

Learning without correspondence

Daniel Hsu *

djhsu@cs.columbia.edu

I will talk about algorithmic and statistical aspects of two inference problems where correspondence information is missing. In the first problem, the goal is to estimate regression coefficients from data where the correspondence between covariates and responses is unknown. There, the natural least squares optimization problem is NP-hard, and there are information-theoretic barriers to recovering the correspondence. I'll describe some efficient algorithms for worst-case and average-case versions of the problem, and also present some lower-bounds. The second problem is a generalization of the (real) phase retrieval problem called correspondence retrieval. I'll describe a sample-optimal algorithm for the noise-free setting and a sample-inefficient method-of-moments algorithm for the noisy setting.

This is based on joint works with Alex Andoni (Columbia), Kevin Shi (Columbia), and Xiaorui Sun (Microsoft Research).