Statistical-Computational Tradeoffs in High-Dimensional Single Index Models

Lingxiao Wang, Zhuoran Yang, Zhaoran Wang

Abstract

- We study the statistical-computational tradeoffs in the single index model (SIM).
- We characterize the computational cost via statistical query model.
- When $\mathrm{Cov}(Y, X^\top \beta^*)$ is small, no algorithm can achieve the information-theoretic limit within polynomial oracle complexity.

Statistical Model

We study the following high-dimensional single index model,

$$Y = \begin{cases} f_1(X^\top \beta^*) + \epsilon, & \text{with probability } \alpha, \\ f_2(X^\top \beta^*) + \epsilon, & \text{with probability } 1-\alpha. \end{cases}$$

- Stein's associations:

$$S_1(Y) = \mathrm{Cov}(Y, X^\top \beta^*), \quad S_2(Y) = \mathrm{Cov}(Y, (X^\top \beta^*)^2).$$

- Link functions: f_1 with nonzero Stein's associations and f_2 with zero first-order and nonzero second-order Stein's associations.
- Mixing probability α controls the magnitude of first-order Stein's association.

Associated Testing Problem

$$H_0\colon \beta^* = 0 \quad \text{versus} \quad H_1\colon \beta^* \neq 0.$$

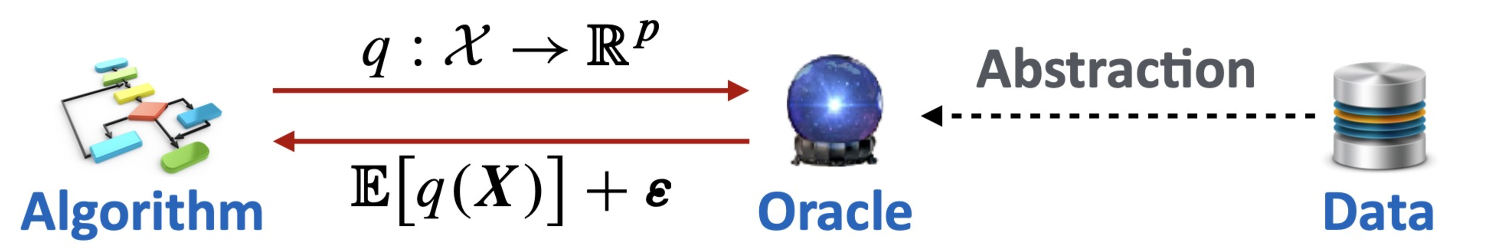
- The difficulty of testing is characterized by the signal-to-noise ratio $\kappa(\beta^*, \sigma) = \|\beta^*\|^2/\sigma^2$.
- Associated parameter spaces:

$$\mathcal{G}_0 = \{(\beta^*, \sigma) \in \mathbb{R}^{d+1}\colon \beta^* = 0\},$$
$$\mathcal{G}_1(s, \gamma_n) = \{(\beta^*, \sigma) \in \mathbb{R}^{d+1}\colon \|\beta^*\|_0 = s, \kappa(\beta^*, \sigma) \geq \gamma_n\}.$$

- Worst-case risk: sum of type-I and type-II errors. Minimax risk: minimal worst-case risk possible given the hardest model.
- Minimax seperation rate (γ_n^*): for $\gamma_n = o(\gamma_n^*)$, minimax risk converges to 1. For $\gamma_n = \Omega(\gamma_n^*)$, minimax risk converges to 0.

Oracle Computational Model

High-Level Computational Primitive *[Nemirovski, Yudin'83; Kearns'93]*



- Algorithm queries a statistical oracle with query function $q : \mathcal{X} \mapsto \mathbb{R}^d$.
- Statistical oracle responds with noise: $\mathbb{E}[q(X)] + \epsilon$.
- Oracle complexity: number of rounds that the algorithm queries the statistical oracle.
- Computational risks and separation rate ($\bar{\gamma}_n^*$): risks and separation rates restricted to algorithms with polynomial query complexity (computationally tractable algorithms).

Lower Bounds

Proposition 1. Let β^* be sparse such that $s = o(d^{1/2-\delta})$ for some absolute constant $\delta > 0$. For

$$\gamma_n = o\left(\sqrt{\frac{s \log d}{n}} \bigwedge \frac{1}{\alpha^2} \cdot \frac{s \log d}{n}\right),$$

any test is asymptotically powerless.

Theorem 1. Let β^* be sparse such that $s = o(d^{1/2-\delta})$ for some absolute constant $\delta > 0$. For

$$\gamma_n = o\left(\left\{\sqrt{\frac{s^2}{n}} \bigwedge \frac{1}{\alpha^2} \cdot \frac{s}{n}\right\} \bigvee \gamma_n^*\right),$$

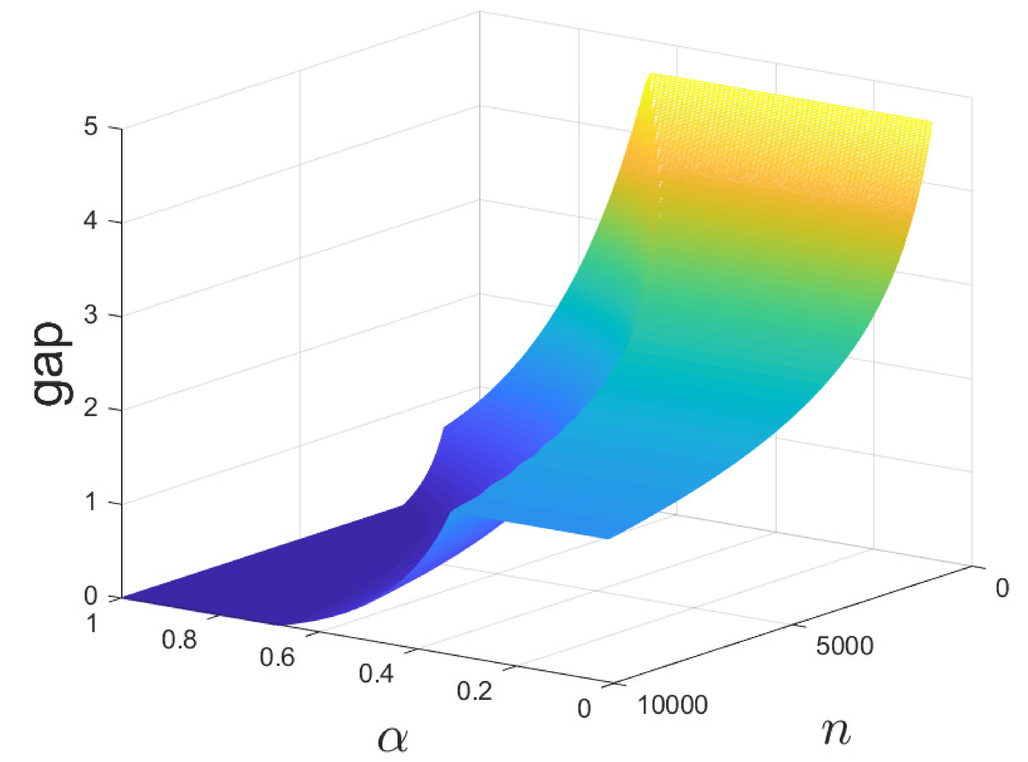
any computational tractable test is asymptotically powerless.

Matching Upper Bounds There exist algorithms and computationally tractable algorithms that attains the lower bounds above.

$$\gamma_n^* = \sqrt{\frac{s \log d}{n}} \bigwedge \frac{1}{\alpha^2} \cdot \frac{s \log d}{n}, \quad \bar{\gamma}_n^* = \sqrt{\frac{s^2}{n}} \bigwedge \frac{1}{\alpha^2} \cdot \frac{s \log d}{n}.$$

Phase Transition

- Gap: the difference between $\bar{\gamma}_n^*$ and γ_n^*.



- For $0 < \alpha \leq ((\log d)^2/n)^{1/4}$, the gap is invariant to α.
- For $(\log^2 d/n)^{1/4} \leq \alpha \leq (s \log d/n)^{1/4}$, a larger α implies a smaller gap.
- For $(s \log d/n)^{1/4} < \alpha \leq 1$, the gap vanishes.

Implication to Parameter Estimation

Theorem 2. For $n = o(s^2/\gamma_n^2 \wedge s \log d/(\gamma_n \cdot \alpha^2))$, any computationally tractable algorithm that estimates β^* is inconsistent. Specifically, it holds that

$$\bar{\mathbb{P}}(\|\widehat{\beta} - \beta^*\|_2 \geq \sigma \cdot \|\beta^*\|_2^{-1} \cdot \gamma_n/4) \geq C,$$

where $C > 0$ is an absolute constant.