The Sixth Symposium on Conformal and Probabilistic Prediction with Applications

COPA 2017

Hosted by Karolinska Institutet, Stockholm, Sweden

Tuesday 13 June to Friday 16 June 2017
Samuelsson Lecture Hall, Karolinska Institutet

Proceedings: `http://proceedings.mlr.press/v60/`
(Proceedings of Machine Learning Research, volume 60)

AstraZeneca

SeRC
Swedish e-Science Research Centre

Schedule

Day 0: Tuesday 13 June

10:00–12:00 Tutorial 1: Henrik Linusson *Introduction to Conformal Prediction* (abstract on p. 9)

12:00–13.30 Lunch

13.30–15:00 Tutorial 2: Marco Capuccini *Conformal Prediction in Spark* (abstract on p. 10)

15.30–17:00 Tutorial 3: Paolo Toccaceli *Venn Predictors* (abstract on p. 10)

Day 1: Wednesday 14 June

9:00–9:30 Registration

9:30–9:40 Welcome Address

9:40–10:40 Session 1 (Chair: Henrik Boström)

- Vladimir Vovk, Jieli Shen, Valery Manokhin, and Min-Ge Xie *Nonparametric Predictive Distributions Based on Conformal Prediction* (abstract on p. 5)
- Paolo Toccaceli and Alexander Gammerman *Combination of Conformal Predictors for Classification* (abstract on p. 5)

10:40–11:00 Coffee Break

11:00–12:00 Session 2 (Chair: Evgeny Burnaev)

- Vladimir V'yugin *Online Aggregation of Unbounded Signed Losses Using Shifting Experts* (abstract on p. 4)
- Yury Yanovich *Asymptotic Properties of Nonparametric Estimation on Manifold* (abstract on p. 4)

12:00–12:30 Poster Presentation Session (Chair: Ulf Norinder)

- the list of posters is on p. 3

12:30–13:30 Lunch

13:30–14:30 **Keynote 1** Prof. Vladimir Vapnik (AI Research Facebook and Columbia University, USA, and Royal Holloway, University of London, UK) *Intelligent Methods of Learning* (Chair: Alexander Gammerman)

14:30–15:00 Coffee Break

15:00–17:00 Session 3 (Chair: Martin Eklund)

- Denis Volkhonskiy, Evgeny Burnaev, Ilia Nouretdinov, Alexander Gammerman, and Vladimir Vovk *Inductive Conformal Martingales for Change-Point Detection* (abstract on p. 6)
- Vladislav Ishimtsev, Ivan Nazarov, Alexander Bernstein, and Evgeny Burnaev *Conformal k-NN Anomaly Detector for Univariate Data Streams* (abstract on p. 8)
- Charalambos Eliades and Harris Papadopoulos *Conformal Prediction for Automatic Face Recognition* (abstract on p. 5)
- Fan Yang, Xiaolu Gan, Huazhen Wang, Lei Feng, and Yongxuan Lai *CP-RAkEL: Improving Random k-labelsets with Conformal Prediction for Multi-label Classification* (abstract on p. 9)

Day 2: Thursday 15 June

9:30–10:00 Registration

10:00–11:40 Session 4 (Chair: Ola Engkvist)

- Henrik Linusson, Ulf Norinder, Henrik Boström, Ulf Johansson, and Tuve Löfstrm̈ *On the Calibration of Aggregated Conformal Predictors* (abstract on p. 6)
- Valery Manokhin *Probabilistic Prediction for Multi-class Classification* (abstract on p. 8)
- Ilia Nouretdinov *Improving Reliable Probabilistic Prediction by Using Additional Knowledge* (abstract on p. 8)

11:40–12:00 Coffee Break

12:00–13:00 Session 5 (Chair: Claus Bendtsen)

- Ilia Nouretdinov *Reverse Conformal Approach for On-line Experimental Design* (abstract on p. 7)
- Mikhail Malyutov and Paul Grosu *SCOT Approximation, Training and Asymptotic Inference* (abstract on p. 9)

13:00–14:00 Lunch

18:00 Conference Dinner (aboard the M/S Riddarholmen)

Day 3: Friday 16 June

9:00–10:00 Keynote 2 Dr Andreas Bender (Cambridge University, UK) *Conformal and Probabilistic Prediction for Chemical and Biological Data* (Chair: Lars Carlsson)

10:00–10:30 Coffee Break

10:30–12:30 Session 6 (Chair: Ola Spjuth) *Novel Directions of Applying Machine Learning in Cheminformatics*

- Staffan Arvidsson, Ola Spjuth, Lars Carlsson, and Paolo Toccaceli *Prediction of Metabolic Transformations Using Cross Venn-ABERS Predictors* (abstract on p. 6)
- Lars Carlsson, Claus Bendtsen, and Ernst Ahlberg *Comparing Performance of Different Inductive and Transductive Conformal Predictors Relevant to Drug Discovery* (abstract on p. 8)
- Ulf Norinder, Fredrik Svensson, and Avid Afzal *Maximizing Gain in HTS Screening Using Conformal Prediction* (abstract on p. 5)
- Ernst Ahlberg, Susanne Winiwarter, Henrik Boström, Henrik Linusson, Tuve Löfström, Ulf Norinder, Ola Engkvist, Oscar Hammar, Claus Bendtsen, and Lars Carlsson *Using Conformal Prediction to Prioritize Compound Synthesis in Drug Discovery* (abstract on p. 7)

12:30–13:00 Closing Remarks

13:00–14:00 Lunch

Posters

- Ulf Norinder, Fredrik Svensson, Scott Boyer, and Andreas Bender *Graceful Handling of Imbalanced Binary Datasets Using Mondrian Conformal Prediction*
- Jiangming Sun, Ola Engkvist, Lars Carlsson, and Hongming Chen *Estimate Prediction Confidence of Bayesian Based Matrix Factorization Method through Conformal Prediction*
- Avid Afzal, Fredrik Svensson, Ines Smit, James Lu, Jerome Mettetal, and Andreas Bender *Prediction of Drug-Drug Interactions in a Bayesian Framework*
- Christina Founti, Valerie J. Gillet, and Jonathan Vessey *Investigating the Effect of Reliability Estimation Methods on the (Perceived) Quality of Predictions*
- Denis Volkhonskiy, Ilia Nouretdinov, Alexander Gammerman, and Pitt Lim *Valid Probabilistic Prediction of Life Status after Percutaneous Coronary Intervention Procedure*
- Giuseppe Rattá and Jesus Vega *Well Calibrated Probability Estimates of the Low and High Confinement Modes Using Venn Predictors*
- Gayane Harutyunyan *The Sufficient Money Fund Level Prediction in Armenian Insurance Companies*

- Paul Grosu, Mikhail Malyutov, Javed Aslam, Hanai Sadaka, Virgil Pavlu, and Tong Zhang *Sparsity Against Exponential Complexity: Parallel Implementation of Separate Testing of Inputs Authorship Attribution via Sparse Stochastic Context Trees Modeling*

Abstracts of talks

Online Aggregation of Unbounded Signed Losses Using Shifting Experts **by Vladimir V. V'yugin** For the decision theoretic online (DTOL) setting, we consider methods to construct algorithms that suffer loss not much more than of any sequence of experts distributed along a time interval (shifting experts setting). We present a modified version of the method of Mixing Past Posteriors which uses as basic algorithm AdaHedge with adaptive learning rate. Due to this, we combine the advantages of both algorithms: regret bounds are valid in the case of signed unbounded losses of the experts, also, we use the shifting regret which is a more optimal characteristic of the algorithm. All results are obtained in the adversarial setting—no assumptions are made about the nature of data source.

We present results of numerical experiments for the case where losses of the experts cannot be bounded in advance. **PMLR 60:3–17**

Asymptotic Properties of Nonparametric Estimation on Manifold **by Yury Yanovich** In many applications, the real high-dimensional data occupy only a very small part in the high dimensional "observation space" whose intrinsic dimension is small. The most popular model of such data is Manifold model which assumes that the data lie on or near an unknown manifold (Data Manifold, DM) of lower dimensionality embedded in an ambient high-dimensional input space (Manifold Assumption about high-dimensional data). Manifold Learning is a Dimensionality Reduction problem under the Manifold assumption about the processed data, and its goal is to construct a low-dimensional parameterization of the DM (global low-dimensional coordinates on the DM) from a finite dataset sampled from the DM.

Manifold Assumption means that local neighborhood of each manifold point is equivalent to an area of low-dimensional Euclidean space. Because of this, most of Manifold Learning algorithms include two parts: "local part" in which certain characteristics reflecting low-dimensional local structure of neighborhoods of all sample points are constructed via nonparametric estimation, and "global part" in which global low-dimensional coordinates on the DM are constructed by solving the certain convex optimization problem for specific cost function depending on the local characteristics. Both statistical properties of "local part" and its average over manifold are considered in the paper. The article is an extension of the paper Yanovich (2016) for the case of nonparametric estimation. **PMLR 60:18–38**

Combination of Conformal Predictors for Classification **by Paolo Toccaceli and Alexander Gammerman** The paper presents some possible approaches to the combination of Conformal Predictors in the binary classification case. A first class of methods is based on p-value combination techniques that have been proposed in the context of Statistical Hypothesis Testing; a second class is based on the calibration of p-values into Bayes factors. A few methods from these two classes are applied to a real-world case, namely the chemoinformatics problem of Compound Activity Prediction. Their performance is discussed, showing the different abilities to preserve of validity and improve efficiency. The experiments show that P-value combination, in particular Fisher's method, can be advantageous when ranking compounds by strength of evidence. **PMLR 60:39–61**

Conformal Prediction for Automatic Face Recognition **by Charalambos Eliades and Harris Papadopoulos** Automatic Face Recognition (AFR) has been the subject of many research studies in the past two decades and has a wide range of applications. The provision of some kind of indication of the likelihood of a recognition being correct is a desirable property of AFR techniques in many applications, such as for the detection of wanted persons or for performing post-processing in automatic annotation of photographs. This paper investigates the use of the Conformal Prediction (CP) framework for providing reliable confidence information for AFR. In particular we combine CP with two classifiers based on calculating similarities between images using Scale Invariant Feature Transformation (SIFT) features. We examine and compare the performance of several nonconformity measures for the particular task in terms of their accuracy and informational efficiency. **PMLR 60:62–81**

Nonparametric Predictive Distributions Based on Conformal Prediction **by Vladimir Vovk, Jieli Shen, Valery Manokhin, and Min-ge Xie** This paper applies conformal prediction to derive predictive distributions that are valid under a nonparametric assumption. Namely, we introduce and explore predictive distribution functions that always satisfy a natural property of validity in terms of guaranteed coverage for IID observations. The focus is on a prediction algorithm that we call the Least Squares Prediction Machine (LSPM). The LSPM generalizes the classical Dempster–Hill predictive distributions to regression problems. If the standard parametric assumptions for Least Squares linear regression hold, the LSPM is as efficient as the Dempster–Hill procedure, in a natural sense. And if those parametric assumptions fail, the LSPM is still valid, provided the observations are IID. **PMLR 60:82–102**

Maximizing Gain in HTS Screening Using Conformal Prediction **by Ulf Norinder, Fredrik Svensson, Avid M. Afzal, and Andreas Bender** Today, screening of large compound collections in high throughput screening campaigns form the backbone of early drug discovery. Although widely applied, this approach is resource and potentially labour intensive. Therefore, improved

computational approaches to streamline screening is in high demand. In this study we introduce conformal prediction paired with a gain-cost function to make predictions in order to maximise the gain of screening campaigns on new screening sets. Our results indicate that using 20% of the screening library as an initial screening set and using the data obtained together with a gain-cost function, the significance level of the predictor that maximise the gain can be identified. Importantly, the parameters for the predictor derived from the initial screening set was highly predictive of the maximal gain also on the remaining data. Using this approach, the gain of a screening campaign can be improved considerably. **PMLR 60:103–117**

Prediction of Metabolic Transformations using Cross Venn-ABERS Predictors **by Staffan Arvidsson, Ola Spjuth, Lars Carlsson, and Paolo Toccaceli** Prediction of drug metabolism is an important topic in the drug discovery process, and we here present a study using probabilistic predictions applying Cross Venn-ABERS Predictors (CVAPs) on data for site-of-metabolism. We used a dataset of 73599 biotransformations, applied SMIRKS to define biotransformations of interest and constructed five datasets where chemical structures were represented using signatures descriptors. The results show that CVAP produces well-calibrated predictions for all datasets with good predictive capability, making CVAP an interesting method for further exploration in drug discovery applications. **PMLR 60:118–131**

Inductive Conformal Martingales for Change-Point Detection **by Denis Volkhonskiy, Evgeny Burnaev, Ilia Nouretdinov, Alexander Gammerman, and Vladimir Vovk** We consider the problem of quickest change-point detection in data streams. Classical change-point detection procedures, such as CUSUM, Shiryaev-Roberts and Posterior Probability statistics, are optimal only if the change-point model is known, which is an unrealistic assumption in typical applied problems. Instead we propose a new method for change-point detection based on Inductive Conformal Martingales, which requires only the independence and identical distribution of observations. We compare the proposed approach to standard methods, as well as to change-point detection oracles, which model a typical practical situation when we have only imprecise (albeit parametric) information about pre- and post-change data distributions. Results of comparison provide evidence that change-point detection based on Inductive Conformal Martingales is an efficient tool, capable to work under quite general conditions unlike traditional approaches. **PMLR 60:132–153**

On the Calibration of Aggregated Conformal Predictors **by Henrik Linusson, Ulf Norinder, Henrik Boström, Ulf Johansson, and Tuve Löfström** Conformal prediction is a learning framework that produces models that associate with each of their predictions a measure of statistically valid confidence. These models are typically constructed on top of traditional machine learning algorithms. An important result of conformal prediction theory is that

the models produced are provably valid under relatively weak assumptions—in particular, their validity is independent of the specific underlying learning algorithm on which they are based. Since validity is automatic, much research on conformal predictors has been focused on improving their informational and computational efficiency. As part of the efforts in constructing efficient conformal predictors, aggregated conformal predictors were developed, drawing inspiration from the field of classification and regression ensembles. Unlike early definitions of conformal prediction procedures, the validity of aggregated conformal predictors is not fully understood—while it has been shown that they might attain empirical exact validity under certain circumstances, their theoretical validity is conditional on additional assumptions that require further clarification. In this paper, we show why validity is not automatic for aggregated conformal predictors, and provide a revised definition of aggregated conformal predictors that gains approximate validity conditional on properties of the underlying learning algorithm. **60:154–173**

Using Conformal Prediction to Prioritize Compound Synthesis in Drug Discovery **by Ernst Ahlberg, Susanne Winiwarter, Henrik Boström, Henrik Linusson, Tuve Löfström, Ulf Norinder, Ulf Johansson, Ola Engkvist, Oscar Hammar, Claus Bendtsen, and Lars Carlsson** The choice of how much money and resources to spend to understand certain problems is of high interest in many areas. This work illustrates how computational models can be more tightly coupled with experiments to generate decision data at lower cost without reducing the quality of the decision. Several different strategies are explored to illustrate the trade off between lowering costs and quality in decisions.

AUC is used as a performance metric and the number of objects that can be learnt from is constrained. Some of the strategies described reach AUC values over 0.9 and outperforms strategies that are more random. The strategies that use conformal predictor p-values show varying results, although some are top performing. **60:174–184**

Reverse Conformal Approach for On-line Experimental Design **by Ilia Nouretdinov** Conformal prediction is a recently developed framework of confident machine learning with guaranteed validity properties for prediction sets. In this work we study its usage in reversed version of the traditional machine learning problem: prediction of objects which can have a given label, instead of usual prediction of labels by objects. It is meant that the label reflect some desired property of the object. For this kind of task, the conformal prediction framework can provide a prediction set that is a set of objects that are likely to have the label. Based on this, we create an on-line protocol of experimental design. It includes a choice criterion based on conformal output, and elements of transfer learning in order to keep the validity properties in on-line regime. **PMLR 60:185–192**

Improving Reliable Probabilistic Prediction by Using Additional Knowledge **by Ilia Nouretdinov** Venn Machine is a recently developed machine learning framework for reliable probabilistic prediction of the labels for new examples. This work proposes a way to extend Venn machine to the framework known as Learning Under Privileged Information: some additional features are available for a part of the training set, and are missing for the example being predicted. We make use of this information by making a taxonomy transfer, where taxonomy is the core detail of Venn Machine framework. The transfer is done from the examples with additional information to the examples without additional information. **PMLR 60:193–200**

Comparing Performance of Different Inductive and Transductive Conformal Predictors Relevant to Drug Discovery **by Lars Carlsson, Claus Bendtsen, and Ernst Ahlberg** We present an evaluation of the impact of transductive, inductive, aggregated and cross inductive mondrian conformal prediction on the validity and efficiency of predictions. The aim of the study is to give guidance to which methods perform best where there is limited data. The evaluation has been made on a large public dataset of Ames mutagenicity data, relevant for drug discovery, a spam dataset and a diverse set of drug discovery datasets. When considering predictions only, the transductive conformal predictor performs the best in terms of validity. If however more information is required, for example interpretation of a prediction, then any of the methods that calculate an averaged p-value should be considered. **PMLR 60:201–212**

Conformal k-NN Anomaly Detector for Univariate Data Streams **by Vladislav Ishimtsev, Alexander Bernstein, Evgeny Burnaev, and Ivan Nazarov** Anomalies in time-series data give essential and often actionable information in many applications. In this paper we consider a model-free anomaly detection method for univariate time-series which adapts to non-stationarity in the data stream and provides probabilistic abnormality scores based on the conformal prediction paradigm. Despite its simplicity the method performs on par with complex prediction-based models on the Numenta Anomaly Detection benchmark and the Yahoo! S5 dataset. **PMLR 60:213–227**

Multi-class probabilistic classification using inductive and cross Venn–Abers predictors **by Valery Manokhin** Inductive (IVAP) and cross (CVAP) Venn–Abers predictors are computationally efficient algorithms for probabilistic prediction in binary classification problems. We present a new approach to multi-class probability estimation by turning IVAPs and CVAPs into multi-class probabilistic predictors. The proposed multi-class predictors are experimentally more accurate than both uncalibrated predictors and existing calibration methods. **PMLR 60:228–240**

SCOT Approximation, Training and Asymptotic Inference **by Mikhail Malyutov and Paul Grosu** Approximation of stationary strongly mixing processes by Stochastic Context Trees (SCOT) models and the Le Cam-Hajek-Ibragimov-Khasminsky locally minimax theory of statistical inference for them is outlined. SCOT is an m-Markov model with sparse memory structure. In our previous papers we proved SCOT equivalence to 1-MC with state space—alphabet consisting of the SCOT contexts. For the fixed alphabet size and growing sample size, the Local Asymptotic Normality is proved and applied for establishing asymptotically optimal inference. We outline what obstacles arise for a large SCOT alphabet size and not necessarily vast sample size. Training SCOT on a large string using clusters of computers and statistical applications are described. **PMLR 60:241–265**

CP-RAkEL: Improving Random k-labelsets with Conformal Prediction for Multi-label Classification **by Fan Yang, Xiaolu Gan, Huazhen Wang, Lei Feng, and Yongxuan Lai** Multi-label conformal prediction has attracted much attention in the conformal predictor (CP) community. In this article, we propose to combine CP with random k-labelsets (RAkEL) method, which is state-of-the-art multi-label classification method for large number of labels. In the framework of RAkEL, the original problem is reduced to a number of small-sized multi-label classification tasks by randomly breaking the initial set of labels into a number of small-sized labelsets, and then label powerset (LP) method is employed on these tasks respectively. In this work, ICP-RF, an inductive conformal predictor based on random forest, is used in each multi-label task in order to get p-values for predictions of the LP model, and then the predictions are aggregated to get a final result. Experimental results on six benchmark datasets empirically demonstrate the calibration property of ICP-RF as LP models, and show that conformal prediction can significantly improve the performances of the proposed approach, which is called RAkEL. However, the validity property of CP does not hold in CP-RAkEL. In the future work we will study how to use some new CP techniques to calibrate the new method. **PMLR 60:266–279**

Abstracts of tutorials

Tutorial 1 *Introduction to Conformal Prediction.* Teacher: Henrik Linusson, University of Borås, Sweden.

How good is your prediction? In risk-sensitive applications, it is crucial to be able to assess the quality of a prediction, however, traditional classification and regression models don't provide their users with any information regarding prediction trustworthiness. In contrast, conformal classification and regression models associate each of their multi-valued predictions with a measure of statistically valid confidence, and let their users specify a maximal threshold of the model's error rate — the price to be paid is that predictions made with a higher confidence cover a larger area of the possible output space. This tutorial aims

to provide its attendees with the knowledge necessary to implement conformal prediction in their daily data science work, be it research or practice oriented, as well as highlight current research topics on the subject.

Since its development the framework has been combined with many popular techniques, such as Support Vector Machines, k-Nearest Neighbours, Neural Networks, Ridge Regression etc., and has been successfully applied to many challenging real world problems, such as the early detection of ovarian cancer, the classification of leukaemia subtypes, the diagnosis of acute abdominal pain, the assessment of stroke risk, the recognition of hypoxia in electroencephalograms (EEGs), the prediction of plant promoters, the prediction of network traffic demand, the estimation of effort for software projects and the back calculation of non-linear pavement layer moduli. The framework has also been extended to additional problem settings such as semi-supervised learning, anomaly detection, feature selection, outlier detection, change detection in streams and active learning. The aim of this symposium is to serve as a forum for the presentation of new and ongoing work and the exchange of ideas between researchers on any aspect of Conformal Prediction and its applications.

Tutorial 2 *Conformal Prediction in Spark.* Teacher: Marco Capuccini, Department of Pharmaceutical Biosciences, Uppsala University, Sweden.

This tutorial will introduce the Spark framework for automation of model building. The Spark framework is a cluster-computing engine for large data processing that makes coding in massively parallel pipelines easy, like never before. Spark applications can be written in Scala, Java, Python and R (we will use Scala in this tutorial), and such applications can run with minor adaption on HPC clusters, cloud computing engines (e.g., Amazon EC2) and local machines. Furthermore, Spark applications are out-of-the-box scalable and fault tolerant. This key features, along with the Spark built-in machine learning library, allows one to code massively parallel pipelines for predictive modeling faster, and they are therefore to be more productive.

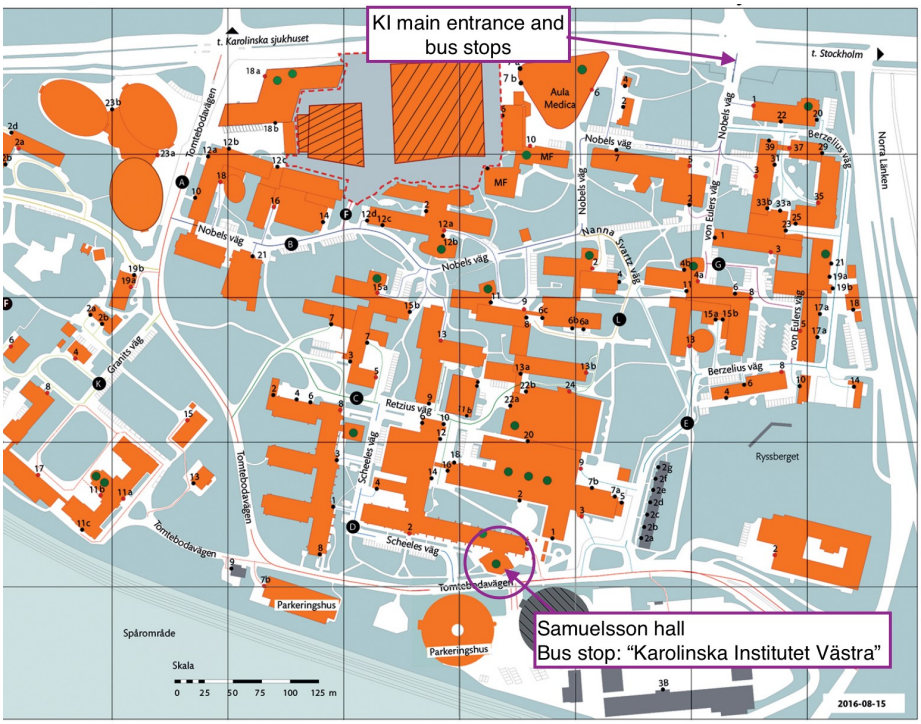
Tutorial 3 *Venn Predictors.* Teacher: Paolo Toccaceli, Royal Holloway, University of London, UK.

Machine Learning is primarily concerned with producing prediction rules, i.e., functions mapping "objects" onto predicted "labels", given a training set of (object,label) examples. Practitioners often focus on the value of the prediction, but overlook its uncertainty. Venn Predictors offer a principled way of assigning a probability to predictions, relying on a minimal set of assumptions. One distinguishing feature of Venn Predictors is that they have a theoretically-backed property of calibration, in the sense that the probabilities reflect the long-term distribution. The tutorial will also introduce Venn-ABERS Predictors, which offer an efficient way to transform the output of a scoring classifier into a prediction probability. A comparison with other methods, such as Platt Scaling, will be discussed using a practical example from a Chemoinformatics application.

Practical information

Map

The Samuelsson Lecture Hall is located at Tomtebodavägen 6 at the Karolinska Institutet Solna Campus. Please note that the tutorials on the 13th of June will be in the Widerströmska huset at Tomtebodavägen 18.



Conference dinner

We will have our conference dinner onboard the M/S Riddarholmen, where we will have a three course dinner while cruising the Stockholm archipelago. M/S Riddarholmen was built in 1916 at Motala Shipyards and renovated in 1990. In 2010 Blidösundsbolaget bought the ship, and she was moved to Skeppsbron in front of the Royal Castle in Stockholm.

The M/S Riddarholmen

Stockholm archipelago

How to find M/S Riddarholmen

The M/S Riddarholmen is located at the quay very close to the royal castle in the Stockholm Old Town. The easiest way to get there is to take the subway to the "Gamla Stan" or "Slussen" stations on the green line and then walk from there or to the station "Kungsträdgården" on the blue line and walk from there (approximately the same walking distance from all three stations).