NYU Machine Learning
Homework 2

Dr. George A. Lentzas

This Homework is due by 11:59pm on Tuesday Nov 12th. 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.

1. For each stock you will try to predict log return at time $t+1$ using log returns from time t to $t-90$. Accordingly create a data frame that includes the target and explanatory variables for all stocks as well as dummy variables for each stock. Print the `summary()` and `head()` of this data frame. Explain why it makes sens to use one model for all stocks instead of 30 individual models.

   [10 Points]

2. Using the Training Set, fit an SVM with a radial kernel to predict the $t+1$ individual stock return. Use cross-validation to select the relevant parameters.

   [15 Points]

3. Using the Training Set, fit a Random Forest to predict the $t+1$ individual stock return.

   [15 Points]

4. Using the Training Set, fit a Boosted Tree to predict the $t+1$ individual stock return. Use cross-validation to select the relevant parameters.

   [15 Points]

5. Explain in detail how you performed cross-validation above, including how you ensured that there is no look ahead bias and how you chose the number of folds.

   [15 Points]

6. Compare the MSE of the different estimated models on the Test Set (do not refit after December 31, 2002). What are your conclusions?

   [15 Points]

7. Use the best performing model above to construct portfolios on the Test Set (do not refit after December 31, 2002). Specifically, at each time period rank the forecasted returns and create a long/short portfolio comprising of the top 5 and bottom 5 forecasted returns. You should assume that transaction costs are 5bps of traded dollar amount and that at each period you allocate 1/10 of your capital to each of the 5 top and 5 bottom forecasts. Your starting capital is $1,000,000$ Dollars, how does your strategy perform?

   [15 Points]