Prediction with Expert Advice for Economics and Finance

Background: On-line Learning and Prediction with Expert Advice

This project belongs to the area of on-line machine learning.

Traditionally machine learning has been developed within the batch paradigm. In this paradigm, a learner is presented with a data sample (called a training set) and needs to find a pattern of a certain kind. The pattern is then applied for prediction on unseen test data.

While the approach of batch learning fits well with the traditional scientific methodology, the on-line learning paradigm is motivated by the development of information technologies and increased speed of information processing. In this paradigm, small chunks of data arrive one by one and the learner must draw insights about the behaviour of the data sequence, obtain predictions for the future, and make decisions (such as investment decisions) based on the available insights. An important feature of this approach is the feedback loop: the learner makes predictions or takes actions and then observes the outcome produced by the environment. On the basis of this output, the learner modifies her insights and makes new predictions/actions, which again confront the environment.

The attitude we take is pragmatic. We do not aim at identifying a "correct" model in any sense but rather at making good quality predictions. This implies a specific approach to the generalisation problem and model selection.

In machine learning, generalisation problem plays a very important role (Hastie, Friedman and Tibshirani, 2017). A model performing well on the training data does not necessarily maintain the same level of performance on the test data because it may overfit the training set. In practice, the standard approach is to use regularisation in the model, e.g., penalise the growth of coefficients in regression or SVM, complexity of a decision tree etc. Theoretical justification of such methods requires assumptions on the nature of the data. If the training and test data follow this assumption, the method can be shown to perform well. For example, if the data following a Gaussian random field with the covariance given by a known kernel then ridge regression produces best predictions on average (Rasmussen and Williams, 2006). In practical situations, one cannot expect the requirements to hold.

A direct counterpart of regularisation for time series data is the use of AIC and BIC in the selection of an ARIMA model (Hyndman and Athanasopoulos, 2018, Section 8.6). Theoretical results on the consistency of these criteria require that the data indeed follows an operating model (e.g., Karagrigoriou et al, 2011).

In on-line learning, making assumptions on the data becomes particularly problematic as the laws governing the data may change with time. One can address this through a complete retraining, but this approach is costly, time-consuming, and hard to analyse at the theoretical level. The methods of prediction with expert advice learning provide a helpful alternative. Instead of picking one model from a pool, one can take *all* models and merge their predictions using methods of prediction with expert advice. The methods come with explicit theoretical guarantees, which assure that in terms of cumulative performance we are not much worse off than the best model from the pool (finite or infinite) regardless of the nature of the data. As long as our pool of models is sufficiently rich for the data, this leads to good performance. In intuitive terms, prediction with expert advice makes sure the learner re-aligns herself with the best performing model in an optimal way. For an overview of methods of prediction with expert advice see the monograph by Cesa-Bianchi and Lugosi (2006) and a recent survey by Kalnishkan (2022).

Kalnishkan et al (2015) suggested an extension of this idea. Methods of prediction with expert advice may be used to select relevant training data. In machine learning, one often finds a trade-off between taking more data to ensure stability of estimates and concentrating on relevant data. When estimating volatility of a stock, how far can we go into the past before the price readings become irrelevant and start to confuse us instead of making our estimate better? With the pattern in the data drifting, what time window do we pick? Methods of prediction with expert advice allow the learner to merge predictions worked out on the basis of different data scopes and converge on the relevant horizon (in time and space) automatically.

Project Aims

The aims of the project are twofold. The theoretical aspect of the project consists of developing novel algorithms for prediction with expert advice and analysing their properties. The practical part consists of developing applications for novel domains, applying algorithms to real-world data, and analysing the results. The specific focus of the project may change depending on the findings and new ideas.

Theoretical Research

Prediction with expert advice gives rise to many interesting theoretical problems of varying difficulty.

Construction of Universal Algorithms

Methods of prediction with expert advice can be applied to construct predictors that perform as well as the best algorithm from a large parametric class. One can think of the resulting prediction algorithms as self-tuning: they optimise their parameters in the process of prediction to perform similarly to the retrospectively best parameter combination. Examples are provided by the Aggregating Algorithm Regression (Vovk-Azoury-Warmuth predictor) by Vovk (2001) and Azoury and Warmuth (2001), which performs as well as any linear regression with respect to the square loss, and Aggregating Algorithm Regression with Changing Dependencies (Busuttil and Kalnshkan, 2008 and Kalnishkan, 2016), which competes with slowly drifting linear regressions. An important goal of the project is construction of universal algorithms motivated by different theoretical scenarios and practical applications.

Algorithms for Prediction w.r.t. Continuous Ranked Probability Score

An interesting recent direction of research is probabilistic prediction with respect to Continuous Ranked Probability Score (CRPS). It is a mixable loss function (Vyugin and Trunov, 2019) but generalises absolute loss, which is not mixable (Dzhamtyrova and Kalnishkan, 2019). This opens new possibilities for situations where absolute loss and other similar losses matter (newsvendor problem, value at risk estimation).

Lower Bounds

A challenging problem is construction of theoretical lower bounds for universal algorithms and, generally, methods of prediction with expert advice. Vovk (2001) proposed a powerful method based on Bernoulli random processes with a hyperparameter and applied it to Aggregating Algorithm Regression. Dzhamtyrova and Kalnishkan (2021) generalise the method to probabilistic prediction with softmax applied on top of regression. The method should be capable of returning further interesting results. One particularly interesting problem is constructing a lower bound for the Aggregating Algorithm Regression with Changing Dependencies.

Practical Applications

Expanding the portfolio of successful applications of on-line learning should raise the awareness of methods of on-line learning among practitioners and feed back to the theoretical study of prediction. Economics and finance provide multiple scenarios generating data that fits the on-line prediction framework.

Value at Risk

Value at risk, VaR (e.g., of a portfolio of stocks) quantifies possible losses given a confidence level. The study of VaR is an important part of modern quantitative finance (see, e.g., the monograph by Alexander, 2009). The approaches for calculating VaR can grouped into model-based and historical. The model-based methods assume that the prices of underlying assets follow a distribution from some family. The dependency on the choice of the model is an obvious weakness here. The historical approach uses methods of non-parametric statistics; it is model-free but needs plenty of data for meaningful estimation. The methods of prediction with expert advice with its special approach to the underlying model are very promising here. Dzhamtyrova and Kalnishkan (2020a, 2020b) applied prediction with expert advice to quantile regression using pinball loss; this approach should be useful for value at risk but further study is needed.

Credit Scoring

The problem of credit scoring (see, e.g., the monograph of Thomas, 2009) addresses risk from a different perspective. It consists of predicting whether an economic agent (an individual or a firm) stays solvent over a period of time. Importantly, this happens against the backdrop of changing economic situation. Changing economic conditions may render past experience irrelevant or require data from some particular period from the past. This is an important potential playground for prediction with expert advice.

Predicting Demand

Predicting demand for goods and services is ever important for economic agents. Capturing the evolution of the demand is a complex problem, which has been attempted from the perspectives of statistics, time series, and machine learning. The methods of on-line learning have also been applied extensively.

Electricity consumption in particular has attracted much attention in the prediction with expert advice community (Devaine et al, 2013, Dzhamtyrova and Kalnishkan 2020b, V'yugin and Trunov, 2022). Levina et al. (2010) applied methods of prediction with expert advice to the Newsvendor problem, a classical operations management problem, which had hitherto been considered in the probabilistic context.

An interesting problem is the study of demand for related products, which can be approached with methods of Kalnishkan et al (2015).

Trading

Trading data is an example of inherently on-line data, where the feedback loop is at the heart of the scenario. There is a history of applications of universal algorithms and prediction with expert advice to investment (Cover and Ordentlich, 1998, Vovk and Watkins 1998), but applications in a real-life context still meet with important problems (Al-baghdadi et al, 2022). Trading on a margin with a possibility to open short positions make a bankruptcy a routine event. An adequate theoretical formalisation of this is lacking.

New Problem Domains

An important part of the project is identifying new potential application domains and seeking potential collaborations. The methods of prediction with expert advice have long developed with purely theoretical goals in mind; a search for more problem domains with bring much needed practical insight and will raise the awareness of methods of on-line learning among practitioners.

Training

The project will provide many opportunities for a young researcher to train and apply his or her skills to theoretical and practical problems in machine learning. While posing challenging mathematical problems, the project aims to get insight from practical applications. The student will have a range of diverging research directions to choose from to suit his or her interests.

The machine learning group at Royal Holloway has a history dating back to 1998, when the Computer Learning Research Centre (now Centre for Reliable Machine Learning) was established. Since 2013, Royal Holloway has been home to a popular family of MSc programmes in Big Data growing in size from 30 students in 2013-14 to 150 in 2020-21. MSc and PhD graduates of the department have pursues a range of successful career paths in academia and industry.

The student on the project will need a strong background in mathematics (analysis, linear algebra, probability and random processes) as well as numerical programming (e.g., using MATLAB, Julia, Python, or R). Familiarity with machine learning and mathematical finance will be an advantage.

Bibliography

Hastie, T., Friedman, J. and Tibshirani, R., 2017. The Elements of Statistical Learning. New York: Springer

Rasmussen, C. and Williams, C., 2006. Gaussian Processes for Machine Learning. MIT Press.

Hyndman, R.J. and Athanasopoulos, G., 2018. Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2.

Cesa-Bianchi, N. and Lugosi, G., 2006. Prediction, Learning, and Games. Cambridge University Press.

Kalnishkan, Y., 2022. Prediction with Expert Advice for a Finite Number of Experts: A Practical Introduction. To appear in *Pattern Recognition.*

Kalnishkan,Y., Adamskiy, D., Chernov, A., and Scarfe, T., 2015. Specialist experts for prediction with side information. In *2015 IEEE international conference on data mining workshop (ICDMW)*, (pp. 1470–1477). IEEE.

Vovk, V., 2001. Competitive on-line statistics. *International Statistical Review*, 69(2), pp.213-248.

Azoury, K.S. and Warmuth, M.K., 2001. Relative loss bounds for on-line density estimation with the exponential family of distributions. *Machine Learning*, *43*(3), pp.211-246.

Busuttil, S. and Kalnishkan, Y., 2007, October. Online regression competitive with changing predictors. In *International Conference on Algorithmic Learning Theory* (pp. 181-195). Springer, Berlin, Heidelberg.

Kalnishkan, Y., 2016, October. An upper bound for aggregating algorithm for regression with changing dependencies. In *International Conference on Algorithmic Learning Theory* (pp. 238-252). Springer

V'yugin, V.V. and Trunov, V.G., 2019, August. Online learning with continuous ranked probability score. In *Conformal and Probabilistic Prediction and Applications* (pp. 163-177). PMLR.

Dzhamtyrova, R. and Kalnishkan, Y., 2019. Competitive Online Regression under Continuous Ranked Probability Score, In *Proceedings of Machine Learning Research*, Vol.105., p. 178-195, Conformal and Probabilistic Prediction and Applications, Golden Sands, Bulgaria.

Alexander, C., 2009. Market Risk Analysis: Value at Risk Models. Wiley.

Dzhamtyrova, R. and Kalnishkan, Y. 2020a. Prediction with Expert Advice for Value at Risk. In *Proceedings of The 2020 International Joint Conference on Neural Networks (IJCNN 2020.)*, IEEE.

Dzhamtyrova, R. and Kalnishkan, Y., 2020b. Competitive online quantile regression. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems, Proceedings of the 18th International Conference IPMU 2020*, Springer.

Thomas, L.C., 2009. *Consumer credit models: Pricing, profit and portfolios*. OUP Oxford.

Devaine, M., Gaillard, P., Goude, Y., and Stoltz, G., 2013. Forecasting electricity consumption by aggregating specialized experts. *Machine Learning* 90 (2) (2013) 231-260.

V'yugin, V. and Trunov, V., 2022. Online aggregation of probability forecasts with confidence. To appear in *Pattern Recognition.*

Levina, T., Levin, Y., McGill, J., Nediak, M. and Vovk, V., 2010. Weak aggregating algorithm for the distribution-free perishable inventory problem. *Operations Research Letters*, *38*(6), pp.516-521.

Cover, T.M. and Ordentlich, E., 1996. Universal portfolios with side information. *IEEE Transactions on Information Theory*, *42*(2), pp.348-363.

Vovk, V. and Watkins, C., 1998, July. Universal portfolio selection. In *Proceedings of the eleventh annual conference on Computational learning theory* (pp. 12-23).

Al-Baghdadi, N., Lindsay, D., Kalnishkan, Y. and Lindsay, S., 2020, August. Practical investment with the long-short game. In *Conformal and Probabilistic Prediction and Applications* (pp. 209-228). PMLR.