LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

INSTITUT FÜR STATISTIK

Can we open the Black Box of Deep Neural Networks? - An Information Theoretic Approach to Validate Deep Learning-Based Algorithms

Gitta Kutyniok

(Department of Mathematics, LMU Munich)

19.05.2021, 15.00 (s.t.)

Online via Zoom
(Meeting-ID: 913-2473-4411; Password: StatsCol21)

We currently witness the impressive success of deep learning in real-world applications, ranging from science to public life. At the same time, such methods still lack a profound theoretical understanding, sometimes even being referred to as ”alchemy”. This poses an enormous challenge to, in particular, sensitive applications. The area of interpretability aims to tackle this problem by identifying those features from the input, which are most relevant for the observed output. In this talk, we provide a theoretical framework for interpreting neural network decisions by formalizing the problem in a rate-distortion framework. The solver of the associated optimization, which we coin Rate-Distortion Explanation (RDE), is then accessible to a mathematical analysis. We will discuss theoretical results as well as present numerical experiments showing that our algorithmic approach outperforms established methods, in particular, for sparse explanations of neural network decisions and is flexible enough to also be applied to challenging modalities.

Biography:
Gitta Kutyniok currently has a Bavarian AI Chair for Mathematical Foundations of Artificial Intelligence at the Ludwig-Maximilians-Universität München. She received her Diploma in Mathematics and Computer Science as well as her Ph.D. degree from the Universität Paderborn in Germany, and her Habilitation in Mathematics in 2006 at the Justus-Liebig Universität Gießen. From 2001 to 2008 she held visiting positions at several US institutions, including Princeton University, Stanford University, Yale University, Georgia Institute of Technology, and Washington University in St. Louis, and was a Nachdiplomslecturer at ETH Zurich in 2014. In 2008, she became a full professor of mathematics at the Universität Osnabrück, and moved to Berlin three years later, where she held an Einstein Chair in the Institute of Mathematics at the Technische Universität Berlin and a courtesy appointment in the Department of Computer Science and Engineering until 2020. Gitta Kutyniok's research work covers, in particular, the areas of applied and computational harmonic analysis, approximation theory, artificial intelligence, compressed sensing, frame theory, imaging sciences, inverse problems, machine learning, numerical analysis of partial differential equations, and applications to life sciences and telecommunication.