A Few Experiments in Time Series Prediction

Alex Dara Infanger
alexinf@stanford.edu

Background & Introduction

Goal: We experiment with a few different methods for the prediction of financial time series. We compare the performance of a classical vector autoregression model with that of a deep fully connected network, and a long short-term memory (LSTM) based recurrent neural network (RNN).

Vector Autoregression

Consider the Gaussian vector autoregression of order *p*,

$$X_{n+1} = A_0 X_n + ... + A_p X_{n-p} + b + \epsilon_{n+1}$$

where the ϵ_j are i.i.d. Gaussian random vectors, the A_i are deterministic matrices and the X_n are random vectors. Here the *j*'th component of X_n contains the returns of stock *j* at time *n*. We show that we can reduce finding the maximum likelihood estimator for the above into a linear regression problem with mean squared error loss,

$$\mathcal{J}_{\text{regularized}} = m^{-1} \sum_{i=1}^{m} ||WX^{(i)} + b - Y^{(i)}||_2^2 + \lambda ||W||_2^2.$$

The correspondence holds for $\lambda=0$, but we also experiment with $\lambda>0$ to test if regularization helps.

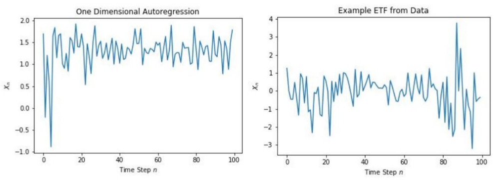

Figure 1: Simulated Autoregression and ETF from Data.

References

[1] Data from *Bloomberg, 2012.* Subscription service. Available for academic use through Stanford University libraries.

Deep Models

Fully Connected Network

We also implement a three layer deep fully connected network. The former two layers have relu activations and the last layer has a linear activation, to allow for the output to be negative.

LSTM Models

We implement two LSTM based RNNs. The first LSTM takes *p* vectors *X_j* as input and outputs *p* vectors, which are fed into a fully connected linear layer. We use mean squared error loss with dropout regularization (probability *q=.5*). We call this LSTM1. The second is the a three layer LSTM model, which is LSTM1's LSTM repeated three times before a linear layer.

PCA Based Preprocessing

Recall the low-rank approximation of a data matrix *D*,

$$D = VW^T + \epsilon.$$

We use this to do (1) signal prediction (predict the columns of *V* instead of that of *D,* where column *j* of *D* is the time series of ETF *j*), and (2) market signal subtraction (remove the market signal from the data).

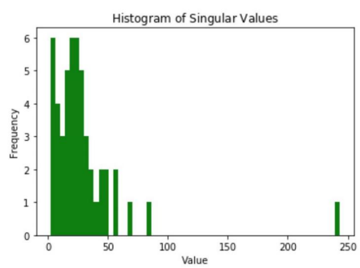

Figure 2: Histogram of Singular Values.

Results

All models perform well on a simulated data set when the noise parameter ϵ_i is sufficiently small. Regularized linear regression performs the best on the original dataset. LSTM1 performs best and most robustly on the market signal prediction problem (post preprocessing).

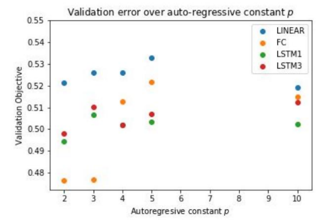

Figure 3: Loss as a function of regression constant p.

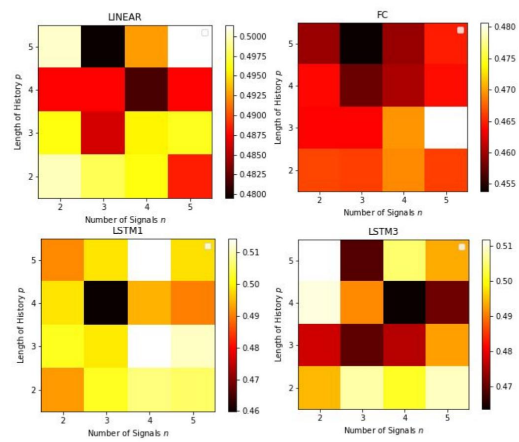

Figure 4: Validation loss as a function of signals and regression constant p.