Proceedings of Machine Learning Research vol 140:1–1, 2022

Non-parametric Inference Adaptive to Intrinsic Dimension

Khashayar Khosravi KHOSRAVI@STANFORD.EDU
Stanford University

Greg Lewis GLEWIS@MICROSOFT.COM
Microsoft Research

Vasilis Syrgkanis VASY@MICROSOFT.COM
Microsoft Research

Abstract

We consider non-parametric estimation and inference of conditional moment models in high dimensions. Formally, we consider the problem of finding a parameter vector $\theta(x) \in \mathbb{R}^p$ that is a solution to a set of conditional moment equations of the form:

$$\mathbf{E}[\psi(Z;\theta(x))|X = x] = 0\,, \tag{1}$$

when given n i.i.d. samples $(Z_1, \ldots, Z_n)$ from the distribution of Z, where $\psi : \mathcal{Z} \times \mathbb{R}^p \to \mathbb{R}^p$ is a known vector valued moment function, $\mathcal{Z}$ is an arbitrary data space, $X \in \mathcal{X} \subset \mathbb{R}^D$ is the feature vector that is included Z.

We show that even when the dimension D of the conditioning variable is larger than the sample size n, estimation and inference is feasible as long as the distribution of the conditioning variable has small intrinsic dimension d, as measured by locally low doubling measures.

Our estimation is based on a sub-sampled ensemble of the k-nearest neighbors (k-NN) Z-estimator. our estimator solves a locally weighted empirical conditional moment equation

$$\hat{\theta}(x) \text{ solves} : \sum_{i=1}^{n} K(x, X_i, S)\, \psi(Z_i; \theta) = 0\,, \tag{2}$$

where $K(x, X_i, S)$ is a *kernel* capturing the proximity of X_i to the target point x. We consider weights $K(x, X_i, S)$ that take the form of an average over B base weights: $K(x, X_i, S) = \frac{1}{B}\sum_{b=1}^{B} K(x, X_i, S_b)\, 1\{i \in S_b\}$, where each $K(x, X_i, S_b)$ is calculated based on a randomly drawn sub-sample S_b of size $s < n$ from the original sample.

We show that if the intrinsic dimension of the covariate distribution is equal to d, then the finite sample estimation error of our estimator is of order $n^{-1/(d+2)}$ and our estimate is $n^{1/(d+2)}$-asymptotically normal, irrespective of D. The sub-sampling size required for achieving these results depends on the unknown intrinsic dimension d. We propose an adaptive data-driven approach for choosing this parameter and prove that it achieves the desired rates. We discuss extensions and applications to heterogeneous treatment effect estimation.

Keywords: non-parametric statistics, inference, intrinsic dimension, conditional moment equation