TU-E2231 Machine Learning in Financial Risk Management
Spring 2025
Ruth Kaila and Eljas Toepfer

Instructions for Exercise 2

1. This exercise is graded passed/failed. The early-bird deadline is Friday, **7.2.2025**. The final deadline is **16.4.2025**.
   To pass the course, the student must complete 6 weekly exercises. However, if the student returns each exercise within 10 days of the date the exercise was given, 5 exercises are sufficient to pass the course.
2. Return 2 files:
   1. A PDF file with
      I. A: Individual work (one page)
      II. B: answers to the questions in the Python files
      III. C: screenshots of two figures generated in Python file 2C related to pairs trading: 'Pair's spread' and 'Pair Trading - Trading Signals and Position'
   2. A PDF file of the whole Python files. (or 3 separate) (see the instructions at the end for guidance on converting Python files to PDF)

A. Written exercise (Individual work)

Write a total of one page (400-500 words) on the following 2 topics:

1. What are the specific problems of applying machine learning to financial data? Why would it be difficult to beat the market with available data and some good algo? What other challenges can you think of?
   Use lecture 2 and your own expertise and then search google; avoid asking ChatGPT because I already know what it says. Using ChatGPT would be a waste of both our time. Once you have answered, you can look what ChatGPT says.
2. What is z-score? What are the advantages and disadvantages of using z-scores when the data are not normally distributed?

B. Python exercises 2

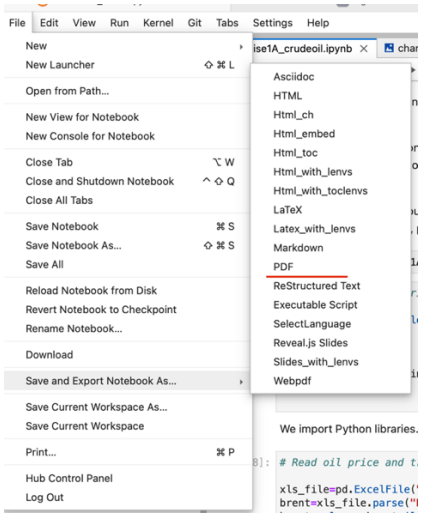
- Exercises 2A on data structures and 2B on labeling are based on Lecture 2 and dePrado's book. In these exercises, the code is given. You should use the same data on S&P500 index in both exercises (given in the folder Data). You should read through the code, make sure you understand everything and answer questions in a separate paper. The code is given to you so that you have the energy to play with it. You can change the parameters where possible and see how it affects the results/charts. This will improve your understanding of the theory presented.
  - IMPORTANT: In the end, task 3 is for those who do the 5/6 credit version of the course.

- Exercise 2C is about pair trading. Read the code, answer the questions, and complete simple pieces of code. Make sure that you understand the logic of taking long and short positions.

- 5/6 credit version of the course: Select a specific sector with several companies. Select two companies that could be used in a paired trade. Find suitable data and repeat exercise 2C. Then explain why you chose this specific sector, how you selected the companies for pair trading, how you selected and prepared the data and what the results of your trading experiment were.

IMPORTANT: Calibrate your answer to your level. Focus on what you think is essential.

How to save the notebook as a pdf file:



Give screenshots of these pictures: 'Pair's spread' and 'Pair Trading - Trading Signals and Position'