Bocconi

6 MAY 2025
12:00 PM
Room 2-E4-SR03
Via Röntgen 1, Milan, 2° floor

Memorization and Generalization in Generative Diffusion under the Manifold Hypothesis

Speaker

Enrico Ventura
Postdoc Researcher
Bocconi University

Abstract

In this talk we examine how Diffusion Models (DMs) memorize and generalize structured data embedded in low-dimensional latent manifolds.

To analyze memorization, we extend the Random Energy Model (REM) framework proposed by [1]. We first derive an explicit expression for the memorization time at which typical trajectories fall into the basins of attraction of training points, revealing that a mitigation of the curse of dimensionality is possible for highly structured data [2]. We also show that the memorization time can depend on the position in the ambient space implying a selective loss of dimensionality where some prominent features of the data are memorized in advance [3].

Turning to generalization, we apply Random Matrix Theory to first show how, progressively in time, DMs geometrically conform to the target distribution supported on the manifold, avoiding manifold overfitting [4]. We further identify a generalization time where the Kullback-Leibler divergence between the empirical and true data distributions reaches a minimum. Our results suggest that optimal generalization emerges during the memorization phase, highlighting an important interplay between memorization and generalization in DMs [2].

Università Bocconi
DEPARTMENT OF COMPUTING SCIENCES