CF969-7-SP-CO
Big Data for Computational Finance

Please refer to the Student's handbook on the School's Policy on Plagiarism and Late Submission.

All deliverables below must be uploaded on FASER by the deadline as independent items, i.e., not bundled together in a zip file.

Part I (55%): Report on Machine Learning in Finance

You are asked to write a report on **one (1)** recent research paper in applications of machine learning for computational finance from the following list:

- "Learning to simulate realistic limit order book markets from data as a World Agent" by A. Coletta, A. Moulin, S. Vyetrenko, and T. Balch.
  Available at: https://arxiv.org/abs/2210.09897
- "FinRL: Deep Reinforcement Learning Framework to Automate Trading in Quantitative Finance" by X.-Y.Liu et al.
  Available at: https://arxiv.org/abs/2111.09395
- "Deep Learning Statistical Arbitrage" by J. Guijjaro-Ordonez, M. Pelger, and G. Zanotti.
  Available at: https://arxiv.org/abs/2106.04028
- "Trading via Selective Classification" by N. Chalkidis and R. Savani.
  Available at: https://arxiv.org/pdf/2110.14914.pdf
- "Trading with the Momentum Transformer: An Intelligent and Interpretable Architecture" by K. Wood, S. Giegerich, S. Roberts, and S. Zohren.
  Available at: https://arxiv.org/abs/2112.08534
- "Deep learning with long short-term memory networks for financial market predictions" by T. Fischer and C. Krauss.
  Available at: https://www.econstor.eu/bitstream/10419/157808/1/886576210.pdf
- "A deep learning framework for financial time series using stacked autoencoders and long-short term memory" by W. Bao, J. Yue , and Y. Rao.
  Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0180944
- "Forecasting CPI Inflation Components with Hierarchical Recurrent Neural Networks" by O. Barkan et al.
  Available at: https://scholar.harvard.edu/files/jbenchimol/files/forecasting-cpi-inflation-components-hrnn.pdf
- "Event prediction within directional change framework using a CNN-LSTM model" by A. Rostamian and J. O'Hara.
  Available at: https://repository.essex.ac.uk/33313/1/Published_Version.pdf
- "Ascertaining price formation in cryptocurrency markets with Deep Learning" by F. Fang, W. Chung, C. Ventre, M. Basios, Leslie Kanthan, L. Lid, and F. Wu.
  Available at https://arxiv.org/abs/2003.00803
- "FinBERT: A Pre-trained Financial Language Representation Model for Financial Text Mining" by Z. Liu, D. Huang, K. Huang, Z. Li, and J. Zhao.
  Available at: https://www.ijcai.org/proceedings/2020/622
- "The Efficient Hedging Frontier with Deep Neural Networks" by Z. Gong, C. Ventre, and J. O'Hara.
  Available at: https://arxiv.org/pdf/2104.05280.pdf

- "Learning about risk: Machine learning for risk assessment" by Nicola Paltrinieri, Louise Comfort, and Genserik Reniers.
  Available at: https://www.sciencedirect.com/science/article/pii/S0925753518311184
- "Short-term bitcoin market prediction via machine learning" by P. Jaquart, D. Dann, and C. Weinhardt.
  Available at: https://www.sciencedirect.com/science/article/pii/S2405918821000027

A report of **at most 1500 words** must be written. The report should both **summarise** and **evaluate** the article. As a guideline, a good report should address satisfactorily the following questions, namely:

- What is the paper about (i.e., what is the topic)?
- How do the authors approach the problem? I.e., what is the method they use?
- What are the results?
- What are strong points in the paper, in your view?
- What are weak points in the paper, in your view?

The report will be assessed attending factors such as its contents and the connection to the key questions outlined above, presentation, organisation, clarity, explanations, references, etc.

Deliverable for Part I: The report as a pdf file

Part II (45%): Optimisation and Machine Learning in Finance – Software

Part IIA (20% of the total mark)

Consider a scenario where an investor has £20,000 to invest in a combination of the following:

- Stock XYZ sells today at £20 per share.
- A European call option to buy 100 shares of stock XYZ at £15 per share exactly six months from today sells for £1,000.
- The investor can also raise additional funds which can be immediately invested, if desired, by selling call options with the above characteristics.
- In addition, a 6-month riskless zero-coupon bond with £100 face value sells for £90.

The investor has decided to limit the number of call options that they buy or sell to at most 50.

The investor considers three scenarios for the price of stock XYZ six months from today: the price will be the same as today, the price will go up to £40, or drop to £12. The investor's best estimate is that each of these scenarios is equally likely.

(13% of the total mark) **Formulate** and **solve** a linear program to determine the portfolio of stocks, bonds, and options that maximises expected profit.

(7%) of the total mark) Suppose that the investor wants a profit of at least £2,000 in any of the three scenarios for the price of XYZ six months from today. **Formulate** and **solve** a linear program that will maximise the investor's expected profit under this additional constraint.

You can use any solver (e.g., Excel solver, Gurobi, gurobipy, linprog) for solving the linear programs.

Deliverables for Part IIA:

- A document explaining the reasoning behind the model formulation as well as what is the answer and its interpretation. It should be either a pdf file or a Jupyter notebook file.
- The source code. E.g., an Excel file in case you are using the Excel solver, a Jupyter notebook file in case you are using linprog or gurobipy, the Gurobi file in case you are using Gurobi.

Part IIB (25% of the total mark)

Consider the .csv file available on Moodle. It contains 1700 observations of 26 financial and accounting metrics for a set of firms in several different industries. For each observation, the last column denotes the rating according to Moody's, while the second-to-last column denotes whether the assets are of investment grade or not; ratings in the set {Aaa, Aa1, Aa2, Aa3, A1, A2, A3, Baa1, Baa2, Baa3} are in an investment grade.

The task is to classify a firm's credit rating into one of the 16 categories and to predict whether a firm is considered investment grade or not. To do so, you should split the dataset in a training set and a test set in a 80%:20% ratio. You are asked to implement

- a linear regression approach with Ridge (or L1) and Lasso (or L2) regularisation to predict whether a firm is in an investment grade or not
- a logistic regression approach with Ridge (or L1) and Lasso (or L2) regularisation to predict whether a firm is in an investment grade or not
- a Neural Networks based approach to classify the firm's rating into one of the rating categories and predict if it is in an investment grade

and discuss how and why were the parameters selected at each model and also discuss what the results demonstrate for the effectiveness and suitability of each approach on this problem.

Any language/software we covered during the modules is acceptable; you are also welcome to use a different one that you might have learnt independently or in some other module.

Your submission will be assessed attending factors such as contents, clarity, explanations, etc. of the Jupyter-like notebook and correctness, techniques and style of the software.

Deliverables for Part IIB:

- The source code (and any executables, if applicable) of your software in Part IIB
- A document presenting and discussing the selection of parameters as well as the results. It should be either a pdf file or a Jupyter notebook file.