領域論壇：數據科學
Session: Data Science

Venue: 數學館 M212

Time	**Speaker**	**Title of the Talk**	**Chair**
13:20–13:45	陳瑞彬	Category Tree Gaussian Process for Computer Experiments with Many-Category Qualitative Factors and Application to Cooling System Design	呂翠珊
13:45–14:10	Pierre-Louis Poirion	Random-subspaces Newton method for unconstrained non-convex optimization	呂翠珊
14:10–14:35	溫啓仲	Is "the seven year itch" real? – Regression analysis of randomized response event time data	呂翠珊
14:35-15:00	Yeonjong Shin	Towards Trustworthy Scientific Machine Learning: Theory, Algorithms, and Applications	呂翠珊

Category Tree Gaussian Process for Computer Experiments with Many-Category Qualitative Factors and Application to Cooling System Design

Ray-Bing Chen
Department of Statistics & Institute of Data Science, National Cheng Kung University
E-mail: rbchen@ncku.edu.tw

Abstract

In computer experiments, Gaussian process (GP) models are commonly used for emulation. However, when both qualitative and quantitative factors are in the experiments, emulation using GP models becomes challenging. In particular, when the qualitative factors contain many categories in the experiments, existing methods in the literature become cumbersome due to the curse of dimensionality. Motivated by the computer experiments for the design of a cooling system, a new tree-based GP is proposed that emulates computer models with many-category qualitative factors, which we call category tree GP. The proposed method incorporates a tree structure to split the categories of the qualitative factors, and GP or mixed-input GP models are employed for modeling the simulation outputs in the leaf nodes. The splitting rule takes into account the cross-correlations between the categories of the qualitative factors, which have been shown by a recent theoretical study to be a crucial element for improving the prediction accuracy. In addition, a pruning procedure based on the cross-validation error is proposed to ensure the prediction accuracy. The application to the design of a cooling system indicates that the proposed method not only enjoys marked computational advantages and produces accurate predictions, but also provides valuable insights into the cooling system by discovering the tree structure.

Random-subspaces Newton method for unconstrained non-convex optimization

Pierre-Louis Poirion

RIKEN Center for Advanced Intelligence Project, Tokyo, Japan

E-mail: pierre-louis.poirion@riken.jp

Abstract

In this talk, we present a randomized subspace regularized Newton method for a non-convex function. We show that our method has global convergence under appropriate assumptions, and its convergence rate is the same as that of the full regularized Newton method. Furthermore, we can obtain a local linear convergence rate, under some additional assumptions, and prove that this rate is the best we can hope when using random subspace.

Is "the seven year itch" real? – Regression analysis of randomized response event time data

Chi-Chung Wen
Department of Mathematics, Tamkang University
E-mail: ccwen@mail.tku.edu.tw

Abstract

The randomized response techniques (RRTs), including the related-question RRT of Warner (1965) and the unrelated-question RRT of Greenberg et al. (1969), have been utilized to reduce under-reporting of sensitive characteristics in survey studies by enhancing privacy protection. Currently, the RRT is mainly applied for prevalence estimation of some sensitive event. This work extends the application of the RRT to the analysis of time-to-event outcome. We apply the proposed method to the Taiwan extramarital sex data surveyed by the RRT to make statistical inferences on the time to the incidence of extramarital sex since marriage. In final, we discuss the authenticity of "the seven year itch" based on the Taiwan survey data.

Towards Trustworthy Scientific Machine Learning: Theory, Algorithms, and Applications

Yeonjong Shin
Department of Mathematical Sciences, KAIST, Korea
E-mail: yeonjong_shin@kaist.ac.kr

Abstract

Machine learning (ML) has achieved unprecedented empirical success in diverse applications. It now has been applied to solve scientific problems, which has become an emerging field, Scientific Machine Learning (SciML). Many ML techniques, however, are very complex and sophisticated, commonly requiring many trial-and-error and tricks. These result in a lack of robustness and interpretability, which are critical factors for scientific applications. This talk centers around mathematical approaches for SciML, promoting trustworthiness. The first part is about how to embed physics into neural networks (NNs). I will present a general framework for designing NNs that obey the first and second laws of thermodynamics. The framework not only provides flexible ways of leveraging available physics information but also results in expressive NN architectures. The second part is about the training of NNs, one of the biggest challenges in ML. I will present an efficient training method for NNs - Active Neuron Least Squares (ANLS). ANLS is developed from the insight gained from the analysis of gradient descent training.