

1163-62-32

Caroline Uhler* (cuhler@mit.edu). *Multi-Domain Data Integration: From Observations to Mechanistic Insights.*

Massive data collection holds the promise of a better understanding of complex phenomena and ultimately, of better decisions. An exciting opportunity in this regard stems from the growing availability of perturbation / intervention data (manufacturing, advertisement, education, genomics, etc.). To obtain mechanistic insights from such data, a major challenge is the integration of different data modalities (video, audio, interventional, observational, etc.). Using genomics and in particular the problem of identifying drugs for the repurposing against COVID-19 as an example, I will first discuss our recent work on coupling autoencoders in the latent space to integrate and translate between data of very different modalities such as sequencing and imaging. I will then present a framework for integrating observational and interventional data for causal structure discovery and characterize the causal relationships that are identifiable from such data. We end by a theoretical analysis of overparameterized autoencoders, characterizing their implicit bias and showing that such networks trained using standard optimization methods implement associative memory. Collectively, our results have major implications for planning and learning from interventions in various application domains. (Received July 15, 2020)