

**Department of Statistics Archive of Data Science Seminars - Michaelmas**  
**Term 2020**

**Monday 18 October 2021, 2-3pm - Rada Mihalcea (University of Michigan)**

**Biography:** Rada Mihalcea is the Janice M. Jenkins Collegiate Professor of Computer Science and Engineering at the University of Michigan and the Director of the Michigan Artificial Intelligence Lab. Her research interests are in computational linguistics, with a focus on lexical semantics, multilingual natural language processing, and computational social sciences. She serves or has served on the editorial boards of the Journals of Computational Linguistics, Language Resources and Evaluations, Natural Language Engineering, Journal of Artificial Intelligence Research, IEEE Transactions on Affective Computing, and Transactions of the Association for Computational Linguistics. She was a program co-chair for EMNLP 2009 and ACL 2011, and a general chair for NAACL 2015 and \*SEM 2019. She currently serves as ACL President. She is the recipient of a Presidential Early Career Award for Scientists and Engineers awarded by President Obama (2009), an ACM Fellow (2019) and a AAAI Fellow (2021). In 2013, she was made an honorary citizen of her hometown of Cluj-Napoca, Romania.

**Title:** TextRank: Bringing Order Into Texts (Revisited)

**Abstract:** TextRank is a framework for the application of graph-based ranking algorithms to structures derived from text, which demonstrates how the synergy between graph-theoretical algorithms and graph-based text representations can result in efficient unsupervised methods for several natural language processing tasks. The original TextRank framework was proposed more than ten years ago, when it was found to be effective for several text processing applications, including word sense disambiguation, extractive summarization, and keyphrase extraction. Since then, TextRank has continued to be used in numerous language processing applications, oftentimes leading to performance comparable to state-of-the-art algorithms, despite being a light unsupervised methodology. In this talk, I will revisit TextRank and talk about some of its successful applications over the recent years.

.....

Monday 8 November 2021, 3-4pm - Suvrit Sra (MIT)

[Website](#)

**Biography:** Suvrit Sra is an Associate Professor in the EECS Department at MIT, and also a core faculty member of the Laboratory for Information and Decision Systems (LIDS), the Institute for Data, Systems, and Society (IDSS), as well as a member of MIT-ML and Statistics groups. He obtained his PhD in Computer Science from the University of Texas at Austin. Before moving to MIT, he was a Senior Research Scientist at the Max Planck Institute for Intelligent Systems, Tübingen, Germany. He has held visiting faculty positions at UC Berkeley (EECS) and Carnegie Mellon University (Machine Learning Department) during 2013-2014. His research bridges a number of mathematical areas such as differential geometry, matrix analysis, convex analysis, probability theory, and optimization with machine learning. He founded the OPT (Optimization for Machine Learning) series of workshops, held from OPT2008–2017 at the NeurIPS (erstwhile NIPS) conference. He has co-edited a book with the same name (MIT Press, 2011). He is also a co-founder and chief scientist of macro-eyes, a global healthcare+AI-for-good startup.

**Title:** Do we understand how to find critical points in nonsmooth optimization?

**Abstract:** Machine learning is full of nonconvex nonsmooth optimization problems, yet almost always the “nonsmoothness” is swept under the rug. In this talk, I will not ignore this key property, and will discuss the computational complexity of finding critical points of a rich class of nonsmooth, nonconvex functions. In particular, the class chosen contains widely used ReLU neural networks as a special case. I will focus on two key ideas: first, that it is impossible to find an  $\epsilon$ -stationary point using first-order methods in finite time; and second, a natural alternative notion of  $(\delta, \epsilon)$ -stationarity. I will describe a formal algorithm (implementable) and its complexity for finding this modified notion of stationarity. Time permitting, I will highlight some open directions and other recent progress too.

.....

Monday 15 November 2021, 3.30-4.30pm - Po-Ling Loh (University of Cambridge)

[Website](#)

**Biography:** I am a Lecturer in the Statistical Laboratory in the Department of Pure Mathematics and Mathematical Statistics at the University of Cambridge. My current research interests include high-dimensional statistics, optimization, network inference, robust statistics, and differential privacy. I am also interested in statistical applications to medical imaging and epidemiology.

**Title:** A modern take on Huber regression

**Abstract:** In the first part of the talk, we discuss the use of a penalized Huber M-estimator for high-dimensional linear regression. We explain how a fairly straightforward analysis yields high-probability error bounds that hold even when the additive errors are heavy-tailed. However, the parameter governing the shape of the Huber loss must be chosen in relation to the scale of the error distribution. We discuss how to use an adaptive technique, based on Lepski's method, to overcome the difficulties traditionally faced by applying Huber M-estimation in a context where both location and scale are unknown.

In the second part of the talk, we turn to a more complicated setting where both the covariates and responses may be heavy-tailed and/or adversarially contaminated. We show how to modify the Huber regression estimator by first applying an appropriate "filtering" procedure to the data based on the covariates. We prove that in low-dimensional settings, this filtered Huber regression estimator achieves near-optimal error rates. We further show that the commonly used least trimmed squares and least absolute deviation estimators may similarly be made robust to contaminated covariates via the same covariate filtering step. This is based on joint work with Ankit Pensia (UW-Madison) and Varun Jog (Cambridge).

.....

**Monday 22 November 2021, 2-3pm - Vitaliy Kurlin (University of Liverpool)**

[Website](#)

**Biography:** Vitaliy Kurlin is a Data Scientist at the Materials Innovation Factory at Liverpool and the Royal Academy Engineering Fellow at the Cambridge Crystallographic Data Centre, who leads the group developing the new area of Periodic Geometry for applications in crystallography and materials science.

**Title:** Mathematical Data Science for solid crystalline materials

**Abstract:** Most real data is ambiguous in the sense that the same real object has too many (often infinitely many) data representations. For example, any periodic lattice can be represented by infinitely many different linear bases. Data Science aims to define equivalence relations on real objects to make their classifications meaningful. The most natural equivalence of solid crystalline materials is rigid motion or isometry, because periodic crystal structures are determined in a rigid form.

This talk will describe recent advances in isometry classifications of finite and periodic point sets. The implemented isometry invariants completely distinguished all (hundreds of thousands) real periodic crystals in the world's largest Cambridge Structural Database. The work is joint with colleagues from the Data Science Theory and Applications group, <http://kurlin.org/index.php#group>

.....

**Monday 6 December 2021, 1-2pm - Negar Kiyavash (Ecole polytechnique fédérale de Lausanne)**

[Website](#)

**Biography:** Negar Kiyavash is the chair of Business Analytics (BAN) at École polytechnique fédérale de Lausanne (EPFL) at the College of Management of Technology. Prior to joining EPFL, she was a faculty member at the University of Illinois, Urbana-Champaign, and at Georgia Institute of Technology. Her research interests are broadly in the area of statistical learning and applied probability with special focus on network inference and causality. She is a recipient of the NSF CAREER and AFOSR YIP awards.

**Title:** Database alignment: fundamental limits and efficient algorithms

**Abstract:** As data collection becomes ubiquitous, understanding the potential benefits as well as the risks posed by the availability of such large amount of data becomes more pressing. Identifying how data from different sources relate to each other, could allow to merge and augment data. On the positive side, this could help for instance in deducting functionality of proteins by comparing protein interaction networks of different species. On the negative side, such alignment could cause unintended exposure of confidential information. A famous case of such breach occurred when customer data from the anonymous Netflix Prize database was revealed through alignment with public IMDB profiles.

In this talk we present information-theoretic results on database alignment, showing how the size of databases and the correlation between their elements determines the success of alignment. Database alignment is closely related to equally interesting problem of network alignment, a generalization of the graph isomorphism problem.

### **Archive of Data Science Seminars - Lent Term 2021**

Monday 31 January 2022, 2 - 3pm - Cynthia Rudin (Duke University)

[Website](#)

**Biography:** Cynthia Rudin is a professor of computer science, electrical and computer engineering, statistical science, mathematics, and biostatistics & bioinformatics at Duke University. She directs the Interpretable Machine Learning Lab, whose goal is to design predictive models with reasoning processes that are understandable to humans. Her lab applies machine learning in many areas, such as healthcare, criminal justice, and energy reliability. She holds an undergraduate degree from the University at Buffalo, and a PhD from Princeton University. She is the recipient of the 2022 Squirrel AI Award for Artificial Intelligence for the Benefit of Humanity from the Association for the Advancement of Artificial Intelligence (the “Nobel Prize of AI”). She is a fellow of the American Statistical Association and a fellow of the Institute of Mathematical Statistics. Her work has been featured in

many news outlets including the NY Times, Washington Post, Wall Street Journal, and Boston Globe.

**Title:** Scoring Systems: At the Extreme of Interpretable Machine Learning.

**Abstract:** With widespread use of machine learning, there have been serious societal consequences from using black box models for high-stakes decisions, including flawed bail and parole decisions in criminal justice, flawed models in healthcare, and black box loan decisions in finance. Interpretability of machine learning models is critical in high stakes decisions.

In this talk, I will focus on one of the most fundamental and important problems in the field of interpretable machine learning: optimal scoring systems. Scoring systems are sparse linear models with integer coefficients. Such models first started to be used ~100 years ago. Generally, such models are created without data, or are constructed by manual feature selection and rounding logistic regression coefficients, but these manual techniques sacrifice performance; humans are not naturally adept at high-dimensional optimization. I will present the first practical algorithm for building optimal scoring systems from data. This method has been used for several important applications to healthcare and criminal justice.

I will mainly discuss work from three papers:

Learning Optimized Risk Scores. Journal of Machine Learning Research, 2019.

<http://jmlr.org/papers/v20/18-615.html>

The Age of Secrecy and Unfairness in Recidivism Prediction. Harvard Data Science Review, 2020.

<https://hdsr.mitpress.mit.edu/pub/7z10o269>

Association of an Electroencephalography-Based Risk Score With Seizure Probability in Hospitalized Patients. JAMA Neurology, 2017.

[https://jamanetwork.com/journals/jamaneurology/fullarticle/2656833?utm\\_source=jps&utm\\_medium=email&utm\\_campaign=author\\_alert-jamanetwork&utm\\_content=author-author\\_engagement&utm\\_term=1m](https://jamanetwork.com/journals/jamaneurology/fullarticle/2656833?utm_source=jps&utm_medium=email&utm_campaign=author_alert-jamanetwork&utm_content=author-author_engagement&utm_term=1m)

.....  
**Monday 7 February 2022, 2 - 3pm - Joseph Salmon (Université de Montpellier)**

[Website](#)

**Biography:** Joseph Salmon is a full professor at Université de Montpellier. Prior to joining Université de Montpellier in 2018, he was a visiting assistant professor at University of Washington (2018) and an assistant professor at Télécom ParisTech (2012 - 2018).

He is specialised in high dimensional statistics, optimization and statistical machine learning. His research interests include convex optimization, sparse regression models and inverse problems from imaging science. Recent contributions include speeding-up standard Lasso solvers and leveraging the noise structure to improve signal estimation. As an associate member with the INRIA Parietal Team, he is also contributing to apply his work to brain imaging challenges.

**Title:** Implicit differentiation for fast hyperparameter selection in non-smooth convex learning.

**Abstract:** Finding the optimal hyperparameters of a model can be cast as a bilevel optimization problem, typically solved using zero-order techniques. We study here first-order methods when the inner optimization problem is convex but non-smooth. We show that the forward-mode differentiation of proximal gradient descent and proximal coordinate descent yield sequences of Jacobians converging toward the exact Jacobian. Using implicit differentiation, we show it is possible to leverage the non-smoothness of the inner problem to speed up the computation. Finally, we provide a bound on the error made on the hypergradient when the inner optimization problem is solved approximately. Results on regression and classification problems reveal computational benefits for hyperparameter optimization, especially when multiple hyperparameters are required.

Joint work with Q. Bertrand, Q. Klopfenstein, M. Massias, M. Blondel, S. Vaiter and A. Gramfort

<https://arxiv.org/abs/2105.01637>

<https://arxiv.org/abs/2002.08943>

.....  
**Monday 14 February 2022, 2 - 3pm - Volkan Cevher (Ecole polytechnique fédérale de Lausanne)**

[Website](#)

**Biography:** Volkan Cevher received the B.Sc. (valedictorian) in electrical engineering from Bilkent University in Ankara, Turkey, in 1999 and the Ph.D. in electrical and computer engineering from the Georgia Institute of Technology in Atlanta, GA in 2005. He was a Research Scientist with the University of Maryland, College Park from 2006-2007 and also with Rice University in Houston, TX, from 2008-2009. Currently, he is an Associate Professor at the Swiss Federal Institute of Technology Lausanne and a Faculty Fellow in the Electrical and Computer Engineering Department at Rice University. His research interests include machine learning, signal processing theory, optimization theory and methods, and information theory. Dr. Cevher is an ELLIS fellow and was the recipient of the Google Faculty Research award in 2018, the IEEE Signal Processing Society Best Paper Award in 2016, a Best Paper Award at CAMSAP in 2015, a Best Paper Award at SPARS in 2009, and an ERC CG in 2016 as well as an ERC StG in 2011.

**Title:** Optimization challenges in adversarial machine learning

**Abstract:** Thanks to neural networks (NNs), faster computation, and massive datasets, machine learning (ML) is under increasing pressure to provide automated solutions to even harder real-world tasks beyond human performance with ever faster response times due to potentially huge technological and societal benefits. Unsurprisingly, the NN learning formulations present a fundamental challenge to the back-end learning algorithms despite their scalability, in particular due to the existence of traps in the non-convex optimization landscape, such as saddle points, that can prevent algorithms from obtaining “good” solutions.

In this talk, we describe our recent research that has demonstrated that the non-convex optimization dogma is false by showing that scalable stochastic optimization algorithms can avoid traps and rapidly obtain locally optimal solutions. Coupled with the progress in representation learning, such as over-parameterized neural networks, such local solutions can be globally optimal.



Unfortunately, this talk will also demonstrate that the central min-max optimization problems in ML, such as generative adversarial networks (GANs), robust reinforcement learning (RL), and distributionally robust ML, contain spurious attractors that do not include any stationary points of the original learning formulation. Indeed, we will describe how algorithms are subject to a grander challenge, including unavoidable convergence failures, which could explain the stagnation in their progress despite the impressive earlier demonstrations. We will conclude with promising new preliminary results from our recent progress on some of these difficult challenges.

.....

**Monday 21 February 2022, 2 - 3pm - Silvia Villa (University of Genoa)**

[Website](#)

**Biography:** Silvia Villa is an associate professor at the University of Genoa, where she works in the Machine Learning Genoa Center. Her research is focused on optimization, and in particular on algorithms for solving machine learning and inverse problems. She is the coordinator of an MSCA ITN European Project on optimization for data science involving 15 PhD students.

**Title:** Iterative regularization for low complexity regularizers

**Abstract:** Iterative regularization exploits the implicit bias of an optimization algorithm to regularize ill-posed problems. Constructing algorithms with such built-in regularization mechanisms is a challenge of modern inverse problems and machine learning, providing both a new perspective on algorithms analysis and significant speed-ups compared to explicit regularization. I will present different iterative regularization methods, depending on the desired regularization and analyze their convergence and stability properties.

.....

**Monday 7 March 2022, 2 - 3pm - Dino Sejdinovic (University of Oxford)**

[Website](#)

**Biography:** I am an Associate Professor in Statistics at the University of Oxford and a Fellow of Mansfield College. I conduct research at the interface between machine learning and statistical methodology.

**Title:** Recent Developments at the Interface Between Kernel Embeddings and Gaussian Processes

**Abstract:** Reproducing kernel Hilbert spaces (RKHS) provide a powerful framework, termed kernel mean embeddings, for representing probability distributions, enabling nonparametric statistical inference in a variety of applications. I will give an overview of this framework and present some of its recent developments which combine RKHS formalism with Gaussian process modelling. Some recent applications include causal data fusion, where data of different quality needs to be combined in order to estimate the average treatment effect, as well as statistical downscaling using potentially unmatched multi-resolution data.

### **Archive of Data Science Seminars - Summer Term 2022**

**Monday 16 May 2022, 4 - 5pm - Krishnakumar Balasubramanian (University of California)**

[Website](#)

**Biography:** Krishna Balasubramanian is an Assistant Professor in the Department of Statistics, University of California, Davis. His recent research interests include stochastic optimization and sampling, reproducing kernel Hilbert space methods, and geometric and topological statistics. His research was/is supported by a Facebook PhD fellowship, and CeDAR and NSF grants.

**Title:** Unified RKHS Methodology and Analysis for Functional Linear and Single-Index Models.

**Abstract:** Functional linear and single-index models are elementary methods in the functional data analysis toolkit and are widely used methods for performing

regression when the covariates are observed as random functions in various applications. In the existing literature, however, constructing the associated estimators and studying their theoretical properties are invariably carried out on a case-by-case basis for the individual model under consideration. In this work, we provide a unified methodological and theoretical framework for estimating the index in the functional linear and single-index models; in the latter case the proposed approach is agnostic to the specification of the link function. On the methodological side, we show that the reproducing kernel Hilbert space (RKHS) based functional linear least-squares estimator, when viewed through the lens of infinite-dimensional Gaussian Stein's identity, also provides an estimator of the index of the single-index model. On the theoretical side, we characterize the convergence rates of the estimator for both the linear and single-index model. Our analysis has several advantages: (i) we do not require restrictive commutativity assumptions on the covariance operator of the random covariates and the integral operator associated with the reproducing kernel, and (ii) we also allow for the true index parameter to lie outside of the chosen RKHS thereby allowing for and quantifying the degree of index misspecification in the models. We recover several existing results as special case of our analysis.

.....  
**Monday 23 May 2022, 2 - 3pm - Rémi Flamary (École Polytechnique)**

[Website](#)

**Biography:** Rémi Flamary is Monge Assistant Professor at École Polytechnique in the Centre de Mathématiques Appliquées (CMAP) and holder of a Chair in Artificial Intelligence from 3IA Côte d'Azur. He was previously Associate Professor at Université Cote d'Azur (UCA) and a member of Lagrange Laboratory, Observatoire de la Cote d'Azur. He received the Dipl.-Ing. in electrical engineering and the M.S. degree in image processing from the Institut National de Sciences Appliquées de Lyon in 2008, and a Ph.D. degree from the University of Rouen in 2011. His current research interests include signal and image processing, and machine learning with a recent focus on application of Optimal Transport theory to machine learning problems.

**Title:** Modeling graphs with optimal transport.

**Abstract:** Optimal Transport (OT) has recently emerged as a powerful and interpretable tool to model and measure similarity between graph objects. In this talk we will introduce the Gromov-Wasserstein divergence and several extensions that have been proposed recently to measure a similarity between weighted graphs. We will discuss two important aspects of OT on graphs: as a divergence between non-registered graphs with potentially different number of nodes and as a transport finding optimal correspondences between those graph nodes. We will then present several applications of those divergences for dictionary learning of graphs, community clustering and graph completion.

.....

**Monday 30 May 2022, 2 - 3pm - Lester Mackey (Microsoft Research New England)**

[Website](#)

**Biography:** Lester Mackey is a Principal Researcher at Microsoft Research, where he develops machine learning methods, models, and theory for large-scale learning tasks driven by applications from climate forecasting, healthcare, and the social good. Lester moved to Microsoft from Stanford University, where he was an assistant professor of Statistics and (by courtesy) of Computer Science. He earned his PhD in Computer Science and MA in Statistics from UC Berkeley and his BSE in Computer Science from Princeton University. He co-organized the second place team in the Netflix Prize competition for collaborative filtering, won the Prize4Life ALS disease progression prediction challenge, won prizes for temperature and precipitation forecasting in the yearlong real-time Subseasonal Climate Forecast Rodeo, received best paper and best student paper awards from the ACM Conference on Programming Language Design and Implementation and the International Conference on Machine Learning, and was elected to the COPSS Leadership Academy.

**Title:** Kernel Thinning and Stein Thinning

**Abstract:** This talk will introduce two new tools for summarizing a probability distribution more effectively than independent sampling or standard Markov chain Monte Carlo thinning:

1. Given an initial  $n$  point summary (for example, from independent sampling or a Markov chain), kernel thinning finds a subset of only square-root  $n$  points with comparable worst-case integration error across a reproducing kernel Hilbert space.
2. If the initial summary suffers from biases due to off-target sampling, tempering, or burn-in, Stein thinning simultaneously compresses the summary and improves the accuracy by correcting for these biases.

These tools are especially well-suited for tasks that incur substantial downstream computation costs per summary point like organ and tissue modeling in which each simulation consumes 1000s of CPU hours.

.....

**Monday 13 June 2022, 2- 3pm - Florence d'Alché-Buc (Télécom Paris)**

[Website](#)

**Biography:** Florence d'Alché-Buc is professor and head of the Image, Data and Signal department of Télécom Paris, a founding member of Institut Polytechnique de Paris (France). She was program co-chair of NeurIPS 2019 and is currently acting as action editor of Journal of Machine Learning Research and IEEE Transactions on Pattern Analysis and Machine Intelligence. She also holds the Data Science & Artificial Intelligence for Digitalized Industry & Services Chair. Her research spans many various directions in Machine Learning with non-parametric models including structured prediction, functional output regression, dynamical systems modeling and more recently, robustness and explainability.

**Title:** Learning to predict complex outputs: a kernel view.

**Abstract:** Motivated by real-world complex output prediction tasks such as link prediction, molecule identification or functional output regression, we propose to leverage the notion of output kernel to take into account the nature of output variables whether they be discrete structures or functions. This approach boils down to encode output data as vectors of the Reproducing kernel Hilbert Space associated to the so-called output kernel. We present a vector-valued kernel machines as well as a deep variant to implement it and discuss different learning problems linked with the chosen loss function. Eventually large scale approaches can be developed using

low rank approximations of the outputs. We illustrate the framework on graph prediction and infinite task learning.