Proceedings of Machine Learning Research 1–1

Computational-Statistical Gaps in Gaussian Single-Index Models

Alex Damian AD27@PRINCETON.EDU
Princeton University
Loucas Pillaud-Vivien LOUCASPILLAUDVIVIEN@ORANGE.FR
École des Ponts ParisTech, CERMICS
Jason D. Lee JASONDLEE88@GMAIL.COM
Princeton University
Joan Bruna BRUNA@CIMS.NYU.EDU
New York University

Abstract

Single-Index Models are high-dimensional regression problems with planted structure, whereby labels depend on an unknown one-dimensional projection of the input via a generic, non-linear, and potentially non-deterministic transformation. As such, they encompass a broad class of statistical inference tasks, and provide a rich template to study statistical and computational trade-offs in the high-dimensional regime.

While the information-theoretic sample complexity to recover the hidden direction is linear in the dimension d, we show that computationally efficient algorithms, both within the Statistical Query (SQ) and the Low-Degree Polynomial (LDP) framework, necessarily require $\Omega(d^{k^\star/2})$ samples, where $k^\star$ is a "generative" exponent associated with the model that we explicitly characterize. Moreover, we show that this sample complexity is also sufficient, by establishing matching upper bounds using a partial-trace algorithm. Therefore, our results provide evidence of a sharp computational-to-statistical gap (under both the SQ and LDP class) whenever $k^\star > 2$. To complete the study, we construct smooth and Lipschitz deterministic target functions with arbitrarily large generative exponents $k^\star$.

Keywords: Single-Index Models, Statistical Queries, Low-Degree Polynomials

If $X \in \mathbb{R}^d$ is a random data point and $Y \in \mathbb{R}$ is its corresponding label, we say that (X, Y) follow a *Gaussian single-index model* with planted direction $w^\star$ if the marginal of X is $N(0, I_d)$, and if $\mathbb{P}[Y|X]$ depends only on the projection of X in the $w^\star$ direction, i.e. $\mathbb{P}[Y|X] = \mathbb{P}[Y|X \cdot w^\star]$. The goal is to recover $w^\star$ given n i.i.d. samples $\{(x_i, y_i)\}_{i=1}^n$ from a Gaussian single-index model.

We begin by defining the *generative exponent* $k^\star$. If $\{h_k\}_{k\geq 0}$ are the Hermite polynomials, then $k^\star$ is defined to be the smallest $k \geq 1$ such that $\mathbb{E}[h_k(X \cdot w^\star)|Y]$ is not identically 0 in $L^2(\mathbb{P}_Y)$.

Our first result shows that while the information-theoretic sample complexity for this problem is $n \gtrsim d$, computationally efficient algorithms in both the statistical query and low-degree polynomial classes require $n \gtrsim d^{k^\star/2}$ samples where $k^\star$ is the generative exponent of the single index model.

Theorem 1 (informal) *Any low-degree polynomial learner and any statistical query algorithm making polynomialy many queries need $n \geq \tilde{\Omega}(d^{k^\star/2})$ samples from a Gaussian single-index model with generative exponent $k^\star$ to recover $w^\star$.*

We supplement this with a matching upper bound that shows this sample complexity is tight:

Theorem 2 (informal) *There exists a polynomial time algorithm that recovers $w^\star$ up to error ϵ given $n \gtrsim d^{k^\star/2} + d/\epsilon^2$ samples from a Gaussian single-index model with generative exponent $k^\star$.*

[*] Extended abstract. Full version appears as [https://arxiv.org/abs/2403.05529]