Workflow:

* + - Remove noisy features
  + How to do this?
    - Correlation matrix
      * More applicable to my PCA stuff
        + Before and after?
    - Decision tree?
      * Made 2, look decent, need to get feature names? Think I can just use lists of string
    - Pick “good” features
  + Select “good” features
    - Use decision tree, covariance matrix, correlation coefficients
    - USED PCA TRANSFORM, cheap easy and works
      * Something happened, no longer works as well? I hate my life
    - Normalize feature values
      * Centering
    - Did this, centered by subtracting mean
      * Scaling (variance or absolute)
    - Have option for variance or absolute scaling
    - Normalize example length
    - Training
    - Testing
    - Evaluation

Activation Functions (use ReLU)

- Data Preprocessing (center & normalize)

- Weight Initialization (use Xavier init)

- Batch Normalization (do it)

- Monitoring the Learning process

- Hyperparameter Optimization (random

sample hyperparams, in log space when

appropriate)

You will submit a 200 – 300 word abstract describing your problem, solution, and (expected) results. The abstract should provide enough information for a student in the course to understand the problem and your approach. The attached document provides tips and examples for writing abstracts. This milestone will be graded how accurately it reflects the material presented, clarity, and comprehensiveness.

1. Motivation or Statement of Problem: Why do we care about the problem? What practical, theoretical, scientific, or artistic gap is your research filling?

The prediction of asset prices in financial markets represents an important problem in finance. Accurately forecasting asset prices could yield numerous real-world benefits that extend beyond asset prices themselves; lower borrowing rates, higher lending rates, lower insurance costs, and higher investment returns.

Currently many research papers exist showing the utility of neural networks (of all types) to help with financial asset prediction. Using domain knowledge from a background in finance, it is the goal of the research is to show that mathematically combining existing features together can increase asset pricing accuracy.

This problem was approached using the proposed machine learning workflow from the lecture notes. A collection of daily stock pricing data for the S&P 500 from 1962 to present was gathered to be analyzed by a recurrent neural network. The data was then analyzed to determine its characteristics.

A covariance matrix and correlation coefficients showed many features contained similarities to one another. These similarities were removed using PCA to transform the data into data with almost no covariance between the features. After this processing, normalization occurred and then the normal routing of training, testing, and evaluation proceeded. Hyperparameters were decided using previous domain standards, monitoring and selection, and random sampling.

Analyses of both the original data, and data that has been selectively engineered are presented. The goal of the research is to show that the domain engineered data has a lower rate of error than the normal data.

It is the author’s opinion that the results are not conclusive. If the results definitively proved feature creation and transformation led to lower error rates, this would show that domain knowledge’s role in the transformation and engineering of features for solving financial asset pricing problems should be an area for further study.