



ELLIS talk

Provable Representation Learning: The Importance of Task Diversity and Pretext Tasks

Qi Lei

Princeton University

Host: Marco Mondelli

Modern machine learning models are transforming applications in various domains at the expense of a large amount of hand-labeled data. In contrast, humans and animals first establish their concepts or impressions from data observations. The learned concepts then help them to learn specific tasks with minimal external instructions. Accordingly, we argue that deep representation learning seeks a similar procedure: 1) to learn a data representation that filters out irrelevant information from the data; 2) to transfer the data representation to downstream tasks with few labeled samples and simple models. In this talk, we study two forms of representation learning: supervised pre-training from multiple tasks and self-supervised learning. Supervised pre-training uses a large labeled source dataset to learn a representation, then trains a simple (linear) classifier on top of the representation. We prove that supervised pre-training can pool the data from all source tasks to learn a good representation that transfers to downstream tasks (possibly with covariate shift) with few labeled examples. We extensively study different settings where the representation reduces the model capacity in various ways. Self-supervised learning creates auxiliary pretext tasks that do not require labeled data to learn representations. These pretext tasks are created solely using input features, such as predicting a missing image patch, recovering the color channels of an image, or predicting missing words. Surprisingly, predicting this known information helps in learning a representation useful for downstream tasks. We prove that under an approximate conditional independence assumption, self-supervised learning provably learns representations that linearly separate downstream targets. For both frameworks, representation learning provably and drastically reduces sample complexity for downstream tasks.

Wednesday, November 24, 2021 04:00pm - 06:00pm

Zoom Link:

<https://istaustria.zoom.us/j/98066215937?pwd=YmZoWDAzME13dC9LU0Jwc1kzWVphQT09>

Meeting ID: 980 6621 5937 Passcode: 177564



This invitation is valid as a ticket for the ISTA Shuttle from and to Heiligenstadt Station.
Please find a schedule of the ISTA Shuttle on our webpage:
<https://ista.ac.at/en/campus/how-to-get-here/> The ISTA Shuttle bus is marked ISTA Shuttle
(#142) and has the Institute Logo printed on the side.