Machine Learning II

Module 5, 2021-2022

Ilya Munerman
Munerman & Partners, Interfax LAB
ivm@munerman.ru

Course description

This course describes how dramatically changes of the information market provides new power tools for financial data processing and analysis. In addition, we will compare classical and contemporary tools performance. The main aim of this course is a species of different country data sets and technics for integration this data for common international data environment.

It also considers the features of algorithms for use online data for the instant response of the model to changing environmental circumstances, taken into the consequences of a pandemic.

The course is designed for listeners which known elementary economics, finances, IT and mathematics and may be able for economists, IT specialists, managers, include MBA and journalists.

Course requirements, grading, and attendance policies

Statistics, mathematics, corporate finance, assets valuation. The course grade is based on four home assignments (30%), case discussions (20%), and final exam (50%).

Course contents

1. Contemporary financial analyses main challenges. Ontology and its role in the quantitative analyses. Main categories: target variables, sustainability, dynamic. Understanding a difference between pattern recognition and prediction technics. Data quality and cognitive biases. Expert's professional crisis caused by data revolution. Neuromorphic computing vs cognitive biases.

2. Big, data mining, data science. Why general universe processing most powerful, then sample analysis? What is learning sample today? Data source survey. Big data infrastructure. Low and ethical problem using big data. Data gathering, storage and retrieval instruments. Cloud technologies. Data base management, API, BI, ERP, final processing systems. Providers survey. From state to startups. Blockchain technology.

3. Main data types. Quote data for technical and fundamental stock market analysis. Labor data. Real estate data. Procurement and Contracts electronic trading platforms data. B2B and other suggestion services. Procurement efficiency analysis. Legal and court information. Court cases liabilities estimation algorithm and prediction. Web-sales and traffic data. Providers, aggregators, efficient and inefficient solutions. Open data hubs: pro et contra. Different solution costs. Credit scores and paydex. Transaction data and cash boxes on-line date. International trade data. Trademarks and intellectual properties. Credit histories.

4. Correlation. Correlation vs causalities. Metrics minding. Linear and nonlinear dependencies. Copulas.

4.1. Practice. Theory and Python implementation. Base statistical issues. Correlation calculation. MSE, MAE, Accuracy, Confusion matrix.

5. Data processing. Modeling. Recognition, prediction and other forms of data processing. From regression to a neural network. Data mining. Cluster analyses and dendrograms. Stochastic processes. Probability vs reliability. Classical ML – regression and clusterization, supervised and unsupervised learning, fuzzy logic and fuzzy c-means, SVM, logistic regression, PCA and another dimension reduction methods, CART, naive Bayes, etc. Ensembles methods – bagging (include random forests), boosting and stacking. Reinforcement learning, genetic algorithm, reinforcement learning for scorings and decision making. Neural networks MLP, GRNN, CNN, RBF etc. Natural language processing.

6. Estimation and model testing. Statistical tests and criterions: R2, ROC-curve, type I and type II errors, graphical analyses. Accuracy, precision, recall, f1-score and other metrics. Ex-post testing. Weight of Evidence (WoE), Information Value (IV), PI, Max profit and other business useful metrics. MPP – model performance predictor.

6.1. Practice. Theory + visualization + Python implementation
Common statistical metrics, Advanced statistical metrics, Business metrics, Model overfitting, Comparison of different metrics on identical samples

7. Interpretable models and suggestion systems. Black and gray boxes. Usage of NLP-pipeline for scoring models explanation. LIME, SHAP, Eli5 etc. Suggestion systems algorithms. Collaborative filtering, content and knowledge based and hybrid systems. ALS and SELF algorithms.

8. Applications. Scorings, rankings, ratings. Tools for modern institutes: crowdsourcing, start-ups, franchising and so on. Social score - pro and contra. Main parameters of credit scores: one year, life time PD, LGD, EAD and their calculation ML algorithms. Credit conveyors. Value of credit and collateral value. Credit value monitoring according to regulation and standards. Basel III and modelling risk estimation. Real estate apps and valuation. Ad valor tax rates. Real estate databases. State and corporate real estate data hubs. Collecting data about all types of procurement: government, commercial, international, planned purchases. Bankruptcy analysis. News proceedings. Data standards problem. Auction houses data. AML and SAR, transaction fraud. B2B and investment roboadvisers. Fraud, failure and delinquency scores. Bankruptcy prediction without corporate finance. Open societies, economic forecast, parameters estimation applications and so on. Visualization, infographics and data journalism. Importance, examples, tools, and applications. Contests and hackathons for provide original salvation. Application for legal tech, prop tech, reg tech, ag tech, health tech etc.

9. Modern challenges. Stock market and cryptocurrency roboadvisors. Investor services development: relevant news delivering, investors profiling, personal investment ideas for the current investor, portfolio optimization with the Black-Litterman model and other methods. COVID model adjustment and priority data for express online analysis in a crisis. Methods for analyzing companies under sanction without quarterly financial reports and other usual data. lowCode and NoCode solution, ML Ops and AutoML.

Course materials

Required textbooks and materials

1. James G., Witten D., Hastie T., Tibshiriani R. (2015) An introduction to statistical learning with applications in R, 6th edition, Springer
2. 2. Brooks C. (2014) Introductory Econometrics for Finance, Third Edition, Cambridge University Press.
3. Ian Goodfellow, Yoshua Bengio and Aaron Courville, Deep Learning, An MIT Press book (2016)
4. Peter Flach, Machine Learning: The Art and Science of Algorithms that Make Sense of Data Cambridge University Press; (2012)
5. IFRS 9 and CECL Credit Risk Modelling and Validation: A Practical Guide with Examples Worked in R and SAS by Tiziano Bellini, Academic Press (2019)

6. **https://ru.coursera.org/learn/machine-learning**

Additional materials

1. https://www.kdnuggets.com/2017/04/10-free-must-read-books-machine-learning-data-science.html
https://www.kdnuggets.com/2018/05/10-more-free-must-read-books-for-machine-learning-and-data-science.html
https://www.kdnuggets.com/2018/11/10-free-must-see-courses-machine-learning-data-science.html
https://www.kdnuggets.com/2018/12/10-more-free-must-see-courses-machine-learning-data-science.html
2. Special course materials will be provided in cloud storage

Academic integrity policy

Cheating, plagiarism, and any other violations of academic ethics at NES are not tolerated.