

Department of Applied and Computational Mathematics and Statistics Colloquium



Arian Maleki

Columbia University

Accurate and Efficient Data Point Removal for High-Dimensional Settings

Consider a model trained with p parameters from n independent and identically distributed observations. To assess a data point's impact on the model, we remove it from the dataset and aim to understand the model's behavior when trained on the remaining data. This scenario is relevant in various classical and modern applications, including risk estimation, outlier detection, machine unlearning, and data valuation. Conventional approaches involve training the model on the remaining data, but these can be computationally demanding. Consequently, researchers often resort to approximate methods. This talk highlights that in high-dimensional settings, where p is either larger than n or at the same order, many approximation methods may prove ineffective. We will present and analyze an accurate approximation method tailored for high-dimensional regimes, elucidating the conditions for its accuracy. In the concluding part of the presentation, time permitting, we will briefly discuss some of the unresolved issues in this domain.

Monday, March 30, 2026

3:30 – 4:30 PM

117 Hayes-Healy Center

Colloquium Tea – 3:15 PM in 117 Hayes-Healy