



Theoretical Statistics and Mathematics Unit, Kolkata
INDIAN STATISTICAL INSTITUTE

SEMINAR

Date: March 24, 2025

Time: 04:15 PM

VENUE:

L - Infinity

(5th Floor, A.N. Kolmogorov Bhavan), ISI Kolkata

TITLE:

Few steps towards statistical theory of foundation models

SPEAKER:

Debarghya Ghoshdastidar

Technical University of Munich, Germany

ABSTRACT:

In this talk, I will consider a theoretical simplification of a foundation model pipeline: Unlabeled augmented data is used to learn a representation map $x \rightarrow f(x)$ via self-supervised pre-training. Subsequently, a downstream predictor is learned with few labeled samples to map the representation $f(x)$ to labels. The main statistical question for this pipeline is how can one characterize the downstream generalization error when the model is primarily learned through self-supervision. I first give an overview of current approaches to derive generalization error bounds, including our recent work that provides PAC-Bayes risk certificates (arXiv:2412.03486), and discuss some limitations of the existing generalization error bounds.

I will then suggest an alternative approach to characterize the generalization error of self-supervised models by exploiting the equivalence between self-supervision and kernel principal component analysis. I will present some of our results in this direction (arXiv:2309.02028, 2403.08673, 2411.11176) and mention preliminary ideas on how to get precise generalization error rates using this approach.

ALL ARE CORDIALLY INVITED