HAUSDORFF CENTER FOR MATHEMATICS

UNIVERSITÄT BONN

Hausdorff School

"Foundational Methods in Machine Learning"

June 6 - 10, 2022

organized by
Franca Hoffmann, Martin Rumpf

Abstracts for Courses

Weinan E (Peking University (China) and Princeton University (US))

Understanding Machine Learning Models in High Dimension

Abstract: In essence, the subject of machine learning is about approximating functions, approximating probability distributions or solving equations such as the Bellman equation in dynamic programming. All these are classical problems that lie at the center of numerical analysis or statistics, except for one key difference: In classical numerical analysis, we are only able to handle problems in very low dimensions whereas machine learning typically aims at problems in much higher dimension. Therefore a natural question is: What is the magic behind machine learning?

The one area in which high dimensional problems have been well studied in numerical analysis is numerical integration. The lesson we learn from this success is that probabilistic approaches are essential in high dimension.

In this series of lectures, we focus on the regression problem in supervised learning, i.e. the problem of approximating high dimensional functions. We try to address the following question: Given a machine learning model, what are the class of functions that can be approximated by this particular model efficiently, in the sense that the convergence rate for the approximation, estimation and optimization errors does not deteriorate as dimensionality goes up? We address this question for three classes of machine learning models: The random feature model, two-layer neural networks and the residual neural network model.

In the process, we will advance several ideas that we feel are important for analyzing functions in high dimension, such as the probabilistic representation of functions and the complexity of function spaces. We will also discuss the important differences for the gradient-based optimization algorithms in low and high dimensions.

The lecture series ends with a discussion about some of the key open theoretical questions in this area. This lecture series is prepared together with Professor Lei Wu at Peking University.

Dejan Slepcev (Carnegie Mellon University (US))

Variational problems and PDE on random structures: analysis and applications to data science

Abstract: Many machine learning tasks, such as clustering, regression, classification, and dimensionality reduction are commonly described as optimization problems. Namely, these tasks are modeled by functionals, defined using the available random sample, which specify the desired properties of the object sought. While the data are often high dimensional, they usually have an intrinsic low-dimensional structure that makes the learning tasks feasible. The intrinsic geometry is often encoded by a graph created by connecting the nearby data points. We will introduce mathematical tools used to study variational problems and PDE-based models posed on random data samples. In particular we will discuss the connection between discrete optimization problems on random samples to their continuum limits. This will be used to establish asymptotic consistency of several important machine learning algorithms. Furthermore we will use these connections to guide the modeling of machine learning tasks.
We will outline the basic elements of the background material on calculus of variations. We will develop connections to nonlocal functionals which serve as intermediate objects between the discrete functionals and their continuum limits. Moreover, we will also consider approaches based on dynamics on graphs and connect these with the evolution equations describing the continuum limits

---

Carola-Bibiane Schönlieb (University of Cambridge (UK))

Data-driven solutions to inverse imaging problems

Abstract: In this short lecture series I will give an introduction to some new concepts that use data-driven components within solutions to inverse imaging problems. Presented methods include data-driven variational models and plug-and-play approaches, learned iterative schemes aka learned unrolling, and learned post-processing. In the first part of the lecture I will give an introduction to inverse problems, classical solution strategies and discuss applications. In the second part we will investigate learned variational models and plug-and-play approaches. In the third part we discuss the idea of unrolling an iterative reconstruction algorithm and turning it into a data-driven reconstruction approach by appropriate parametrisation and optimisation. Throughout presenting these methodologies, we will discuss their theoretical properties and provide numerical examples for image denoising, deconvolution and computed tomography reconstruction. The lecture series will finish with a discussion of open problems and future perspectives.
This lecture series is mainly based on Arridge, S., Maass, P., Öktem, O., Schönlieb, C. B. (2019). Solving inverse problems using data-driven models. Acta Numerica, 28, 1-174.

Rachel Ward (The University of Texas (US))

An introduction to optimization in large-scale machine learning

Abstract: We will give an introduction to optimization for large-scale machine learning. We will start with reviewing the basic convergence theory for gradient descent. We will then motivate the case for stochastic gradient descent in the large-scale setting, and discuss convergence theory for SGD. Finally, we will discuss extensions of the basic SGD algorithm which are important for robustness and accelerated convergence: adaptive gradient SGD and momentum.

Abstracts for Contributed Talks

Jing An (Max Planck Institute for Mathematics in the Sciences (Germany))

Resampling vs. Reweighting for faithful stochastic optimization

Abstract: Resampling and reweighting are two commonly-used techniques to rebalance biasedly-sampled data sets. In this talk, I will discuss resampling vs. reweighting in two types of data science problems when stochastic gradient algorithms are used. The first problem is that when training models, to correct bias in the data sets, resampling is often observed to outperform reweighting. The reason behind this phenomenon can be explained by using tools from dynamical stability and stochastic asymptotics. The second problem we consider is that when a data set has feature disparities, its corresponding loss landscape will cause stochastic gradient descent to experience different variances at different minima. In order to mitigate such bias and perform faithful optimization, we propose a combined resampling-reweighting scheme to balance the variance at local minima. Further quantitative results about how the combined resampling-reweighting strategy improves stability and local convergence will also be discussed in this talk.

Cristina Cipriani (Technische Universität München (Germany))

A Mean-Field Optimal Control Approach to the Training of NeurODEs

Abstract: NeurODEs are a specific type of neural networks which contain shortcut connections that allow interpreting their training as a stochastic optimal control problem. Starting from this concept and extending it to its mean-field version, we derive first order optimality conditions in the form of a mean-field version of the Pontryagin Maximum Principle based on a generalized Lagrange multiplier theorem on convex sets of spaces of measures. The resulting training method provides a unique control solution, which is Lipschitz continuous. Finally, some explanatory and easy-to-read numerical examples will give insights into the resulting algorithm.

Stephen Moore (University of Cape Coast (Ghana))

Transmission dynamics of Epidemiological models: Ordinary and Fractional Differential Equations

Abstract: In this talk, we present the mathematical analysis for the transmission dynamics of epidemiological models. We will consider a COVID-19 model using data from Ghana. In the first part, we will consider the analysis and numerics for a determinstic model. In the second part, we will extend the analysis to the Caputo fractional order model. Finally, we will present numerical results for the fractional order model and discuss possible future directions.

Rodrigo Iza Teran (Fraunhofer Institute SCAI (Germany))

Spectral Approach for Effective Surrogate Modelling

Abstract: Constructing interpretable reduced order models for Finite Element Analysis using a learning approach for realistic engineering data where several input parameters change is challenging.

State of the art methods are based on the Proper Orthogonal Decompositions (POD) and variations of it. This approach relies on a truncated SVD decomposition of a set of simulation snapshots. A drawback of this approach is that it relies on computing projections on the principal axes of an m-dimensional ellipsoid that best fits the data co-variance in a least square sense. This is optimal if the simulated dynamics is linear but non-linear dynamics will in general produce data that will lie near or on a curved manifold. Despite this, such cases do have a natural linear structure, namely the Hilbert space of square-integrable functions on an intrinsic manifold provided adequate a linear operator can be found for this case. We propose an approach that under certain assumptions compute such a linear operator, the Laplace-Beltrami operator, that is independent of the 3D Finite Element results and depends only on the underlying geometry. A low dimensional latent space representation can be gained by projecting the simulation results to the eigenvectors of the Laplace-Beltrami operator and keeping only the projections with highest variance. This projections are shown to represent similarities between simulation fields so that clusters can be found in the simulations. Furthermore, they correlate with determined sensible input variables making them suitable as an intermediate representation to compute predictors for the Finite Element results. Example applications for CFD and Forming simulations are presented to demonstrate the approach.

Rachelle Sambayan (University of the Philippines (Philippines))

Kernel Matrix Completion

Abstract: In this study, in order to improve the accuracy of machine learning, methodologies for completing multiple incomplete data representations have been developed. In data analysis including pattern recognition and clustering, each data object is not limited to a single representation. In many applications, several measurement methods are available to objects to be analyzed, yielding multiple data representations. For example, in the task of function prediction of proteins in cells, each protein can be represented with its amino acid sequence, cubic structure, interactions with other proteins, and expression data. Each data representation provides useful information for function prediction. Proteins that have homology in their amino acid sequences are likely to have same functions in cells. The cubic structures of proteins determine functions. Many function mechanisms in cells depend on multiple interacted proteins. Proteins with same functions express the same conditions in cells. Thus, each of these representations is informative for function prediction. Research reports have shown that analysis accuracy is improved by combining multiple data representations. However, an issue in data analysis based on multiple data representations is that data examples lacking any representation cannot be included in machine learning. In this study, several new methods for completing incomplete data are presented. To assess the effectiveness of the new data completion methods, experiments on real-world data are carried out. The results are then reported.

Christian Offen (Paderborn University (Germany))

Learning Euler–Lagrange Dynamics from Data

Abstract: To identify equations of motions that govern a dynamical system from observed motions is an important task in a variety of applications including optimal control of mechanical systems. While the differential equations themselves might be unknown, prior knowledge about their structure is often available, for instance that these arise from a variational principle, i.e. they are Euler–Lagrange equations to some (a priori unknown) Lagrangian. This relates to many qualitative features of dynamical systems, such as the presence of conservation laws. It is, therefore, crucial to incorporate variational structure into learning algorithms for dynamical systems such that the learned model shares important features with the exact physical system.
