Mini Project (15%)

Portfolio Return Prediction

In this mini-project, you are going to construct a model that allow you to pick the most profitable portfolio in the following week. That is, based on the data available at time t, we want to decide which of the 30 portfolios will generate the highest return. The training dataset is the ldpa30_train.csv. The input features are beta_mkt, beta_hml, beta_smb, and sigma. Our target variable is the return in t+1, week_return1.

You may split the data in ldpa30_train.csv into 5-fold or 10-fold so that you can run your own testing. After you have decided the best model, use the data in ldpa30_test_blind to predict the best portfolio. The input features, including beta_mkt, beta_hml, beta_smb, and sigma can be found in the ldpa30_test_blind. Note that there are no week_return1 column in the ldpa30_test_blind file.

You should submit a short report (1-2 pages) explaining how you construct your model and what is the best model you used to predict portfolio return. Submit your report, together with your program and output files to the TA. Your output file should be in csv format. The first column of your csv file should be the week index and the second column should be the portfolio number that is predicted to have the highest return.

For example, if you predict the winners of week 371 to 375 are portfolio 2, 18, 29, 1, 5, then the output file should look like:

week_index,group

371,2

372,18

373,29

374,1

375,5

Your output will be used to compute a loss function, which is defined as the sum of the gaps between the best-performing portfolio to the selected portfolio. That is, if the best-performing portfolio at week i has a return of 1.21 and the portfolio of your choice has a return of 1.1, then the loss for week i is 1.21 – 1.1 = 0.1. The sum of the loss across all testing weeks is your overall performance. Your score to this mini-project is linked to the relative performance of all mini-project participants.