Bocconi

Università Commerciale
Luigi Bocconi

Data Integration: Challenges and Opportunities for Interpolation Learning under Distribution Shifts

Pragya Sur
Harvard University

Thursday, October 2nd, 2025
12:00 pm **Room 3-E4-SR03** Via Roentgen 1 Milano

Abstract

Min-norm interpolators emerge naturally arise as implicit regularized limits of modern machine learning algorithms. Recently, their out-of-distribution risk was studied when test samples are unavailable during training. However, in many applications, a limited amount of test data is typically available during training. Properties of min-norm interpolation in this setting are not well understood. In this talk, I will present a characterization of the risk of pooled min-L2-norm interpolation under covariate and concept shifts. I will show that the pooled interpolator captures both early fusion and a form of intermediate fusion. Our results have several implications. For example, under concept shift, adding data always hurts prediction when the signal-to-noise ratio is low. However, for higher signal-to-noise ratios, transfer learning helps as long as the shift-to-signal ratio lies below a threshold that I will define. Our results also show that under covariate shift, if the source sample size is small relative to the dimension, heterogeneity between domains improves the risk. Time permitting, I will introduce a novel anisotropic local law that allows to achieve some of these characterizations and is of independent interest in random matrix theory. This is based on joint work with Kenny Gu, Yanke Song and Sohom Bhattacharya.

Department of Decision
Sciences
Via Röntgen 1 – 20136
Milano

Tel. 02 5836.5632
Fax 02 5836.5630