NYU Machine Learning
Homework 3

Dr. George A. Lentzas

This Homework is due by 5pm on Tuesday December 17th. Sumbit your results in .pdf format in Classes. Include your code in the Appendix. Good Luck!

Install the package `rugarch` in an R session and import the dataset `dji30ret` (showing the Dow Jones 30 Constituents daily log returns from March 16 1987 to February 3 2009) into a data frame. Split your sample into a Training Set: from March 16, 1987 to December 31, 2002 and a Test Set from January 1st 2003 to February 3, 2009. Install Keras and Tensorflow to your computer and make sure you can use it to estimate Neural Networks.

All code for this Homework should be in Keras and should be included in the Appendix.

For each stock you will try to predict log return at time $t+1$ using log returns from time t to $t-30$. Accordingly create a data frame that includes the target and explanatory variables for all stocks as well as dummy variables for each stock.

1. Using the Training Set, fit an a 30-20 Neural Network (i.e. 2 hidden layers, the first with 30 units and the second with 20 units) to predict the $t+1$ individual stock return. In this part of the exercise do not use Dropout, instead use a validation set and choose an early stopping in terms of the number of epochs. Choose the optimization algorithm you think is appropriate and a mini-batch size of 20. Report the mean absolute error for the test set.

   [20 Points]

2. Using the Training Set, fit a 15-10-5 Neural Network to predict the $t+1$ individual stock return. This time use a Dropout rate of 0.3 and no validation. Report the mean absolute error for the test set.

   [20 Points]

3. Using the Training Set, fit 15-10-5 Neural Network to predict the $t+1$ individual stock return. This time do not use Dropout but instead use L2 reguralization. Use cross-validation to choose the appropriate reguralization parameter. Explain how you performed cross validation to avoid leakage. Report the mean absolute error for the test set.

   [30 Points]

4. Using the Training Set, fit 15-10-5 Neural Network to predict the $t+1$ individual stock up/down move (1 for positive return and 0 for negative return). Again use a Dropout rate of 0.3 and no validation. Report the classification accuracy for the test set. (Note that this is a classification problem.)

   [20 Points]

5. Compare the mean absolute error of the different estimated models on the Test Set. What are your conclusions?

   [10 Points]