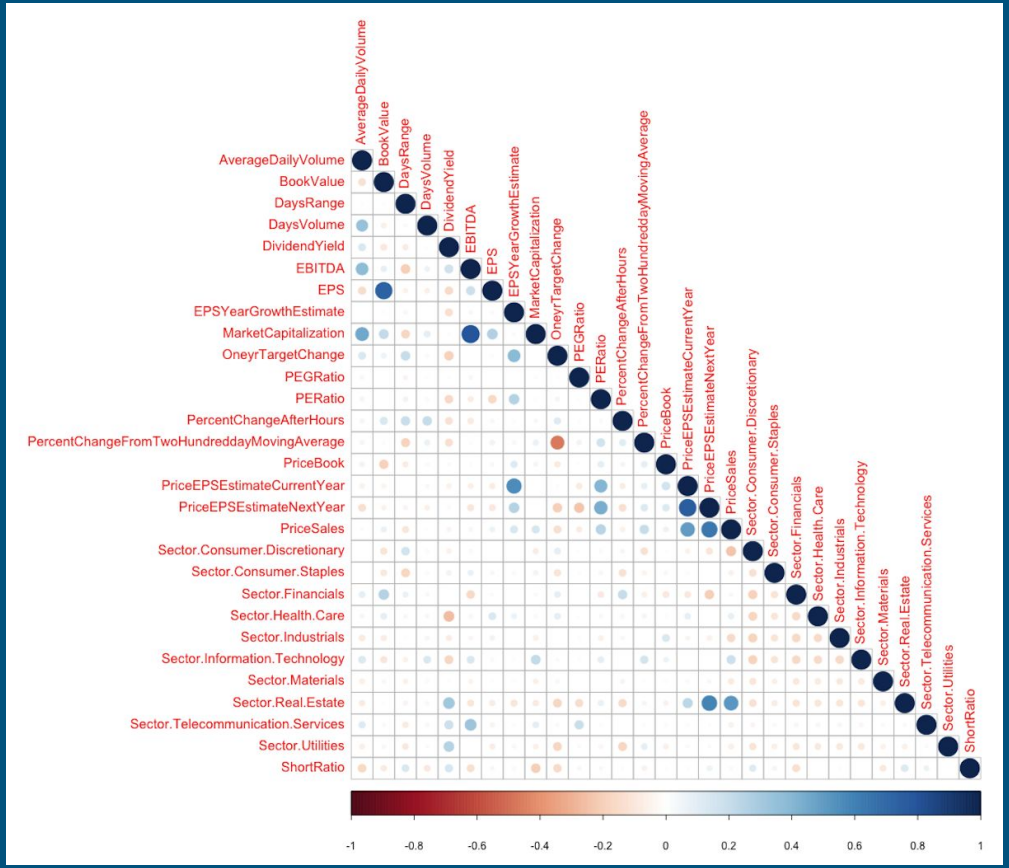
Infrastructure

- Yahoo Finance API
- Filtering & Creating Attributes
- Scaling Data

Dataset

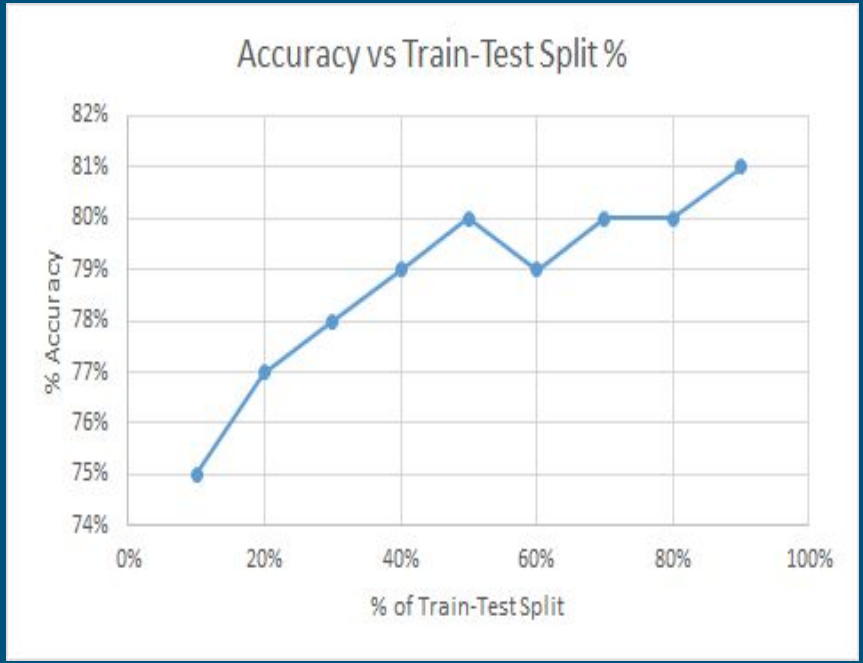
- 28 Attributes
- Common Financial Metrics
- %Change200MovingAvg Label
- Binary Labels (-1, 1)
- Scaled Features

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$



Cross Validation

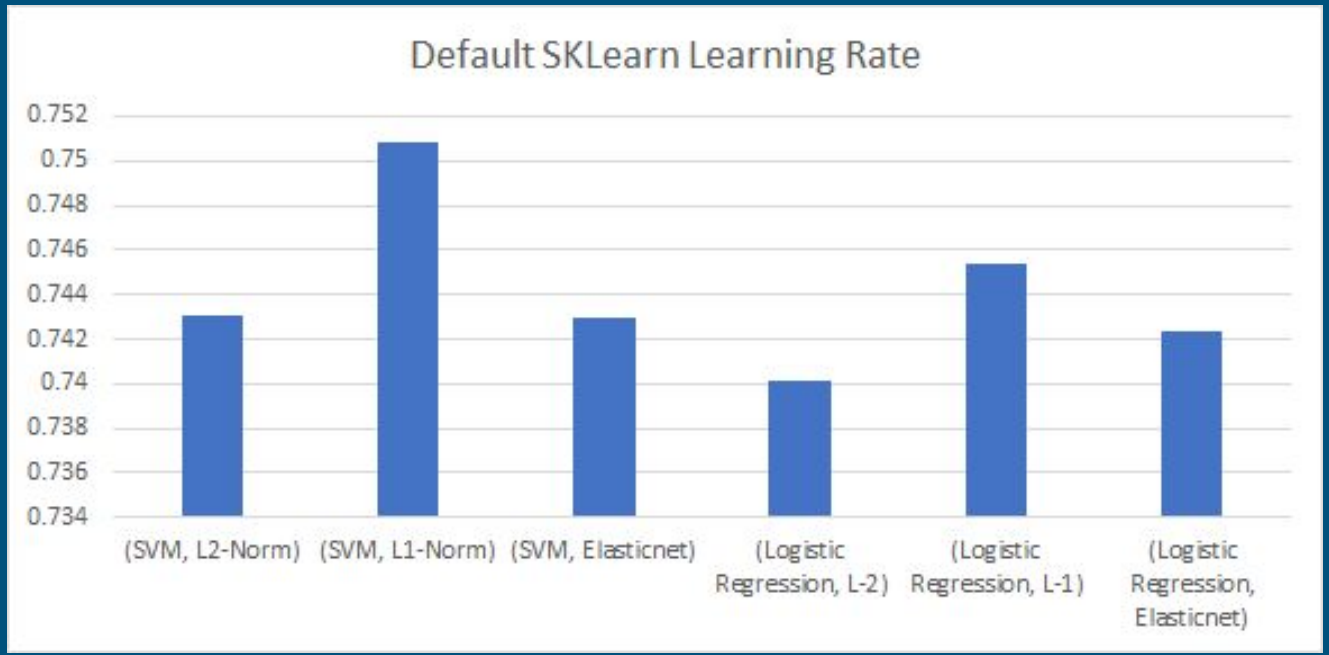
- 10 Fold Sampling
- Varied Train-Test Split
- Final models used:
  - 70% Training
  - 30% Testing



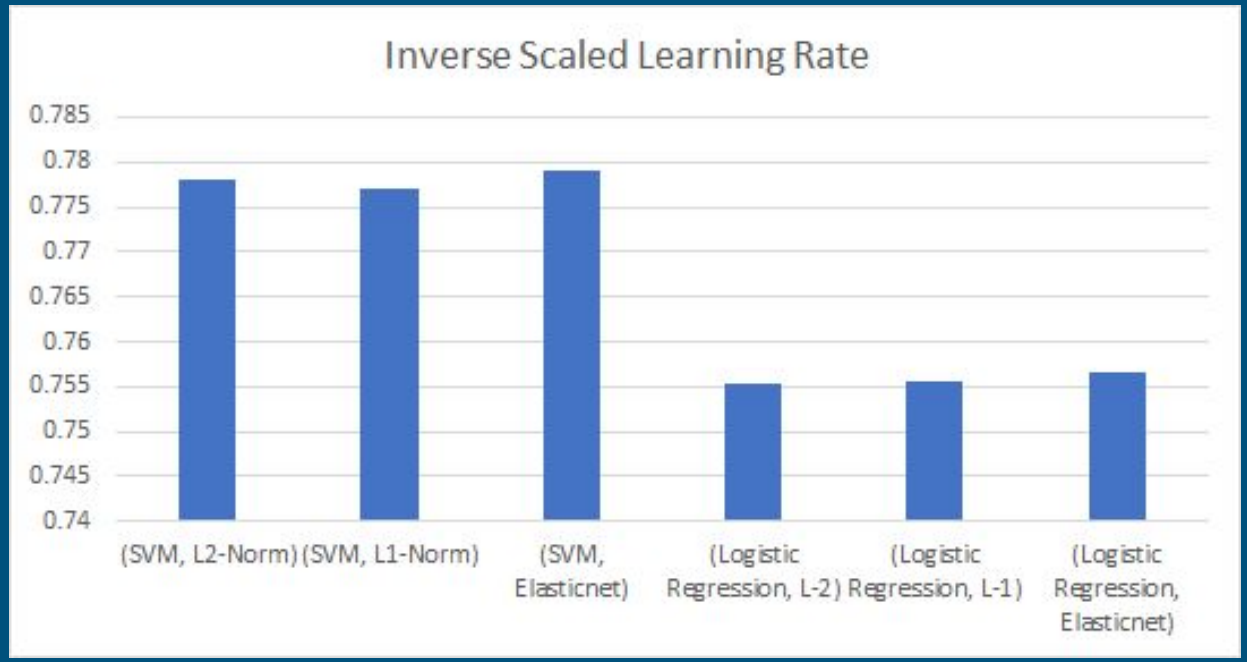
Convex Classification Models

- Optimized with Stochastic Gradient Descent
- Varied Learning Rates, Loss Functions, and Distance Functions
- Learning Rates: SKLearn Default, Constant, Inverse Scaling
- Loss Functions: Hinge (Soft Margin) Linear SVM, Logistic Regression
- Penalties: L1, L2, ElasticNet

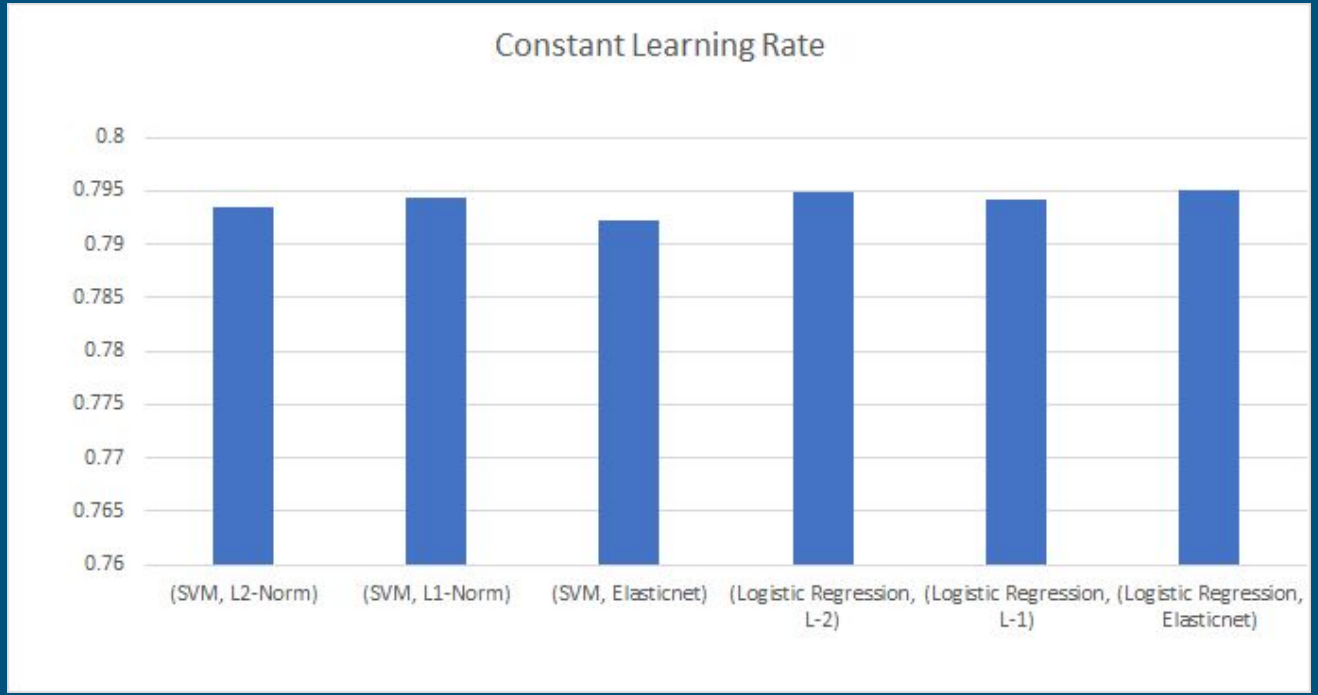
SVM and Logistic Regression Classification



Inverse Scaled Learning Rate



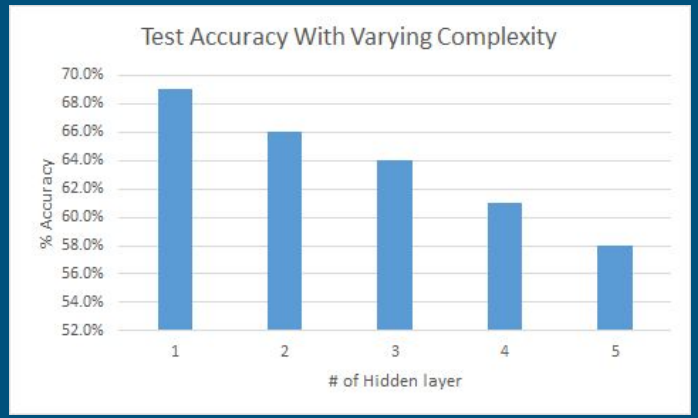
Constant Learning Rate



Non Convex Models

Neural Networks

- 28 Nodes input layer
- Varied hidden layers
  - 3 nodes per hidden layer



Naive Bayes

- Based on Bayes' theorem assuming iid among predictors.
- Find the probability of each observation being on a certain class
- Approximately 65% Accuracy

Conclusion

- We achieve 80% Classification Accuracy with SGD SVM
- Surpassed default SKLearn Accuracy by ~5% using calculated Lipschitz-based gamma
- Showed benefits of Convex Models (SVM) vs Non Convex Models (NN).

Thank you!