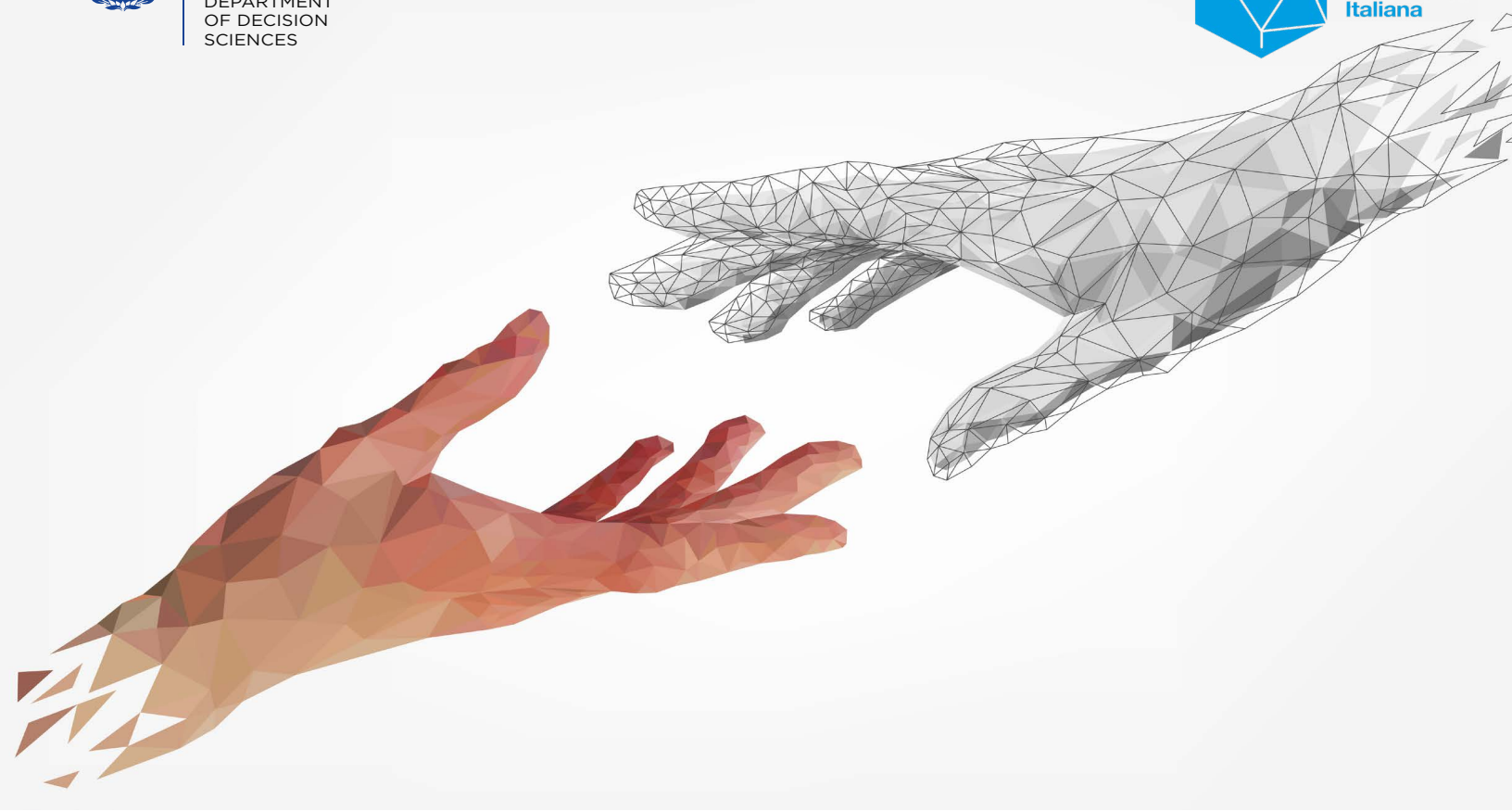
Università Bocconi
DEPARTMENT OF DECISION SCIENCES

Unione Matematica Italiana



MATHEMATICS FOR ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

17-19 JANUARY 2024

Bocconi University | Room AS01-02, Via Röntgen 1 Milano

PROGRAM

CHECK-IN is required to access the event at Bocconi University.

UNIONE MATEMATICA ITALIANA, *Math AI&ML group*
BOCCONI UNIVERSITY, *Department of Decision Sciences*
PRIN 2022 project: *Gradient flows and Non-Smooth Geometric Structures with Applications to Optimization and Machine Learning*

https://dec.unibocconi.eu/mathematics-artificial-intelligence-and-machine-learning

WORKSHOP MATH AI&ML: PROGRAM AT A GLANCE

Wednesday 17			Thursday 18			Friday 19		
			9:00	**Mezard**		9:00	**Giannotti**	
			9:50	**Poggio**		9:50	*Minibreak*	
						10:00	Borghi (AS1)	Quercioli (AS2)
			10:40	*Coffee break*		10:30	Franco (AS1)	D'Inverno (AS2)
			11:10	**Buffa**		11:00	*Coffee break*	
			11:50	**Barra**		11:30	Consolo (AS1)	Hadayat (AS2)
						12:00	Meanti (AS1)	Selicato (AS2)
			12:30	*Lunch*		12:30	Laforgue (AS1)	Cornacchia (AS2)
						13:00	Zhu (AS1)	Malaspina (AS2)
13:00	*Registration*					13:30	*Conclusion, lunch*	
13:45	*Opening*							
14:00	**Quarteroni**		14:00	Papagiannouli (AS1)	Brivio (AS2)			
14:50	*Minibreak*		14:30	Datres (AS1)	Bucarelli (AS2)			
15:00	Molinari (AS1)	Della Santa (AS2)	15:00	Fachechi (AS1)	Mapelli (AS2)			
15:30	Scagliotti (AS1)	Vergani (AS2)	15:30	Alessandrelli (AS1)	Riccobelli (AS2)			
16:00	*Coffee break*		16:00	*Coffee break*				
16:30	Beraha (AS1)	Regazzoni (AS2)	16:30	Camilli (AS1)	Cavinato (AS2)			
17:00	Bonasera (AS1)	Boccardo (AS2)	17:00	Manzan (AS1)	Cambiaghi (AS2)			
17:30	Burzacchi (AS1)	Tomasetto (AS2)	17:30	Albanese (AS1)	Pagani (AS2)			
18:00	Thériault (AS1)	Audone (AS2)	18:00	Centonze (AS1)	Centorrino (AS2)			

Wednesday, January 17

13:00	*Registration*	
13:45	*Opening*	
14:00	**Alfio Quarteroni:** *From Problem Solving to Problem Setting: Revisiting the Role of Mathematicians in the AI Era*	
14:50	*Minibreak*	
	AS1	AS2
15:00	**Cesare Molinari:** *Implicit Regularization*	**Francesco Della Santa:** *Discontinuity Detection via Graph-Informed Neural Networks and Sparse Grids*
15:30	**Alessandro Scagliotti:** *Adversarial training as minimax optimal control problems*	**Andrea Mario Vergani:** *Latent space vector arithmetic for representation learning interpretability in medical imaging*
16:00	*Coffee break*	
16:30	**Mario Beraha:** *Recent advances in distributional data in the Wasserstein space*	**Francesco Regazzoni:** *Scientific Machine Learning augmentation of physics-based models in Computational Cardiology*
17:00	**Lorenzo Bonasera:** *Learning optimal sparselets for event-interval sequence classification*	**Francesco Boccardo:** *Reinforcement Learning with thermal fluctuations at the nano-scale*
17:30	**Arianna Burzacchi:** *Monitoring road infrastructure in developing countries: an object-oriented classification approach from satellite images and road network graphs*	**Matteo Tomasetto:** *Real-time optimal control of parametrized systems b deep learning-based reduced order models*
18:00	**Robin Thériault:** *Dense Associative Memory in the Teacher-Student Setting*	**Gianluca Audone:** *Feature Selection for Time Series: From the Sea's Depths to Space (and beyond)*

Thursday, January 18

9:00	**Marc Mezard:** *Statistical Physics of Generative Diffusion*	
9:50	**Tomaso Poggio:** *The Computable Mathematics of Intelligence*	
10:40	*Coffee break*	
11:10	**Francesca Buffa:** *AI and Big Data in Life Sciences*	
11:50	**Adriano Barra:** *Bio-inspired machine learning*	
12:30	*Lunch*	
	AS1	AS2
14:00	**Katerina Papagiannouli:** *Optimisation analysis with deep linear neural networks using the Bures-Wasserstein Loss*	**Simone Brivio:** *Error estimates for POD-DL-ROMs: a deep learning framework for reduced order modeling of nonlinear parametrized PDEs*
14:30	**Massimiliano Datres:** *A two-scale complexity measure for stochastic neural networks*	**Maria Sofia Bucarelli:** *On Generalization Bounds for Projective Clustering*
15:00	**Alberto Fachechi:** *A Hopfield-like setup to address generalization and overfitting*	**Alessia Mapelli:** *Multi-outcome feature selection via anomaly detection autoencoders for radiogenomic in breast cancer patients*
15:30	**Andrea Alessandrelli:** *Disentangling overlapping inputs with Tripartite Associative Memory: a Statistical Mechanics approach*	**Davide Riccobelli:** *Mathematical modelling of brain tumour growth: model order reduction and patient-specific parameter estimation*
16:00	*Coffee break*	
16:30	**Francesco Camilli:** *Fundamental limits of overparametrized shallow neural networks*	**Lara Cavinato:** *The biological counterpart of radiomics in pancreatic cancer: a preliminary simulation study*
17:00	**Gianluca Manzan:** *Efficiency limits of Restricted Boltzmann Machines in teacher-student frameworks*	**Sara Cambiaghi:** *CT-scan outpatient scheduling under uncertain durations and emergency arrivals*
17:30	**Linda Albanese:** *Learning scheme for dense associative memory*	**Stefano Pagani:** *Physics-informed neural networks for computational medicine in cardiovascular diseases*
18:00	**Martino Centonze:** *Statistical Mechanics of Learning via Reverberation in Bidirectional Associative Memories*	**Veronica Centorrino:** *Biologically Plausible Neural Networks for Sparse Reconstruction: a Normative Framework*

Friday, January 19

9:00	**Fosca Giannotti:** *Towards a synergistic human-machine interaction and collaboration: XAI and Hybrid Decision Making Systems. State-of-the-art and research questions*	
9:50	*Minibreak*	
10:00	**Giacomo Borghi:** *Optimization in Wasserstein space via measure-valued agents*	**Nicola Quercioli:** *A topological model for partial equivariance in deep learning and data analysis*
10:30	**Nicola Rares Franco:** *Deep autoencoders for reduced order modeling of PDEs parametrized by random fields: theoretical results and outlook*	**Giuseppe Alessio D'Inverno:** *VC dimension of Graph Neural Networks with Pfaffian activation functions*
11:00	*Coffee break*	
11:30	**Antonio Consolo:** *Soft regression trees: a model variant and a decomposition training algorithm*	**Fathi Hedayat:** *Selection of functional predictors and smooth coefficient estimation for scalar-on-function regression models*
12:00	**Giacomo Meanti:** *Efficient non-parametric learning with random projections*	**Laura Selicato:** *Bi-level optimization for hyperparameters tuning in sparse low-rank learning algorithms*
12:30	**Pierre Laforgue:** *Multitask Learning with No Regret: from Improved Confidence Bounds to Active Learning*	**Elisabetta Cornacchia:** *Provable Advantage of Curriculum Learning on Parity Targets with Mixed Inputs*
13:00	**Yichen Zhu:** *Fast Approximation and Optimal Contraction for Vecchia Gaussian Processes*	**Greta Malaspina:** *Inexact Newton methods with matrix approximation for nonlinear least-squares and systems*
13:30	*Conclusion, lunch*	

KEYNOTE AND INVITED LECTURES

Adriano Barra:	*Thursday 18, 12:00, AS1*
Francesca Buffa:	*Thursday 18, 11:10, AS1*
Fosca Giannotti:	*Friday 19, 9:00, AS1*
Marc Mezard:	*Thursday 18, 9:00, AS1*
Tomaso Poggio:	*Thursday 18, 9:50, AS1*
Alfio Quarteroni:	*Wednesday 17, 14:00, AS1*

Bio-inspired machine learning

Adriano Barra

Università del Salento (on leave) and Sapienza Università di Roma

Thursday 18, 12:00, AS1

Abstract: I will provide a statistical mechanical picture of shallow networks showing how learning in biological neural networks (e.g. in the Hopfield model) and standard machine learning via gradient descent (e.g. on the Boltzmann machine) ultimately convey the same information after training. This diminishes the gap in our understanding between biological and artificial information processing networks.

AI and Big Data in Life Sciences

Francesca Buffa
Bocconi University, Milano

Thursday 18, 11:10, AS1

Abstract: From genomics to imaging, from biomedical analyses to the ability to predict a patient's response to a particular treatment. If used well, AI can turbocharge researchers and doctors in finding answers to common diseases. I will first provide some general discussion points on the current applications, including critical points and bottlenecks, and then present some of our recent work modelling heterogeneous populations of cells as networks of regulatory networks.

Towards a synergistic human-machine interaction and collaboration: XAI and Hybrid Decision Making Systems. State-of-the-art and research questions

Fosca Giannotti
Scuola Normale Superiore, Pisa

Friday 19, 9:00, AS1

Abstract: Black box AI systems for automated decision making, often based on machine learning over (big) data, map a user's features into a class or a score without exposing the reasons why. This is problematic not only for the lack of transparency, but also for possible biases inherited by the algorithms from human prejudices and collection artifacts hidden in the training data, which may lead to unfair or wrong decisions. The future of AI lies in enabling people to collaborate with machines to solve complex problems. Like any efficient collaboration, this requires good communication, trust, clarity and understanding. Explainable AI addresses such challenges and for years different AI communities have studied such topic, leading to different definitions, evaluation protocols, motivations, and results. Explaining to humans how AI reasons is only a part of the problem, we must then be able to design AI systems that understand and collaborate with humans: Hybrid decision-making systems aim at leveraging the strengths of both human and machine agents to overcome the limitations that arise when either agent operates in isolation.

This lecture provides a reasoned introduction to the work of Explainable AI (XAI) to date, A special focus will be on paradigms in support of synergistic human-machine interaction and collaboration to improve joint performance in high-stake decision-making as for example methods aimed at engaging users with factual and counterfactual or other high-level explanations. The lecture will highlight the steps needed for promoting such collaboration and seamless interaction maintaining the human responsibility of the choice through a progressive disclosure to prevent cognitive overload. We will discuss three distinct paradigms of collaboration, characterized by a different degree of human agency will be discussed: i) human oversight, with a human expert monitoring AI prediction; ii) learning to defer, in which the machine learning model is given the possibility to abstain from making a prediction when it receives an instance where the risk of making a misprediction is too large; iii) collaborative and interactive learning, in which human and AI engage in communication to integrate their distinct knowledge and facilitate the human's ability to make informed decisions.

This lecture is a joint work with: Clara Punzi, Mattia Setzu and Roberto Pellungrini.

Statistical Physics of Generative Diffusion

Marc Mezard
Bocconi University, Milano

Thursday 18, 9:00, AS1

Abstract: Generative models, in which one trains an algorithm to generate samples 'similar' to those of a data base, is a major new direction developed in machine learning in the recent years. In particular, generative models based on diffusion equations have become the state of the art, notably for image generation. However, the reasons for this spectacular technological success are not well understood, and neither are its limitations.

After an introduction to this topic, the talk will focus on the behavior of generative diffusion in the high-dimensional limit, where data are formed by a very large number of variables. Using methods from statistical physics, we explain the various dynamical regimes that occur during the generation.

The Computable Mathematics of Intelligence

Tomaso Poggio
MIT, Cambridge

Thursday 18, 9:50, AS1

Abstract: Is mathematics invented or discovered? Why it has such an unreasonable ef- fectiveness in describing aspects of our physical world? Recent progress in formulating fundamental principles underlying the stunning success of deep learning (DL) provides a new light on this age-old questions. I will discuss in particular why deep network seem to escape the curse of dimensionality. The answer lies in a key property of all functions that are efficiently Turing computable: they are compositionally sparse. This property enables the use of deep (and sparse) networks — the engine powering deep networks and more recent systems such as LLMs.

From Problem Solving to Problem Setting
Revisiting the Role of Mathematicians in the AI Era

Alfio Quarteroni

Politecnico di Milano, EPFL, Lausanne, Accademia dei Lincei

Wednesday 17, 14:00, AS1

Abstract: Problem setting is a critical precursor to problem solving. It involves the art of formulating the right problem statement. The importance of this phase is underscored by the fact that without a well-defined problem, finding the right tools and techniques for problem solution becomes a cumbersome and often futile endeavor. This transition from problem setting to problem solving is integral to the larger paradigm of knowledge development. While AI tools have made tremendous strides in recent years, they remain dependent on the foundation laid by human intelligence. Mathematicians, with their ability to discern patterns and relationships, data and variables, play a vital role in this stage.

CONTRIBUTED TALKS

Linda Albanese:	*Thursday 18, 17:30, AS1*
Andrea Alessandrelli:	*Thursday 18, 15:30, AS1*
Gianluca Audone:	*Wednesday 17, 18:00, AS2*
Mario Beraha:	*Wednesday 17, 16:30, AS1*
Francesco Boccardo:	*Wednesday 17, 17:00, AS2*
Lorenzo Bonasera:	*Wednesday 17, 17:00, AS1*
Giacomo Borghi:	*Friday 19, 10:00, AS1*
Simone Brivio:	*Thursday 18, 14:00, AS2*
Maria Sofia Bucarelli:	*Thursday 18, 14:30, AS2*
Arianna Burzacchi:	*Wednesday 17, 17:30, AS1*
Sara Cambiaghi:	*Thursday 18, 17:00, AS2*
Francesco Camilli:	*Thursday 18, 16:30, AS1*
Lara Cavinato:	*Thursday 18, 16:30, AS2*
Martino Centonze:	*Thursday 18, 18:00, AS1*
Veronica Centorrino:	*Thursday 18, 18:00, AS2*
Antonio Consolo:	*Friday 19, 11:30, AS1*
Elisabetta Cornacchia:	*Friday 19, 12:30, AS2*
Giuseppe Alessio D'Inverno:	*Friday 19, 10:30, AS2*
Massimiliano Datres:	*Thursday 18, 14:30, AS1*
Francesco Della Santa:	*Wednesday 17, 15:00, AS2*
Alberto Fachechi:	*Thursday 18, 15:00, AS1*
Nicola Rares Franco:	*Friday 19, 10:30, AS1*
Fathi Hedayat:	*Friday 19, 11:30, AS2*
Pierre Laforgue:	*Friday 19, 12:30, AS1*
Greta Malaspina:	*Friday 19, 13:00, AS2*
Gianluca Manzan:	*Thursday 18, 17:00, AS1*
Alessia Mapelli:	*Thursday 18, 15:00, AS2*
Giacomo Meanti:	*Friday 19, 12:00, AS1*
Cesare Molinari:	*Wednesday 17, 15:00, AS1*
Stefano Pagani:	*Thursday 18, 17:30, AS2*
Katerina Papagiannouli:	*Thursday 18, 14:00, AS1*
Nicola Quercioli:	*Friday 19, 10:00, AS2*
Francesco Regazzoni:	*Wednesday 17, 16:30, AS2*
Davide Riccobelli:	*Thursday 18, 15:30, AS2*
Alessandro Scagliotti:	*Wednesday 17, 15:30, AS1*
Laura Selicato:	*Friday 19, 12:00, AS2*
Robin Thériault:	*Wednesday 17, 18:00, AS1*
Matteo Tomasetto:	*Wednesday 17, 17:30, AS2*
Andrea Mario Vergani:	*Wednesday 17, 15:30, AS2*
Yichen Zhu:	*Friday 19, 13:00, AS1*

Learning scheme for dense associative memory

Linda Albanese
University of Salento

Thursday 18, 17:30, AS1

Abstract: In this talk we focus on the dense associative memory, a neural network introduced by Hopfield and Krotov in 2018. Then we move to the description of two possible extensions of it, namely the aforementioned network with the integration of unsupervised and supervised settings, two different learning schemes well-known in machine learning. We solve the model through Guerra's interpolation, a technique introduced for the first time by Francesco Guerra for spin glasses and later imported to neural networks; in the end, we focus on their mainly features.

Disentangling overlapping inputs with Tripartite Associative Memory: a Statistical Mechanics approach

Andrea Alessandrelli
Università di Pisa

Thursday 18, 15:30, AS1

Abstract: This work explores the synergy of Artificial Intelligence and Statistical Mechanics, focusing on the Bidirectional-Associative Memory (BAM) as a hetero-associative extension of the Hopfield model. We extend BAM to handle three or more patterns simultaneously, successfully applying it to taxonomic datasets. Leveraging statistical mechanics, interpolation, and phase diagrams, we characterize key computational features and optimize network pre configuration. Numerical results on hierarchical synthetic datasets demonstrate the model's ability to disentangle input signals, with theoretical predictions aligning perfectly with simulations. This Statistical Mechanics approach not only configures network parameters optimally but also holds promise for à priori optimization of deep learning architectures linking it with the internal structure of the dataset under analysis.

Feature Selection for Time Series: From the Sea's Depths to Space (and beyond)

Gianluca Audone
Politecnico di Torino

Wednesday 17, 18:00, AS2

Abstract: Feature selection plays a crucial role in prediction algorithms. The aim of this talk is to build a robust framework and give a set of tools to conduct the exploratory data analysis needed when dealing with Time Series. The starting point will be ergodic stationary processes through which we rigorously define time series. We introduce the methodologies for trend identification and seasonality estimation to then focus on the insights given from auto-correlation and cross-correlation analyses which help understanding the interplay and dependencies among different features.

Using data from the Comprehensive Nuclear Test Ban Treaty Organization (CTBTO) and space weather data we show some result given by the introduced framework. CTBTO operates a global International Monitoring System, with 11 hydroacoustic stations around the globe located in the deep-sea sound channel. Continuous measurements provide up to 20 years of sound pressures at frequencies of up to 100 Hz. These relatively long timescales allow investigating the effects of climate over that period. The solar activity recorded over 14 years consisting of solar wind, geomagnetic and energetic indices measured in situ L1 and, magnetospheric ring current intensity measured on Earth.

This presentation offers insights into the intricate realm of climate studies and space weather prediction, showcasing the significance of meticulous data exploration and feature selection in harnessing the predictive power of time series data.

Recent advances in distributional data in the Wasserstein space

Mario Beraha
Politecnico di Milano

Wednesday 17, 16:30, AS1

Abstract: This talk provides an overview of recent developments in statistical analysis of probability distributions using the 2-Wasserstein metric. Focusing on distribution on the real line, we introduce a novel class of projected methods for Principal Component Analysis (PCA) and regression within the Wasserstein space. Compared to previously proposed methods, ours enjoy comparable performance at a fraction of the computational cost. We then explore the case of measures supported on the unit circle, developing a a framework for PCA, which builds on a detailed investigation of the optimal transportation problem for measures on the unit circle.

Reinforcement Learning with thermal fluctuations at the nano-scale

Francesco Boccardo

MaLGa, Department of Civil, Chemical and Environmental Engineering, University of Genoa

Wednesday 17, 17:00, AS2

Abstract: Learning to choose actions to achieve a goal is a fundamental process in engineering sciences such as robotics and in biology for individual and collective behavior. This process can be modeled within the framework of Reinforcement Learning. However, for micro- and nano-scale devices, thermal fluctuations hamper the learning process.

We show that in this regime, while optimal actions should bring an improvement proportional to the small ratio of the applied force times a length-scale over the temperature, the learned improvement is smaller and proportional to the square of this small ratio. Consequently, the efficiency of learning, which compares the learning improvement to the theoretical optimal improvement, drops to zero. Nevertheless, we show that one can circumvent these limitations by learning at a lower temperature and transferring the learned actions to the high-temperature regime.

Our results are illustrated with simulations of the control of the shape of small particle clusters by means of an external field, and should apply to a wide class of Markov Decision Processes such as nano-navigation, actuation of nano-machines, and the controlled organization of assemblies of colloids, active particles or nano-robots.

Our results rely on the use of a full numerical solution of the optimization problem obtained by Dynamic Programming, in order to evaluate and interpret the performance of basic Reinforcement Learning algorithms.

Learning optimal sparse-lets for event-interval sequence classification

Lorenzo Bonasera
Università di Pavia

Wednesday 17, 17:00, AS1

Abstract: Event-interval sequences are multivariate series of events that occur over time. The classification of event-interval sequences has gained increasing attention among researchers in the field of time series analysis due to their broad applicability. This paper focuses on the optimal extraction of interpretable features from event-interval sequences to construct classifiers. The current state-of-the-art interpretable feature is called e-lets, that is the event-interval counterpart of time series shapelets. In this work, we propose a novel approach to interpretable classification of event-interval sequences based on sparse-lets, a new generalization of the e-lets temporal feature. Our approach utilizes genetic algorithms to learn optimal sparse-lets. By producing sparse and interpretable features, our approach improves the trustworthiness and robustness of its predictions. We evaluate the performance of our method through experiments conducted on benchmark datasets and compare our results against other state-of-the-art methods. The experimental results demonstrate that our method is a viable competitor in terms of performance while also providing trustworthy and interpretable outcomes.

Optimization in Wasserstein space via measure-valued agents

Giacomo Borghi
RWTH Aachen University

Friday 19, 10:00, AS1

Abstract: Optimization over probability measures is a common tasks in Machine Learning as many objects of interest like images, data clouds, or uncertainties, can be modelled as measures. Recent works aim to implement a gradient flow dynamics defined in the Wasserstein space to design first-order local optimization methods.

In this talk, we present a novel approach based on an interacting multi-agent system for global optimization. The measure-valued agents self-organize through a consensus-type dynamics to concentrate around a global minimum of the objective functional. Numerically, the dynamics is implemented via efficient optimal transport solvers. We will also discuss an application to variational inference problems and future research directions.

The talk is based on joint work with Michael Herty, Andrey Stavitskiy and José Carrillo.

Error estimates for POD-DL-ROMs: a deep learning framework for reduced order modeling of nonlinear parametrized PDEs

Simone Brivio

MOX - Dipartimento di Matematica - Politecnico di Milano

Thursday 18, 14:00, AS2

Abstract: POD-DL-ROMs represent a novel framework in the field of reduced order modeling involving (i) a prior linear dimensionality reduction through Proper Orthogonal Decomposition (POD) in pursuance of computational efficiency, (ii) an autoencoder to achieve a nonlinear compression of the POD coefficients into a set of few latent coordinates, and (iii) a dense neural network to reconstruct the function mapping the parameters onto the latent coordinates. In this context, we aim at providing practical criteria to suitably design the POD-DL-ROM architecture and appropriately tune the most impactful hyperparameters. To do that, we propose rigorous error estimates for POD-DL-ROM, shedding light over three main contributions, namely, the sampling criterion employed to generate the training dataset, the influence of the POD dimension, and the complexity of the neural network in terms of number of active weights.

On Generalization Bounds for Projective Clustering

Maria Sofia Bucarelli
Sapienza University of Rome

Thursday 18, 14:30, AS2

Abstract: Given a set of points, clustering consists of finding a partition of a point set into clusters such that the center to which a point is assigned is as close as possible. Most commonly, centers are points themselves, which leads to the famous -median and -means objectives. One may also choose centers to be dimensional subspaces, which gives rise to subspace clustering. In this paper, we consider learning bounds for these problems. That is, given a set of samples drawn independently from some unknown, but fixed distribution , how quickly does a solution computed on converge to the optimal clustering of ? We give several near optimal results. In particular, For center-based objectives, we show a convergence rate of $\tilde{O}\left(\sqrt{k/n}\right)$. This matches the known optimal bounds of [Fefferman, Mitter, and Narayanan, Journal of the Mathematical Society 2016] and [Bartlett, Linder, and Lugosi, IEEE Trans. Inf. Theory 1998] for k-means and extends it to other important objectives such as k-median. For subspace clustering with j-dimensional subspaces, we show a convergence rate of $\tilde{O}\left(\sqrt{\frac{kj^2}{n}}\right)$. These are the first provable bounds for most of these problems. For the specific case of projective clustering, which generalizes k-means, we show a convergence rate of $\Omega\left(\sqrt{\frac{kj}{n}}\right)$ is necessary, thereby proving that the bounds from [Fefferman, Mitter, and Narayanan, Journal of the Mathematical Society 2016] are essentially optimal.

Monitoring road infrastructure in developing countries: an object-oriented classification approach from satellite images and road network graphs

Arianna Burzacchi
Politecnico di Milano

Wednesday 17, 17:30, AS1

Abstract: The information on road pavement is fundamental for the development of an efficient and resilient transportation system, but it is rarely available in road network databases of developing countries. The present research, developed within the Safari Njema project of Politecnico di Milano, aims at monitoring the road pavement infrastructure in African cities and supporting the design of strategic solutions for paratransit urban mobility. In this talk, I will present a supervised classification method for road pavement that recognizes surface type of road segments starting from satellite images and geographical locations in the road network. From the former, pixels related to street surface are extracted in a two-step data-driven process, and the latter is used to build a road network graph where nodes are roads and edges are present whenever two roads are adjacent. Subsequently, surface labels are predicted as paved or unpaved employing k-Nearest Neighbors (k-NN) algorithm for graphs of objects: each road is mathematically modeled as an object representing the distribution of its pixels in the RGB space; then such objects serve as nodes of the graph of the road network. The k-NN classification of road type is then performed by looking at the most similar roads in terms of both similarity between objects and closeness in the graph of objects. The proposed approach is proven to be accurate, low-cost, and can be straightforwardly extended from the single case study of Greater Maputo to other cities.

CT-scan outpatient scheduling under uncertain durations and emergency arrivals

Sara Cambiaghi
University of Pavia

Thursday 18, 17:00, AS2

Abstract: This study aims to optimize the scheduling of Computed Tomography (CT) exams at the Emergency Department (ED) of the Radiodiagnostics Institute of Policlinico San Matteo in Pavia, Italy. The case study is characterized by the presence of different patient flows, that is outpatients, inpatients, and emergencies, sharing the same resources. These patient flows differ mainly in the dynamicity of their scheduling, with outpatients being placed on a waiting list and their appointment planned weekly, inpatients being managed in real time to guarantee their exam within 24-48 hours, and emergency patients who have the maximum priority and must be reported as soon as possible to guarantee a short ED length of stay.

We present a preliminary study that focuses on the outpatient scheduling problem, which is formalized through an optimization model using predictions provided by machine learning to deal with the main uncertainty factors. Firstly, we estimate the execution and reporting time for each exam based on CT exams conducted in the period between June 2021 and November 2022. Interpretable and non-interpretable regression algorithms were utilized to describe the current situation and predict the duration of execution and reporting activities. The insights gained from the data analysis are exploited by a multi-objective mathematical programming model to minimize direct waiting times for outpatient patients and completion times for emergency patients from the ED and the intensive care unit.

A quantitative analysis assesses the effectiveness of the proposed optimization and machine learning approach to improve the efficiency of the CT-scan service concerning different possible preferences of decision-makers taking into account the most important patient-centered and facility-centered performance indices.

Fundamental limits of overparametrized shallow neural networks

Francesco Camilli
The Abdus Salam International Centre for Theoretical Physics

Thursday 18, 16:30, AS1

Abstract: In this talk I shall provide an information-theoretical analysis of a two-layer neural network trained on a relatively small dataset compared to the network size. The dataset is generated by a teacher network with the same architecture. The main finding of this study is the asymptotic equivalence of the two-layer neural network and a suitably tuned one-layer network, i.e. a Generalized Linear Model, for which the Mutual Information between the weights and the training set is known. This result in turn yields the Bayes-optimal generalization error, which serves as a lower bound for any neural network with the mentioned architecture. The proof relies on rigorous Mathematical Physics tools used in the study of spin glasses, such as the interpolation scheme, and it is guided by recently conjectured Gaussian Equivalence Principles. With respect to the existing literature, which is either non-rigorous, or restricted to the case of the learning of the last layer's weights only, the proof addresses the learning of all the network parameters. While our techniques are primarily applicable to the regime of small datasets, it offers the advantage of being self-contained, simple, and it constitutes an independent proof of a Gaussian Equivalence Principle.

Based on a joint work with D. Tieplova and J. Barbier: arXiv:2307.05635.

The biological counterpart of radiomics in pancreatic cancer: a preliminary simulation study

Lara Cavinato
Politecnico di Milano

Thursday 18, 16:30, AS2

Abstract: Radiomics extracts quantitative features from medical images and use them to predict clinical outcomes. However, to fully realize the potential of radiomics, these features need to be associated with a biological meaning. Linking radiomics features to underlying biological processes is still an open matter, yet crucial to achieve a deeper understanding of disease mechanisms and identify potential biomarkers. In this work, we employ a multidisciplinary approach that integrates radiomics and other fields, to correlate radiomics features with biological characteristics of different tissues affected by pancreatic cancer at different stages. We developed and implemented a simulation framework for estimating the spatiotemporal uptake throughout the tissues of the 18F-fluorothymidine tracer ([18F]FLT). The framework consisted of several steps including the creation of the computational domain from CD31 WSI and the implementation of a spatiotemporal model to describe the [18F]FLT uptake via Partial Differential Equations (PDE). From time-varying uptake maps of the tissues, we simulated the dynamic PET imaging and extracted 43 radiomic feature maps describing the tissue texture. Each of these feature maps was correlated with tissue biological characteristics (spatial location of vessel, stroma, and cell density), cell proliferation index and overall tracer uptake, and the overall tracer uptake with carcinoma grade describing tumour aggressiveness. We used Pearson correlation to quantify the relationship between the variables. The proposed framework represents a sound pipeline for evaluating the biological meaning of radiomics. With a properly extended sample size, this work can provide valuable insights into the biological counterpart of the texture analysis.

Statistical Mechanics of Learning via Reverberation in Bidirectional Associative Memories

Martino Centonze
University of Salento

Thursday 18, 18:00, AS1

Abstract: Statistical Mechanics provides useful mathematical techniques that can be used to derive important information about the capacity of a certain type of neural networks, the energy based models. These models are a type of Associative Memories which are nowadays receiving new attention, given their link with RBMs and modern Transformers, which lie at the heart of Deep Learning. A new approach adopts the Guerra's interpolation technique to specifically derive the phase diagram of Associative Memories in the learning phase, where the network stores its memories in the available basins of attraction of its energy function. In this talk I present new results in this direction, with the application of these concepts to the Bidirectional Associative Memory (BAM), first introduced by Kosko in the 80's. The BAM is a bipartite model able to store and retrieve coupled patterns, following Pavlov's Classical Conditioning. A full statistical mechanical picture of supervised and unsupervised learning is provided (at the replica symmetric level of description) obtaining phase diagrams, thresholds for learning and signal-to-noise outcomes. The results are confirmed by Monte Carlo simulations and offer a complete viewpoint on the functioning of this network architecture.

Biologically Plausible Neural Networks for Sparse Reconstruction: a Normative Framework

Veronica Centorrino
Scuola Superiore Meridionale

Thursday 18, 18:00, AS2

Abstract: We present a normative top/down framework to analyze and design continuous-time biologically plausible recurrent neural networks tackling sparse reconstruction problems. These problems, which are ubiquitous across a wide range of domains from signal processing to neuroscience, involve approximating a given input stimulus from a dictionary using a set of sparse (active) neurons and can be formalized as a regularized least squares optimization problem. First, by leveraging monotone operator theory, we relate the solutions of this problem to the equilibria of the recently introduced proximal gradient dynamics, which can be interpreted as a continuous-time recurrent neural network. Then, we employ contraction theory - a powerful control-theoretical tool - to characterize the convergence properties of the dynamics. Specifically, we show that: (i) the dynamics is positive and hence state variables can be interpreted as firing rates; (ii) under a suitable assumption on the dictionary, it exhibits linear-exponential convergence, i.e., the convergence rate is at worst linear and then, after a transient, it becomes exponential. The talk is then concluded by exploring the possibility of embedding Hebbian learning in the framework. Throughout the talk, the results are illustrated via numerical experiments.

References:

[1] V. Centorrino, A. Gokhale, A. Davydov, G. Russo, and F. Bullo, "Positive Competitive Networks for Sparse Reconstruction", 2023, https://arxiv.org/abs/2311.03821

[2] V. Centorrino, A. Gokhale, A. Davydov, G. Russo, and F. Bullo, "Euclidean Contractivity of Neural Networks with Symmetric Weights", IEEE Control Systems Letters, March 2023. https://ieeexplore.ieee.org/document/10130086/citations?tabFilter=papers#citations

[3] A. Davydov, V. Centorrino, A. Gokhale, G. Russo, and F. Bullo, "Contracting dynamics for time-varying convex optimization", 2023, https://arxiv.org/abs/2305.15595

[4] V. Centorrino, F. Bullo, and G. Russo, "Modelling and contractivity of neural-synaptic networks with Hebbian learning", 2022, https://arxiv.org/abs/2204.05382

[5] V. Centorrino, F. Bullo, and G. Russo, "Contraction Analysis of Hopfield Neural Networks with Hebbian Learning", IEEE Conf. on Decision and Control, Cancun, Mexico, December 2022, https://ieeexplore.ieee.org/abstract/document/9993009

Soft regression trees: a model variant and a decomposition training algorithm

Antonio Consolo
Politecnico di Milano

Friday 19, 11:30, AS1

Abstract: Decision trees are widely used for classification and regression tasks arising in a variety of application fields. In [3] Blanquero et al. recently proposed a continuous nonlinear optimization formulation to train multivariate soft regression trees, which can account for sparsity and fairness. For any given input vector, the prediction is defined as the summation of the leaf nodes outputs (linear regressions over the input features) weighted by the probability that the input vector falls into the corresponding leaf node. We propose and investigate a soft regression tree variant where, for every input vector, we have a linear prediction for every leaf node with the associated probability, and the actual prediction is given by the leaf node reached by following from the root the branches with the highest probability. Our formulation is well-suited not only to decomposition but also to impose fairness constraints. The decomposition training algorithm we present includes an ad-hoc initialization strategy and a heuristic for the reassignment of the input vectors along the branch nodes of the tree. Under mild assumptions, we also establish asymptotic convergence guarantees. The results obtained on 15 datasets from the UCI and KEEL repositories indicate that our model variant and decomposition algorithm yield trees with higher accuracy than the formulation in [3], and lead to significant speed-up in training time and similar accuracy compared with the discrete optimization approach described by Bertsimas and Dunn in [1,2].

References:

[1] Dunn, J.W.: Optimal trees for prediction and prescription, PhD Thesis, Massachusetts Institute of Technology (2018).

[2] Bertsimas, D. and Dunn, J.: Machine learning under a modern optimization lens. Dynamic Ideas LLC (2019).

[3] Blanquero, R., Carrizosa, E., Molero-Rìo, C. and Morales, D.R.: On sparse optimal regression trees. European Journal of Operational Research, 299(3),1045-1054 (2022).

Provable Advantage of Curriculum Learning on Parity Targets with Mixed Inputs

Elisabetta Cornacchia
MIT

Friday 19, 12:30, AS2

Abstract: Experimental results have shown that curriculum learning, i.e., presenting simpler examples before more complex ones, can improve the efficiency of learning. Some recent theoretical results also showed that changing the sampling distribution can help neural networks learn parities, with formal results only for large learning rates and one-step arguments. Here we show a separation result in the number of training steps with standard (bounded) learning rates on a common sample distribution: if the data distribution is a mixture of sparse and dense inputs, there exists a regime in which a 2-layer ReLU neural network trained by a curriculum noisy-GD (or SGD) algorithm that uses sparse examples first, can learn parities of sufficiently large degree, while any fully connected neural network of possibly larger width or depth trained by noisy-GD on the unordered samples cannot learn without additional steps. We also provide experimental results supporting the qualitative separation beyond the specific regime of the theoretical results.

VC dimension of Graph Neural Networks with Pfaffian activation functions

Giuseppe Alessio D'Inverno
University of Siena

Friday 19, 10:30, AS2

Abstract: Graph Neural Networks (GNNs) have emerged in recent years as a powerful tool to learn tasks across a wide range of graph domains in a data-driven fashion; based on a message passing mechanism, GNNs have gained increasing popularity due to their intuitive formulation, closely linked with the Weisfeiler-Lehman (WL) test for graph isomorphism, to which they have proven equivalent [1, 2]. From a theoretical point of view, GNNs have been shown to be universal approximators, and their generalization capability (namely, bounds on the VC dimension [3]) has recently been investigated for GNNs with piecewise polynomial activation functions [4]. The aim of our work is to extend this analysis on the VC dimension of GNNs to other commonly used activation functions, such as sigmoid and hyperbolic tangent, using the framework of Pfaffian function theory. Bounds are provided with respect to the usual parameters (depth, number of neurons, input size) as well as with respect to the number of colors resulting from the 1-WL test applied on the graph domain. The theoretical analysis is supported by a preliminary experimental study.

References:

[1] Christopher Morris, Martin Ritzert, Matthias Fey, William L Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and leman go neural: Higher-order graph neural networks. In Proceedings of the AAAI conference on artificial intelligence, volume 33, pages 4602–4609, 2019.

[2] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? arXiv preprint arXiv:1810.00826, 2018.

[3] Franco Scarselli, Ah Chung Tsoi, and Markus Hagenbuchner. The vapnik–chervonenkis dimen- sion of graph and recursive neural networks. Neural Networks, 108:248–259, 2018.

[4] Christopher Morris, Floris Geerts, Jan Tönshoff, and Martin Grohe. WI meet VC. arXiv preprint arXiv:2301.11039, 2023.

A two-scale complexity measure for stochastic neural networks

Massimiliano Datres
University of Trento

Thursday 18, 14:30, AS1

Abstract: Over-parametrized deep learning models are achieving outstanding performances in solving several complex tasks such as image classification problems, object detection and natural language processing. Despite the risk of overfitting, these parametric models show impressive generalization after training. Hence, defining appropriate complexity measures becomes crucial for understanding and quantifying the generalization capabilities of deep learning models. In this talk, I will introduce a new notion of complexity measure, called two-scale effective dimension (2sED), which is a box-covering dimension related to a metric induced by the Fisher information matrix of the parametric model. I will then show how the 2sED can be used to derive a generalization bound. Furthermore, I present an approximation of the 2sED for Markovian models, called lower 2sED, that can be computed sequentially layer-by-layer with less computational demands. Finally, I present experimental evidence that the post-training performance of given parametric models is related both with 2sED and the lower 2sED.

Discontinuity Detection via Graph-Informed Neural Networks and Sparse Grids

Francesco Della Santa
Politecnico di Torino

Wednesday 17, 15:00, AS2

Abstract: In this presentation, we consider the task of detecting the discontinuity interfaces of piece-wise continuous functions, a challenge of primary importance for many scientific and engineering domains. Our focus is on identifying these interfaces, drawing inspiration from previous works based on sparse grids (e.g., Jakeman et al. 2011) and Convolutional Neural Networks (Wang et al. 2022).

In recent years, we observed the integration of Neural Networks (NNs) into discontinuity detection problems, such as the latter work aforementioned. Building on the direction of Wang et al., our work modifies and extends their methodology, moving from regular grids and Convolutional Neural Networks (CNNs) to sparse grids and NNs based on graphs. The innovation lies in the introduction of Graph-Informed Neural Networks (GINNs, Berrone et al. 2022), trained to detect troubled points in a sparse grid by leveraging the graph structure inherent in these grids.

In this presentation we illustrate how GINNs take advantage of the sparse grid's graph structure. The methodology, introduces a recursive algorithm for sparse grid-based detectors, demonstrating finite termination and convergence properties. Trained on synthetic data generated by deterministic detectors, GINNs prove to be a cost-effective and fast discontinuity detection method, especially when applied to high-dimensional functions.

Experiments validating the proposed approach show the efficiency and generalization abilities of GINNs in detecting discontinuity interfaces, even for functions beyond those used in training.

A Hopfield-like setup to address generalization and overfitting

Alberto Fachechi

Department of Mathematics G. Castelnuovo, Sapienza Università di Roma

Thursday 18, 15:00, AS1

Abstract: In this talk, we consider a Hopfield-like model emerging as a fixed point solution of a machine learning procedure with minimal technical requirements. In this scenario, L2-regularization for the network parameters can be mapped to the "dreaming" mechanism in associative neural networks. With this relation, we can apply standard techniques of machine retrieval for analyzing the generalization capabilities of the associative network and address the emergence of the overfitting phenomenon.

Deep autoencoders for reduced order modeling of PDEs parametrized by random fields: theoretical results and outlook

Nicola Rares Franco
MOX, Department of Mathematics, Politecnico di Milano

Friday 19, 10:30, AS1

Abstract: Deep Learning is having a remarkable impact on the literature of Reduced Order Models (ROMs) for Partial Differential Equations (PDEs), offering powerful alternatives capable of addressing complex problems where traditional methods may fall short. A crucial role in this context is played by deep autoencoders, which, by leveraging on the nonlinear capabilities of neural networks, can significantly enhance the dimensionality reduction process. Under this paradigm, several successful approaches have emerged, collectively known as Deep Learning-based Reduced Order Models (DL-ROMs). However, when it comes to stochastic problems parameterized by random fields, the current understanding of DL-ROMs is mostly based on empirical evidence: in fact, their theoretical analysis is currently limited to the case of PDEs depending on a finite number of (deterministic) parameters. In this talk, we discuss a few extensions of such theory, focusing on those cases in which the stochasticity is generated by random fields. In particular, we derive explicit error bounds that can guide domain practitioners in the complex task of designing effective deep autoencoders. We conclude by presenting some numerical experiments, and by anticipating some novel results about secondary aspects (expressivity of convolutional neural networks, training strategies etc.).

Selection of functional predictors and smooth coefficient estimation for scalar-on-function regression models

Fathi Hedayat

Department of Operations and Decision Systems, Université Laval (Québec, Canada)

Friday 19, 11:30, AS2

Abstract: Despite their importance in obtaining reliable and interpretable models, variable selection techniques are still underdeveloped in the framework of scalar-on-function regression models – in which several functional variables are employed to predict a scalar response. The available functional variable selection methods are generally based on a grouping strategy, which consists in transforming each functional coefficient in a group of scalars through basis expansion. This approach lacks theoretical guarantees on the selection of the relevant predictors and is not effective when a large number of variables are present. We propose a methodology for selecting relevant functional predictors, while simultaneously providing accurate smooth (or, more in general, regular) estimates of the functional coefficients. We suppose that the functional predictors lie in a real separable Hilbert space, while the functional coefficients belong to a specific subspace of this Hilbert space. Such subspace can be a Reproducing Kernel Hilbert Space (RKHS). Importantly, this ensures the desired regularity characteristics, such as smoothness or periodicity, for the functional-coefficient estimates. Coefficient estimates are obtained by an adaptive penalized least square algorithm which employs functional subgradients to efficiently solve the minimization problem and allows us to establish asymptotic properties of our estimators. In particular, we prove that our method satisfies the functional oracle property. Performance in terms of variable selection and accuracy of coefficient estimation is assessed through simulations.

In collaboration with: Marzia A. Cremona (Department of Operations and Decision Systems), Federico Severino (Department of Finance, Insurance and Real Estate), Université Laval (Québec, Canada).

Keywords: variable selection; functional data analysis; scalar-on-function regression; RKHS.

Multitask Learning with No Regret: from Improved Confidence Bounds to Active Learning

Pierre Laforgue
Università degli Studi di Milano

Friday 19, 12:30, AS1

Abstract: Multitask learning is a powerful framework that enables one to simultaneously learn multiple related tasks by sharing information between them. Quantifying uncertainty in the estimated tasks is of pivotal importance for many downstream applications, such as online or active learning. In this work, we provide novel multitask confidence intervals in the challenging agnostic setting, i.e., when neither the similarity between tasks nor the tasks' features are available to the learner. The obtained intervals do not require i.i.d. data and can be directly applied to bound the regret in online learning. Through a refined analysis of the multitask information gain, we obtain new regret guarantees that, depending on a task similarity parameter, can significantly improve over treating tasks independently. We further propose a novel online learning algorithm that achieves such improved regret without knowing this parameter in advance, i.e., automatically adapting to task similarity. As a second key application of our results, we introduce a novel multitask active learning setup where several tasks must be simultaneously optimized, but only one of them can be queried for feedback by the learner at each round. For this problem, we design a no-regret algorithm that uses our confidence intervals to decide which task should be queried. Finally, we empirically validate our bounds and algorithms on synthetic and real-world (drug discovery) data.

Inexact Newton methods with matrix approximation for nonlinear least-squares and systems

Greta Malaspina

Dipartimento di ingegneria industriale, Università degli studi di Firenze

Friday 19, 13:00, AS2

Abstract: We present linesearch inexact Newton based methods for nonlinear systems of equations and nonlinear least-squares problems. At each iteration of the methods, the random models are formed by approximating the matrices involved in the deterministic local models using sampling and sketching strategies. We analyze the proposed methods by studying the expected number of iterations needed to reach a desired level of accuracy in the first-order optimality conditions, under suitable assumption on the quality of the stochastic approximation. Finally, we study the numerical performance of the methods and compare it with their deterministic counterparts.

Efficiency limits of Restricted Boltzmann Machines in teacher-student frameworks

Gianluca Manzan
Department of Mathematics, University of Bologna

Thursday 18, 17:00, AS1

Abstract: Unsupervised Machine learning with Boltzmann machines is the inverse problem of finding a suitable Gibbs measure to approximate an unknown probability distribution from a training set consisting of a large amount of samples. The minimum size of the training set necessary for a good estimation depends on both the properties of the data and of the machine. We investigate this problem in a controlled environment where a Teacher Restricted Boltzmann machine (T-RBM) is used to generate the dataset and another student machine (S-RBM) is trained with it. We consider different classes of unit priors and weight regularizers and we analyze both the informed and mismatched cases, viewed as the amount of information the student receives about the teacher model. We describe the results in terms of phase transitions in the posterior distribution, interpreted as a statistical mechanics system.

In the analysis we give special attention to the Hopfield model scenario, where the problem is expressed in terms of phase diagrams, describing the zoology of the possible working regimes of the entire environment. In this present case it is possible to observe the differences between memorization and learning approach. In particular, when data become large and confused the learning methodology overcomes memorization.

Multi-outcome feature selection via anomaly detection autoencoders for radiogenomic in breast cancer patients

Alessia Mapelli
Politecnico di Milano and Human Technopole

Thursday 18, 15:00, AS2

Abstract: Severe long-term side effects occur in approximately 5In this work we propose a Deep Learning-based multivariate FS method for high-dimensional data in the presence of high-order interaction and noise.

The proposed method expands on the Deep Sparse AutoEncoder Ensemble (DSAEE) proposed in [1] by introducing multi-endpoint FS and controlling imputation noise with a tailored denoising training procedure. In the original DSAEE, an ensemble of autoencoders is trained to reconstruct the control group, and tested on a mixed population of controls and cases. The most discriminant features are selected comparing the distribution of reconstruction errors between groups. Multivariate FS is accomplished by redefining the control group (i.e. patients without any LT), and training an ensemble for each endpoint, to eventually merge the sets of selected features. We confirm via simulations that our methodology outperforms the univariate one in identifying highly informative features for class separation. We then test it on the RADPrecise Breast Cancer Cohort, and incorporate the selected variants into clinical risk models through the interaction-aware method for polygenic risk scoring (PRSi) [2]. PRSi integration determined a significant increase in models' performance, highlighting the predictive and descriptive qualities of the variants selected by our multivariate DSAEE.

This work introduces a methodology for FS in a multi-outcome binary framework that effectively addresses the challenges of genomic studies. The model's clinical applicability is a crucial aspect of its significance.

References:

[1] M. C. Massi, F. Gasperoni, F. Ieva et al.: "Feature selection for imbalanced data with deep sparse autoencoders ensemble", Statistical Analysis and Data Mining: The ASA Data Science Journal, 15(3), June 2022, 376-395.

[2] N. R. Franco, M.C. Massi, F. Ieva et al.: "Development of a method for generating SNP interaction-aware polygenic risk scores for radiotherapy toxicity", Radiotherapy and Oncology, 159, June 2021, 241-248

Efficient non-parametric learning with random projections

Giacomo Meanti
Istituto Italiano di Tecnologia

Friday 19, 12:00, AS1

Abstract: A key bottleneck of modern machine learning is its computational footprint, which hampers the environmental sustainability of many techniques. In this talk, we will present results showing that many nonparametric algorithms can be compressed while provably preserving their performance guarantees. We will show how theoretical guarantees inform efficient software design. Further, we will discuss several applications where the proposed methods lead to state-of-the-art accuracy with orders of magnitude less compute.

Implicit Regularization

Cesare Molinari
UniGE

Wednesday 17, 15:00, AS1

Abstract: The goal of machine learning is to achieve a good prediction exploiting training data and some property of the model, that may be known or not a-priori. The most common methods to reach the last objective are explicit and implicit regularization. The first technique is applicable only when the desired property is known a-priori. Then an appropriate regularizer is explicitly introduced to find, among all the solutions, a good generalizing one. The second technique, i.e. implicit regularization, is the main subject of the talk. It is based on the inductive bias intrinsically induced by the method chosen to fit the data, both in the parametrization and in the optimisation process. Recently, the success of learning is related to re- and over-parameterization (that are widely used - for instance - in neural networks applications) and to simple minimization algorithms. However, there is still an open question of how to find systematically what is the inductive bias hidden behind.

The goal of this talk is taking a step in this direction, studying extensively many reparameterizations used in the state of the art and providing a common structure to analyze the problem in a unified way. We show that gradient descent on the empirical loss for many reparameterizations is equivalent, in the original problem, to a generalization of mirror descent. The mirror function depends on the reparameterization and introduces an inductive bias, which plays the role of the regularizer. Our theoretical results provide asymptotic behavior and convergence in the simplified setting of linear models.

Physics-informed neural networks for computational medicine in cardiovascular diseases

Stefano Pagani
MOX, Department of Mathematics, Politecnico di Milano

Thursday 18, 17:30, AS2

Abstract: In many cardiovascular applications, patient data provides clinicians with only partial insights into disease characteristics and mechanisms, limiting the opportunities for developing personalized therapeutic approaches. Physics-informed machine learning can help overcome this limitation by leveraging physical knowledge encoded in parametric differential models, enabling efficient and accurate solutions to inverse or parameter/field estimation problems. Specifically, physics-informed neural networks (PINNs) efficiently approximate input-output maps that reconstruct spatio-temporal quantities of interest, such as blood velocity and pressure, mechanical displacement or electric potential fields. NN's weights and biases, along with unknown parameters or fields of the differential models, are optimized by minimizing a loss function consisting of a data fidelity term, the residual of the differential equations describing the underlying phenomenon, and additional regularization terms. In biomedical applications, however, the optimal design of this loss function and the NN architecture is very challenging.

In this talk, we present our numerical strategies to address these challenges in two cardiovascular applications. Specifically, we perform numerical experiments involving the reconstruction of aortic velocity, pressure and wall shear stress fields from 4D flow MRI, and the estimation of heterogeneous passive mechanical properties coupled with the approximation of displacement fields from a limited amount of noisy data. We first analyze PINN accuracy and robustness with respect to noise on synthetic datasets generated by numerical simulations, searching for the best configuration that balances efficiency and accuracy. We then show numerical results on clinical data to evaluate the potential of the method in providing personalized models for the diagnosis of cardiovascular diseases.

Optimisation analysis with deep linear neural networks using the Bures-Wasserstein Loss

Katerina Papagiannouli
University of Pisa

Thursday 18, 14:00, AS1

Abstract: We consider a deep matrix factorization model of covariance matrices trained with the Bures-Wasserstein distance. While recent works have made advances in the study of the optimization problem for overparametrized low-rank matrix approximation, much emphasis has been placed on discriminative settings and the square loss. In contrast, our model considers another type of loss and connects with the generative setting. We characterize the critical points and minimizers of the Bures-Wasserstein distance over the space of rank-bounded matrices. The Hessian of this loss at low-rank matrices can theoretically blow up, which creates challenges to analyze convergence of gradient optimization methods. We establish convergence results for gradient flow using a smooth perturbative version of the loss as well as convergence results for finite step size gradient descent under certain assumptions on the initial weights.

Joint work with P. Bréchet, J. An, and G. Montufar, International Conference in Machine Learning 2023.

A topological model for partial equivariance in deep learning and data analysis

Nicola Quercioli
Università di Bologna

Friday 19, 10:00, AS2

Abstract: The success of geometric deep learning lies in its effective reduction of dimensionality, achieved by incorporating an inductive bias based on pre-existing knowledge of data symmetries. Concurrently, as geometric deep learning techniques advance, the theory of group equivariant non-expansive operators (GENEOs) has emerged. Recognizing that dataset symmetries may not always be accurately represented as a group, there is a compelling motivation to extend GENEOs to encompass partial equivariance. In this presentation, we introduce a topological framework designed to explore measurement spaces, whose domains are subject to the influence of specific self-maps, along with the space of partially equivariant non-expansive operators (P-GENEOs) connecting these measurement spaces. We define pseudo-metric structures on these spaces and elucidate their key properties. Notably, the space of P-GENEOs retains many advantageous characteristics of the GENEO space, including compactness and convexity.

Scientific Machine Learning augmentation of physics-based models in Computational Cardiology

Francesco Regazzoni
MOX, Dep. of Mathematics, Politecnico di Milano

Wednesday 17, 16:30, AS2

Abstract: The development of computational models in the cardiovascular field is a challenging research area, where the need for accurate responses in short timeframes conflicts with the complexity of the underlying physical processes and the great anatomical and functional variability among patients. In this context, physics-based models require long times and computational resources for the numerical discretization of multi-scale and multi-physics systems of differential equations, while data-driven methods rarely achieve high accuracy and generalization capabilities. In this talk, we present scientific machine learning methods that integrate physical knowledge with data-driven techniques to accelerate the evaluation of differential models and address many-query problems - such as sensitivity analysis, robust parameter estimation, and uncertainty quantification - in cardiovascular applications. To speed up input-output evaluations, we develop emulators of time-dependent processes capable of predicting spatial outputs and accounting for geometric variability from patient to patient. Our methods also enable data-driven learning of mathematical models for the slow-scale remodeling associated with processes whose fast scale is well characterized by physics-based models. Numerical results demonstrate that these scientific machine learning methods enhance efficiency and accuracy in approximating quantities of interest, as well as in solving parameter estimation and uncertainty quantification problems.

Mathematical modelling of brain tumour growth: model order reduction and patient-specific parameter estimation

Davide Riccobelli

MOX - Dipartimento di Matematica, Politecnico di Milano

Thursday 18, 15:30, AS2

Abstract: The physical parameters driving the growth and the recurrence of Glioblastoma Multiforme (GBM) are highly specific to each patient. For this reason,finding a mathematical model that allows an accurate parameter estimation from neuroimaging data becomes of fundamental importance if we want to propose a computational framework that could help clinicians with decision-making. In general, parameter estimation requires a high computational cost, often unsuitable for clinical use if not assisted with techniques to reduce the complexity of starting problem. In this talk, we propose a diffuse interface model of GBM growth based on mixture theory, which consists of a Cahn-Hilliard equation coupled with a reaction-diffusion equation to describe the evolution of the nutrient for cancer cells. The specificity of each patient is modelled via a set of numerical parameters, which dictate the peculiar growth of the tumour and whose prediction is the final objective of the proposed methodology. Starting from a full-order discretization of the proposed model based on the finite element method, we obtain a reduced-base model (ROM) through the proper orthogonal decomposition technique (POD). For the solution of the PDEs system, we use FEniCSx, a powerful computing platform, while its reduction is computed thanks to the Python library RBniCSx. By means of a neural network-based approach, we build a map between the parameter space and the solution in the reduced space that describes the concentration of the tumour over time. As a tool to estimate patient-specific parameters, we propose an approach based on a second neural network trained to predict the parameters of the model based on the tumour distribution in two separate time instants. In this way, the computational effort focused on the training phase, that requires the FOM to be solved several times starting from the same initial condition, is balanced by the rapidity in the estimation of the parameters once the information on the actual evolution of the tumour at a second time instant is available. Such an approach may be exploited in clinical practice to deduce the parameters of the model from imaging data

Adversarial training as minimax optimal control problems

Alessandro Scagliotti
Munich Center for Machine Learning (MCML) and TU Munich

Wednesday 17, 15:30, AS1

Abstract: In this talk, we address the adversarial training of neural ODEs from a robust control perspective. This is an alternative to the classical training via empirical risk minimization, and it is widely used to enforce reliable outcomes for input perturbations. Neural ODEs allow the interpretation of deep neural networks as discretizations of control systems, unlocking powerful tools from control theory for the development and the understanding of machine learning. In this specific case, we formulate the adversarial training with perturbed data as a minimax optimal control problem, for which we derive first order optimality conditions in the form of Pontryagin's Maximum Principle. We provide a novel interpretation of robust training leading to an alternative weighted technique, which we test on a low-dimensional classification task.

Bi-level optimization for hyperparameters tuning in sparse low-rank learning algorithms

Laura Selicato

Consiglio Nazionale delle Ricerche - Istituto di Ricerca sulle Acque

Friday 19, 12:00, AS2

Abstract: Over the past decade, machine learning has emerged as one of the main innovation drivers. Its research community is expanding at an unprecedented speed, thanks to the growing need of building accurate, reliable, and interpretable models, that respond to the multitude of data generated, through increasingly sophisticated algorithms. All these algorithms require the configuration of hyperparameters (HPs), i.e. parameters that govern the learning approach. HPs tuning is a crucial problem in the field of the learning process. Indeed, the selection of the HPs configuration has an important impact on the final performance. The main goal of the hyperparameter optimization (HPO) problem is to automate the search process, thereby improving the generalization performance of models and enabling a more flexible design of the underlying learning algorithms. The common strategies to tackle this problem, grid and random search, are time-consuming since they are driven by some performance metrics, commonly measured by cross-validation. A reliable approach, when the objective function is known, is to transform the HPs optimization into a bi-level optimization problem that can be solved by gradient descent techniques. The challenge for this approach is the estimation of the gradient with respect to the HPs. In this work, we present a new mathematical framework for solving the optimization of penalty HPs in Nonnegative Matrix Factorization (NMF) based on bi-level techniques. Given the natural nonnegativity of real data, we focus on decompositions of nonnegative matrices, and on penalty HPs, which turn out to be useful to emphasize intrinsic properties in the data, such as sparsity.

Dense Associative Memory in the Teacher-Student Setting

Robin Thériault
Scuola Normale Superiore di Pisa

Wednesday 17, 18:00, AS1

Abstract: Modern Hopfield networks are known for their feature to prototype transition and adversarial robustness. However, previous theoretical studies have been mostly concerned with their storage capacity. We bridge this gap by calculating the phase diagram of p-body Hopfield networks in the teacher-student setting, uncovering ferromagnetic phases reminiscent of the prototype and feature phases of modern Hopfield networks. We show that using a larger p for the student than the teacher gives the student an extensive tolerance to noise. We then derive a closed-form expression measuring the adversarial robustness of such a student at zero temperature.

Real-time optimal control of parametrized systems by deep learning-based reduced order models

Matteo Tomasetto
Politecnico di Milano

Wednesday 17, 17:30, AS2

Abstract: Many optimal control problems require suitable strategies in order to steer instantaneously the considered dynamics. Moreover, the control action needs to be updated whenever the underlying scenario undergoes variations, as often happens in applications. Full-order models based on, e.g., Finite Element Method, are not suitable for these settings due to the computational burden. In addition, conventional reduced order modeling techniques, such as the Reduced Basis method, are linear, intrusive and not efficient in addressing nonlinear time-dependent dynamics. We thus propose a nonlinear, non-intrusive technique based on Deep Learning-based Reduced Order Models (DL-ROMs) to control rapidly parametrized PDEs under different scenarios. Both low-dimensional and distributed control actions can be considered. In the former case, the physical constraint is replaced by a DL-ROM to speedup optimization. In the latter, optimal full-order snapshots are generated and reduced by deep autoencoders, while neural networks are exploited to learn the parameter-to-solution map. After data generation, dimensionality reduction and neural networks training in an offline phase, optimal control strategies can be rapidly retrieved online for all the scenarios of interest. The speedup and the high accuracy of the proposed approach have been assessed on different PDE-constrained optimization problems, ranging from diffusion-advection PDEs to unsteady Navier-Stokes equations.

Latent space vector arithmetic for representation learning interpretability in medical imaging

Andrea Mario Vergani
Politecnico di Milano, Human Technopole

Wednesday 17, 15:30, AS2

Abstract: Deep representation learning allows the extraction of latent factors from medical imaging data. In this domain, given a fixed dimension for semantic latent vector, diffusion autoencoder models provide sensibly higher reconstruction quality than variational autoencoder (VAE)-based methods, thanks to the introduction of the stochastic latent, having the same dimension as input images; however, contrary to some VAE variants (e.g., factor VAE), the diffusion representation is not disentangled, causing latent factors to be less straightforwardly interpretable through direct correlations with the traits of interest. In turn, we know that poor interpretability limits clinical applications. This work bridges the gap between diffusion latent factor representation and clinically meaningful phenotypic traits through latent space vector arithmetic. Specifically, we extracted individuals' cardiac semantic latent vectors of size 128 per individual (input long axis heart magnetic resonance acquisitions of size 128x128 x 50 time-steps), obtaining a test structural similarity index measure (SSIM) of 0.95 for the diffusion autoencoder (VAE variants showed SSIM around 0.80). To enhance interpretability, we manipulated the latent factors proportionally to their weights obtained in regressing known cardiovascular phenotypes; moreover, we validated our findings by latent space vector interpolation (convex combination and spherical linear interpolation for semantic and stochastic codes, respectively). In this way, we were able to more clearly associate the latent representation by diffusion model with medical phenotypes such as left ventricular and aorta measures. To the best of our knowledge, this is the first work about diffusion autoencoders and their latent space interpretability with real-world medical imaging data.

Fast Approximation and Optimal Contraction for Vecchia Gaussian Processes

Yichen Zhu
Department of Decision Science

Friday 19, 13:00, AS1

Abstract: Gaussian processes are widely applied in spatial statistics due to its flexibility and automatic uncertainty quantification, but its computation suffers from the $O(n^3)$ computational complexity. Vecchia approximations provide a scalable computational solution to Gaussian processes, but the absence of theoretical foundation leads to unclear guidance in specifying the underlying graphical model. We provide theoretical support for Vecchia approximations of Gaussian processes from two aspects: 1. Considering them as approximations of the original Gaussian processes, we bound rate of approximation in Wasserstein distance; 2. Directly analyzing them as Gaussian process methods, we prove the posterior rate of contraction. Under Matern covariance function, such rate matches the minimax rate of regression when the truth belongs to some Hölder space.