Machine Learning in Finance II part

Prerequisites:

Knowledge of 'Machine Learning in Finance I', 'Time Series Analysis', and 'Reproducible Research'.

Short description

The course covers more advanced methods of machine learning: boosting models, neural networks, and Bayesian time series models. Both theoretical background and practical empirical applications in finance are discussed. Practical part covers problems of regression and classification problems, processing and forecasting of sequences, time-series analysis and deployment of methods in the cloud environment.

Full description

The course consists of 3 chapters divided according to the class of presented algorithms: 1) boosting models 2) (deep) neural network models 3) bayesian time series models. It is conducted in the form of interactive laboratories with the use of case studies which are carried out in parallel with the lecture part.

Chapter 0. One class repeating the most important modeling techniques from the Machine Learning in Finance I course

Chapter 1. Boosting models

1. AdaBoost
2. Gradient Boosting
3. eXtreme Gradient Boosting
4. Light Gradient Boosting Machine
5. CatBoost

Chapter 2. Neural network models

1. Multilayer Perceptrons
2. Recurrent Neural Network
3. Convolution Neural Network*
4. Attention mechanism in Neural Network

Chapter 3. Foundation Models for Tabular and Time Series Data

1. Tabular data models e.g. TabPFN
2. Time series data models e.g. Chronos, Moirai-MoE, TimesFM

Chapter 4. Ensembling methods*

Bibliography:

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An Introduction to Statistical Learning. Springer, New York, NY
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.
- Chollet, F. (2017). Deep learning with Python. Manning Publications.
- Stevens, E., Antiga, L., & Viehmann, T. (2020). Deep learning with PyTorch. Manning Publications.
- Intel (2018). Deep learning. Retrieved from https://www.intel.com/content/www/us/en/developer/learn/course-deep-learning.html
- Intel (2018). Time-Series Analysis. Retrieved from https://www.intel.com/content/www/us/en/developer/learn/course-time-series-analysis.html

Along with additional literature assigned to the case studies.

Learning Outcomes:

After completing the course, the students will have structured and reliable knowledge on boosting models, neural networks, and Bayesian time series models. They will be able to apply them for both regression and classification problems. They will know the theoretical foundations of these algorithms, as well as have programming skills allowing them to deploy the models in practice, also in the cloud framework. They will also know how to interpret results and explain how they work to other non-technical people.

K_W01, K_U01, K_U02, K_U03, K_U04, K_U05, KS_01.

Assessment methods and assessment criteria:

The final grade consists of three elements:

1. **Theoretical Exam** – 10 open-ended questions (34%, mid-December/January).

2. **Machine Learning Project** – prepared in pairs, with an extended report in a Python notebook. The report should include code blocks that allow full reproduction of the analysis. The dataset is selected by the students (subject to tutor approval, e.g., from Kaggle).
    - Initial submission: mid-December
    - Final submission: mid-January
    - Project code will be reviewed with each pair of students during the last two classes.

    (33%)
3. **Pop Quizzes** – about 4 per semester (33%).

To pass, a minimum of **50% in each element** is required.