Mathematics of Machine Learning
Les mathématiques de l'apprentissage machine
(Org: **Simone Brugiapaglia** (Concordia) and/et **Tim Hoheisel** (McGill))

NICHOLAS BARNFIELD, Harvard

MEHDI DAGDOUG, McGill University
Agnostic Model-Assisted Estimation in Finite Population Sampling

The aim of survey sampling is to estimate finite population parameters. In many settings, auxiliary information is available at the population level and can be incorporated through modeling within the model-assisted framework. Establishing the asymptotic properties of such estimators has traditionally been carried out on a method-by-method basis and has remained an active area of research over the past 25 years. Although these estimators often share similar properties, the techniques used to analyze them are typically tailored to the specific prediction method. A general theoretical framework applicable to arbitrary prediction methods is currently lacking.

In this work, we address this gap by introducing a novel form of conditional cross-fitting adapted to the dependence arising in survey sampling. Leveraging this approach, we develop a unified theory for the asymptotic behavior of model-assisted estimators based on general prediction methods. In particular, we establish the asymptotic normality of the proposed estimators and the consistency of a corresponding cross-fitting variance estimator, thereby enabling the construction of asymptotically valid confidence intervals. The proposed approach is shown to be asymptotically optimal in that it reaches the Godambe-Joshi lower bound. Simulation studies further demonstrate the favorable finite-sample performance of the proposed methodology.

ARIEL GOODWIN, Cornell University
Minimum enclosing ball algorithms in geodesic spaces

Finding the smallest ball enclosing a bounded set of points is a fundamental optimization problem in computational geometry and operations research, where it is sometimes known as the minimax facility location problem. Algorithms for solving this problem in Euclidean space are plentiful, but there are fewer viable options in non-Euclidean spaces like Riemannian manifolds and geodesic metric spaces. In this talk we highlight the role that geodesics have played in the design of minimum enclosing ball algorithms on non-Euclidean spaces, dating back to a finitely-convergent algorithm on tree graphs. We demonstrate that a simple greedy geodesic-following algorithm, known to converge in the Riemannian case, admits a simple complexity analysis in geodesic metric spaces satisfying an upper curvature bound. In the case of nonpositive curvature, we also propose a projection-based algorithm that generates certifiable lower bounds on the minimum enclosing ball radius.

MAROUANE IL IDRISSI, UQAM

SOPHIE MORIN, Polytechnique Montréal

RACHEL MORRIS, Concordia

JANOSCH ORTMANN, UQAM

STEFANIA PETRA, Heidelberg

KUNLUN QI, Michigan State

GABRIEL RIOUX, Imperial

MIN WANG, Houston

XUEMENG WANG, Simon Fraser University
Into the danger zone: stable extrapolation in high-dimensional function and operator learning

Out-of-distribution (OOD) generalization is a central challenge in scientific machine learning (SciML). We study regression problems in which the test distribution differs from the training distribution and ask: under what assumptions on the target function or operator is stable extrapolation possible, and how far beyond the training domain can one extrapolate? Existing theory often controls the test error through additive penalties measuring the discrepancy between the training and test distributions. Such guarantees can be considerably more pessimistic than the generalization observed empirically.

Under holomorphic regularity assumptions on the target, we identify classes of functions and operators for which the OOD generalization error converges at algebraic rates. This phenomenon stems from the increasing smoothness of higher-index coordinates, leading to what we term a 'blessing of high dimensionality' We derive explicit rates for arbitrary test measures supported on suitable domains and quantify how the admissible domain depends on the underlying regularity. Our extrapolation guarantees are independent of the test distribution, depending only on its support. We establish matching rates for polynomial approximation, deep neural networks, and deep neural operators. In addition, we present extensive numerical experiments across a range of functions and operators that corroborate the predicted rates.

LAURA WEIDENSAGER, Simon Fraser

ANDRE WIBISONO, Yale