Mathematics of Machine Learning
Mathématiques de l'apprentissage automatique

(Org: **Simone Brugiapaglia** (Concordia University), **Vakhtang Putkaradze** (University of Alberta) and/et **Hamid Usefi** (Memorial University of Newfoundland))

OSAMA BATAINEH, Univ. of Saskatchewan
Imprecise Probabilities for Cybersecurity Applications

In cybersecurity and cryptanalysis, the measurement of cyber-risk is important and crucial for protection against cyberattacks. In cyber threats, probabilistic models can be thought of, and selected to measure the risk of occurrence of cyberattacks and threats. Imprecise probabilities are used to present the differences in prior beliefs amongst cryptanalysts, on cyber breaches and their probabilities of occurrence. Imprecise probabilities do capitalize the prediction margin of several types of cyber-risk, and can also give the cryptanalyst the opportunity to reduce it. For each threat/attack, there will be lower and upper bound probability estimates, based on implementing Bayesian methods with sets of prior probability distributions. Prior changes will be investigated to test on their impact on posterior distributions of risky cyberattacks. Furthermore, with imprecise probabilities, there is a window to evolve higher Bayesian methods for reducing uncertainty on protection and prediction against cyberattacks.

MAXIM BAZHENOV, UC San Diego
Sleep: from biological to artificial systems

Artificial neural networks are known to exhibit a phenomenon called catastrophic forgetting, where their performance on previously learned tasks deteriorates when learning new tasks sequentially. In contrast, human and animal brains possess the remarkable ability of continual learning, enabling them to incorporate new information while preserving past memories. Empirical evidence indicates that sleep plays a crucial role in the consolidation of recent memories and safeguarding against catastrophic forgetting of previously acquired knowledge. Here we tested the hypothesis that implementing a sleep-like phase in artificial neural networks can protect old memories during new training and alleviate catastrophic forgetting. Sleep was implemented as off-line training with local unsupervised Hebbian plasticity rules and noisy input. In an incremental learning framework, sleep was able to recover old tasks that were otherwise forgotten. Previously learned memories were replayed spontaneously during sleep, forming unique representations for each class of inputs. Representational sparseness and neuronal activity corresponding to the old tasks increased while new task related activity decreased. In the weight space, sleep moved the system towards the region representing the intersection of the loss function minima for individual tasks. Our study sheds light on a potential synaptic weight dynamics strategy employed by the brain during sleep to enhance memory performance for continual learning.

NICK DEXTER, Florida State University
Sample-Efficient Active Learning Strategies for Deep Learning in Scientific Computing

We consider active learning strategies for recovering an unknown object from training data using a given model class. In the active learning scenario, one has the flexibility to choose where to sample the ground truth (or oracle) so as to enhance the generalization performance of the learning algorithm. We introduce a unified framework for this problem that allows for objects in Hilbert spaces, general types of (random) linear measurements as training data and general types of nonlinear model classes. We establish learning guarantees for this framework which provide explicit relations between the amount of training data and properties of the model class to ensure near-best generalization bounds. We demonstrate the efficacy of our framework for gradient-augmented learning with polynomials, Magnetic Resonance Imaging (MRI) using generative models, adaptive sampling for solving PDEs using Physics-Informed Neural Networks (PINNs), and operator learning for uncertainty quantification.

ANTHONY GRUBER, Sandia National Laboratories
Learning metriplectic systems and other bracket-based dynamics

The metriplectic formalism is a useful framework for constructing and explaining phenomenological models of physical phenomena. However, general metriplectic equations of motion are highly complicated, relying on delicate compatibility conditions involving the kernels of algebraic brackets. This talk discusses a recent method for machine-learning provably metriplectic dynamics from data in a way that is (1) universally approximating, (2) admits an error estimate, and (3) scales optimally with respect to the number of learnable parameters. Through finite-dimensional benchmark examples, it is shown that the proposed method is fully expressive and capable of reliably learning metriplectic dynamics, even in cases where only partial state data is observed.

SAMIR KARAM, Concordia University
Physics-informed deep learning and compressive collocation for high-dimensional diffusion-reaction equations

On the forefront of scientific computing, Deep Learning (DL), i.e., machine learning with Deep Neural Networks (DNNs), has emerged a powerful new tool for solving Partial Differential Equations (PDEs). It has been observed that DNNs are particularly well suited to weakening the effect of the curse of dimensionality, a term coined by Richard E. Bellman in the late '50s to describe challenges such as the exponential dependence of the sample complexity, i.e., the number of samples required to solve an approximation problem, on the dimension of the ambient space. However, although DNNs have been used to solve PDEs since the '90s, the literature underpinning their mathematical efficiency in terms of numerical analysis (i.e., stability, accuracy, and sample complexity) is only recently beginning to emerge. In this talk, we leverage recent advancements in function approximation using sparsity-based techniques and random sampling to develop and analyze an efficient high-dimensional PDE solver based on DL. We show, both theoretically and numerically, that it can compete with a novel stable and accurate compressive spectral collocation method. In particular, we demonstrate a new practical existence theorem, which establishes the existence of a class of trainable DNNs with suitable bounds on the network architecture and a sufficient condition on the sample complexity, with logarithmic scaling in dimension, such that the resulting networks stably and accurately approximate a diffusion-reaction PDE with high probability.

KAMYAR KHODAMORADI, University of Regina
Parameterized Approximation for Robust Clustering in Discrete Geometric Spaces

We consider the well-studied Robust (k,z)-Clustering problem, which generalizes the classic k-Median, k-Means, and k-Centre problems. Given a constant $z \geq 1$, the input to Robust (k,z)-Clustering is a set P of n weighted points in a metric space (M,δ) and a positive integer k. Further, each point belongs to one (or more) of the m many different groups $S_1, S_2, \ldots, S_m$. Our goal is to find a set X of k centres such that $\max_{i\in[m]}\{\sum_{p\in S_i} w(p)\delta(p,X)^z\}$ is minimized. This problem arises in the domains of robust optimization [Anthony, Goyal, Gupta, Nagarajan, Math. Oper. Res. 2010] and in algorithmic fairness, for which a tight (under GAP-ETH) $(3^z+\epsilon)$-approximation algorithm exists [Goyal, Jaiswal, Inf. Proc. Letters, 2023].

Motivated by the tight lower bounds for general discrete metrics, we focus on geometric spaces such as the (discrete) high-dimensional Euclidean setting and metrics of low doubling dimension, which play an important role in data analysis applications. First, for a universal constant $\eta_0 > 0.0006$, we devise a $3^z(1-\eta_0)$-factor FPT approximation algorithm for discrete high-dimensional Euclidean spaces thereby bypassing the lower bound for general metrics. We complement this result by showing that even the special case of k-Centre in dimension $\Theta(\log n)$ is $(\sqrt{3/2}-o(1))$-hard to approximate for FPT algorithms. Finally, we complete the FPT approximation landscape by designing an FPT $(1+\epsilon)$-approximation scheme (EPAS) for the metric of sub-logarithmic doubling dimension.

ANASTASIS KRATSIOS, McMaster University and the Vector Institute
Pathwise Generalization bounds for Transformers

We derive non-asymptotic statistical guarantees in this setting through bounds on the *generalization* of a transformer network at a future-time t, given that it has been trained using $N \leq t$ observations from a single perturbed trajectory of a Markov process. Under the assumption that the Markov process satisfies a log-Sobolev inequality, we obtain a generalization bound which effectively converges at the rate of $\mathcal{O}(1/\sqrt{N})$. Our bound depends explicitly on the activation function (Swish, GeLU, or $\tanh$ are considered), the number of self-attention heads, depth, width, and norm-bounds defining the transformer architecture.

Joint work: Blanka Horvath and Yannick Limmer (Oxford Math), Xuwei Yang (McMaster), and Raeid Saqur (U. Toronto and Princeton).

MARTINA NEUMAN, University of Vienna
Efficient Learning Using Spiking Neural Networks Equipped With Affine Encoders and Decoders

We study the learning problem associated with spiking neural networks. Specifically, we consider hypothesis sets of spiking neural networks with affine temporal encoders and decoders and simple spiking neurons having only positive synaptic weights. We demonstrate that the positivity of the weights continues to enable a wide range of expressivity results, including an efficient sorting property, a rate-optimal approximation of smooth functions or approximation without the curse of dimensionality. Moreover, positive-weight spiking neural networks are shown to depend continuously on their parameters which facilitates classical covering number-based generalization statements. Finally, we observe that from a generalization perspective, contrary to feedforward neural networks or previous results for general spiking neural networks, the depth has little to no adverse effect on the generalization capabilities.

OPEN PROBLEM SESSION,

VAKHTANG PUTKARADZE, University of Alberta
Lie-Poisson Neural Networks (LPNets): Data-Based Computing of Hamiltonian Systems with Symmetries

Physics-Informed Neural Networks (PINNs) have received much attention recently due to their potential for high-performance computations for complex physical systems, including data-based computing, systems with unknown parameters, and others. However, applications of these methods to predict the long-term evolution of systems with little friction, such as many systems encountered in space exploration, oceanography/climate, and many other fields, need extra care as the errors tend to accumulate, and the results may quickly become unreliable. We provide a solution to the problem of data-based computation of Hamiltonian systems utilizing symmetry methods. Many Hamiltonian systems with symmetry can be written as a Lie-Poisson system, where the underlying symmetry defines the Poisson bracket. For data-based computing of such systems, we design the Lie-Poisson neural networks (LPNets). We consider the Poisson bracket structure primary and require it to be satisfied exactly, whereas the Hamiltonian, only known from physics, can be satisfied approximately. By design, the method preserves all special integrals of the bracket (Casimirs) to machine precision. LPNets yield an efficient and promising computational method for many particular cases, such as rigid body or satellite motion (the case of SO(3) group), Kirchhoff's equations for an underwater vehicle (SE(3) group), and others. We also discuss symmetry-reduced computations for cases of incomplete symmetry reduction, such as the dynamics of coupled rigid bodies.

Joint work with Chris Eldred (Sandia National Lab), Francois Gay-Balmaz (CNRS and ENS, France), and Sophia Huraka (U Alberta). The work was partially supported by an NSERC Discovery grant.

YIFAN SUN, Stony Brook Institution
Learning over very large graphs

Today, many important learning applications harness data in the form of large graphs. Companies like Amazon, Google, and Facebook use graphs to link similar or related entities; these graphs often have millions or billions of nodes, and are not stored on a single server, but over distributed systems. On the other hand, many graph learning methods today rely on an offline large matrix inversion, where the matrix is the size of the graph itself; this is computationally infeasible in the aforementioned application.

In this talk, we will consider graph learning through the lens of online node label prediction, and its close relationship to fast Laplacian matrix inversion. We introduce *local methods*, of whose complexity is independent of the graph size, and show its promise in large graph learning; the most famous is the approximate page-rank algorithm used in many web applications.

We then discuss the fundamental issues in developing local graph methods, such as acceleration, parallelization, and their integration in scalable large graph learning.

SANDRA ZILLES, University of Regina
Formal Models of Active Learning from Contrastive Examples

Machine learning can greatly benefit from providing learning algorithms with pairs of contrastive training examples—typically pairs of instances that differ only slightly, yet have different class labels. Intuitively, the difference in the instances serves as a means of explaining the difference in the class labels. This presentation proposes a theoretical framework in which the effect of various types of contrastive examples on active learners is studied formally. The focus is on the sample complexity of learning concept classes and how it is influenced by the choice of contrastive examples. Specific concept classes we study consist either of geometric concepts or of Boolean functions. Interestingly, we reveal a connection between learning from contrastive examples and the classical model of self-directed learning. (Joint work with Yuxin Chen, Farnam Mansouri, Hans U. Simon, and Adish Singla.)