



Variational Learning for Large Deep Networks

Thomas Möllenhoff
(RIKEN, Tokyo)

10/07/2024, 3.00pm

Department of Statistics, Ludwigstr. 33, Room 144
and online via Zoom (Link)
(Meeting-ID: 683 0699 4223; Password: StatsCol23)

Thomas Möllenhoff presents extensive evidence against the common belief that variational Bayesian learning is ineffective for large neural networks. First, he shows that a recent deep learning method called sharpness-aware minimization (SAM) solves an optimal convex relaxation of the variational Bayesian objective. Then, he demonstrates that a direct optimization of the variational objective with an Improved Variational Online Newton method (IVON) can consistently match or outperforms Adam for training large networks such as GPT-2 and ResNets from scratch. IVON's computational costs are nearly identical to Adam but its predictive uncertainty is better. He shows several new use cases of variational learning where he improves fine-tuning and model merging in Large Language Models, accurately predict generalization error, and faithfully estimate sensitivity to data.

Biography:

Thomas Möllenhoff received his PhD in Informatics from the Technical University of Munich in 2020. From 2020 to 2023, he was a post-doc in the Approximate Bayesian Inference Team at RIKEN. Since 2023 he works at RIKEN as a tenured research scientist. His research focuses on optimization and Bayesian deep learning and has been awarded several times, including the Best Paper Honorable Mention award at CVPR 2016 and a first-place at the NeurIPS 2021 Challenge on Approximate Inference.