



Theoretical Statistics and Mathematics Unit

Indian Statistical Institute

203 B.T. Road, Kolkata - 700108

SEMINAR

Date: July 23, 2026

Time: 04:15 PM

VENUE:

L-Infinity, 5th Floor, A.N. Kolmogorov Bhavan, ISI Kolkata

SPEAKER:

Annesha Ghosh

Stat-Math Unit, ISI Kolkata

TITLE:

A scalable version of MADD for big-data classification

ABSTRACT:

Distance-based classifiers are very popular, and the Euclidean distance is one of the most commonly used metrics. However, classifiers based on Euclidean distance often perform poorly in high-dimensional settings due to issues such as distance concentration, violations of neighborhood structures, and the presence of hubs. In high-dimension, low-sample-size (HDLSS) situations, a data-driven semi-metric called the Mean Absolute Difference of Distances (MADD) is known to circumvent these issues. But one major problem with MADD is that its computational complexity increases quadratically with the training sample size. As a result, the application of MADD becomes computationally challenging for big datasets that have both a high dimension as well as a large number of observations. In this paper, we propose a scalable version of MADD that significantly reduces its computational complexity while retaining its advantages. This speed-up is achieved by selecting a representative set during the computation of MADD. Further speed-ups are achieved by using Random Fourier Features, particularly when the sample size is very large. We establish that our proposed methods achieve performance similar to MADD but at only a fraction of its computational time, both theoretically and numerically. Our approach broadens the scope of MADD, allowing its use for big data with a very large number of observations.

ALL ARE CORDIALLY INVITED

[Click here for More Seminar Information](#)