Problem Set 4

Samuel B. Hopkins

Last updated November 20, 2024

Due: 11/??, 11:59pm.
Please typeset your solutions in LaTeX.

Problem 1 (Sparse robust mean estimation)**.** In this problem, we will solve a sparse version of robust mean estimation. Let $\mu \in \mathbb{R}^d$ be an unknown k-sparse vector, in that only k of its entries are non-zero. First $n = \widetilde{\Omega}(k^2(\log d)/\varepsilon^2)$ samples $v_1, \dots, v_n \in \mathbb{R}^d$ are drawn from $\mathcal{N}(\mu, \mathrm{Id})$. Then an adversary alters εn of the samples and reorders them arbitrarily. We observe the resulting dataset $v'_1, \dots, v'_n$. Our goal will be to give an algorithm for estimating μ from these samples.

(a) Let $\overline{v} = \frac{1}{n}\sum_{i=1}^n v_i$. Prove that with 0.99 probability, for all k-sparse vectors $u \in \mathbb{R}^d$ with $\|u\| = 1$,

$$\langle u, \overline{v} - \mu \rangle^2 \leq \varepsilon^2 \,.$$

(b) Define $\Sigma = \frac{1}{n}\sum_{i=1}^n (v_i - \overline{v})(v_i - \overline{v})^T$. Prove that with 0.99 probability, $|\Sigma_{ij}| \leq 1/k$ for $i \neq j$ and $|\Sigma_{ii} - 1| \leq 1/k$ for all $i, j \in [d]$.

(c) Consider the following system, which we call $\mathcal{S}$, with scalar variables $w_1, \dots, w_n$ and d-dimensional variables $z, z_1, \dots, z_n$

$$\begin{aligned}
&w_i^2 = w_i \\
&\sum_{i=1}^n w_i \geq (1-\varepsilon)n \\
&w_i(z_i - v'_i) = 0 \\
&\overline{z} = \frac{1}{n}\sum_{i=1}^n z_i \,,\ \Sigma = \frac{1}{n}\sum_{i=1}^n (z_i - \overline{z})(z_i - \overline{z})^T \\
&-\frac{1}{k} \leq \Sigma_{ij} \leq \frac{1}{k} \quad \text{for all } i \neq j \\
&-\frac{1}{k} \leq \Sigma_{ii} - 1 \leq \frac{1}{k} \quad \text{for all } i
\end{aligned}$$

Prove that with 0.99 probability, there is a feasible solution to this system where the w_i are indicators of the clean samples and the z_i are the actual clean samples.

From now on, assume that the events in (a), (b), (c) hold.

(d) Now we consider the SoS relaxation of the system $\mathcal{S}$. Let $u \in \mathbb{R}^d$ be an arbitrary k-sparse vector with $\|u\| = 1$. Prove that

$$\mathcal{S} \vdash_2 \sum_{i=1}^{n} \langle u, z_i - v_i \rangle^2 \le 10n(1 + \langle u, \overline{z} - \mu \rangle^2)$$

where recall v_i are the clean samples drawn from $N(\mu, I)$.

(e) Let $u \in \mathbb{R}^d$ be an arbitrary k-sparse vector with $\|u\| = 1$. Use part (c) to prove that

$$\mathcal{S} \vdash_4 \langle u, \overline{z} - \overline{v} \rangle^2 \le 100\varepsilon(1 + \langle u, \overline{z} - \mu \rangle^2)$$

(f) Use part (e) to deduce that

$$\mathcal{S} \vdash_4 \langle u, \overline{z} - \mu \rangle^2 \le O(\varepsilon)\,.$$

Put everything together to show that there is a polynomial time algorithm that takes the samples $v'_1, \ldots, v'_n$ and with probability 0.9, outputs a k-sparse $\widehat{\mu}$ such that $\|\mu - \widehat{\mu}\| \le O(\sqrt{\varepsilon})$.

Problem 2. Recall the *planted clique* problem, with the "null distribution" $\mathcal{N} = G(n, 1/2)$, and the "planted distribution" $\mathcal{P}$ obtained by drawing G from $G(n, 1/2)$, and adding a uniformly random k-clique. It is believed that for k significantly smaller than $O(\sqrt{n})$ (say $O(n^{1/2-\varepsilon})$), it is computationally hard to distinguish these two distributions. In this question, we will establish this computational hardness for the restricted class of algorithms based on low-degree polynomials.

Concretely, set $k = O(n^{1/2-\varepsilon})$ for some (small) constant $\varepsilon > 0$, and $D \le C \log n$ for some (large) constant $C > 0$. Recall the degree-D χ^2-divergence, defined by

$$\sqrt{\chi^2_{\le D}(\mathcal{P} \| \mathcal{N})} = \max_{\substack{F:\{\text{set of graphs on } n \text{ vertices}\} \to \mathbb{R} \\ F \text{ degree} \le D \text{ polynomial} \\ F \text{ not identically } 0}} \frac{\mathbb{E}_{\mathcal{P}}[F] - \mathbb{E}_{\mathcal{N}}[F]}{\sqrt{\mathbf{Var}_{\mathcal{N}}[F]}}.$$

Further recall that this maximum is attained by the function $\left(\frac{\mathcal{P}}{\mathcal{N}}\right)^{\le D}$, where $\frac{\mathcal{P}}{\mathcal{N}}$ is the likelihood ratio $\frac{\mathcal{P}}{\mathcal{N}}(G) = \frac{\mathcal{P}(G)}{\mathcal{N}(G)}$ and the notation $f^{\le D}$ means the projection of f to the space of degree D polynomials. This resulting maximum is equal to

$$\chi^2_{\le D}(\mathcal{P} \| \mathcal{N}) = \left\| \left(\frac{\mathcal{P}}{\mathcal{N}}\right)^{\le D} - 1 \right\|_2^2,$$

with the notation $\|f\|_2^2 = \mathbb{E}_{\mathcal{N}} f^2$.

(a) Let $g = \left(\frac{\mathcal{P}}{\mathcal{N}}\right)^{\le D}$ be a polynomial of degree (at most) D in the variables $(x_e)_{e \in \binom{[n]}{2}}$, where $x_e = 1$ if e is an edge in the graph, and -1 otherwise. Express g in terms of its Fourier coefficients as $g = \sum_{\alpha : |\alpha| \le D} \widehat{g}_\alpha x^\alpha$. Determine $\widehat{g}_\alpha$.

(b) Show that in the given parameter regime of k, D, $\chi^2_{\le D}(\mathcal{P} \| \mathcal{N}) = \|g - 1\|^2 = o(1)$.