FA 690: Machine Learning in Finance
Portfolio Selection and Time Series
Due Tuesday, October 4 3PM Eastern Time

Introduction

For your first assignment for the semester, you will work *individually* to use machine learning to build a tool to do predictive analytics or portfolio optimization on financial data. Broadly your projects will consist of three inter-related topics:

1. accessing and discussing applicable financial data;
2. selecting an appropriate machine learning method; and
3. results and analysis.

You will get the most out of the project if you interact with Professor Feinstein during this assignment, especially when planning a topic.

Project Components

Data Collection and Discussion

Download financial data from the database of your choice (e.g., Bloomberg, WRDS, *Yahoo Finance*, ...). You should spend time deciding on the appropriate data to analyze and whether it is sufficient for the method you will want to implement.

Machine Learning Methods

Given the data you have collected, choose a machine learning method to analyze your data. This can be as simple as a linear regression, but could also be a recurrent neural network (for instance). Give serious thought to your proposed method as you will need to justify your choice.

Results and Analysis

Implement your methodology on your collected data in order to test the results. You may want to compare your chosen methodology to a simple baseline in order to determine performance. You should remark on whether your methodology appears suitable to answer the desired question; statistical analysis is *strongly* encouraged.

Report Details

You will submit your final write-up, which should include all of the information detailed below. This should be presented in roughly the order given, but your write-up need not have corresponding sections or bullet points. The write-up should be about 5-7 double-spaced pages, Times New Roman 12pt font. This does not include any appendices (of, e.g., your *Jupyter Notebook*) you may wish to include. Any external resources used should have clear citations and a reference page at the end of your work. This report **must** be submitted in pdf format; your code may be requested if not clear in the document so please keep that available.

1. **Overview** of the problem statement.
2. Detailed description of the **data collected** and why it is appropriate for the problem being considered. Mention any data cleaning if required.
3. Detailed description of the **machine learning method** and why it is appropriate for the problem being considered. If comparison to a *baseline* model is to be studied, provide the details of this methodology as well.
4. Describe the **results** obtained by your methodology on the data. Analyze these results to provide a recommendation.
5. **Next steps**: What do you recommend as a result of your analysis? Do you suggest attempting different algorithms or a larger test or more data? etc... What else could be done with the problem, but time did not permit?

Project Presentation

On Tuesday, October 4, seven (7) [subject to change based on actual enrollment] of your classmates will give **15 minute presentations** (not including questions). Slides must be uploaded to Canvas prior to presentations and are due by 3:30PM on Tuesday, October 4. If you are a presenter and have to attend remotely, you will need to upload a recording of your presentation to Canvas as well in case of internet connection errors. The class time on October 4 will be used for *live* presentations. Selection of presenters will be announced on Canvas on Tuesday, September 20. If you strongly prefer to present (or not present) in this first group, please contact me as soon as possible so that I can try to accommodate any request. I cannot guarantee all requests will be granted. (All students must present *once* this semester, so if you do not present this first time, you will need to for either Assignment 2 and 3.) Beyond any such requests, selection will be made through a random number generator from the official course list.