Homework 1

October 22, 2019

Problem 1: **Completing the picture for univariate Gaussian robust mean estimation.** Recall that in Lecture 1 we showed that the median recovered the mean of a Gaussian with additively corrupted samples from a Gaussian $\mathcal{N}(\mu, 1)$ to error $O(\varepsilon)$, and in Lecture 2 we showed that no algorithm can do better than $\Omega(\varepsilon)$ given corrupted samples from $\mathcal{N}(\mu, 1)$. Here we'll extend these bounds, and show that $\Theta(\varepsilon)$ is the right answer, for all these models of corruption.

(a) Verify that for ε sufficiently small, the median still achieves $O(\varepsilon)$ error to μ with high probability, given ε-corrupted samples from $\mathcal{N}(\mu, 1)$.

(b) Show that for any $\varepsilon < 1/2$ and any two distributions D_1, D_2, if $\mathrm{d_{TV}}(D_1, D_2) = \varepsilon/(1-\varepsilon)$, then there exists a distribution U so that $U = (1-\varepsilon)D_1 + \varepsilon N_1 = (1-\varepsilon)D_2 + \varepsilon N_2$ for some noise distributions N_1, N_2. Conclude that no algorithm can learn the mean of a Gaussian with variance 1 given ε-obliviously additively corrupted samples to error better than $\Omega(\varepsilon)$.

Problem 2: **Population level spectral signatures.** In this problem we will prove Lemma 4.1 from Lecture 4, step-by-step. We reproduce the lemma below for completeness.

Lemma 0.1. *Let $\varepsilon \in [0, 1/2)$, and let $\delta > 0$. Let D be a distribution over $\mathbb{R}^d$ with mean μ and covariance $\Sigma \preceq I$. Let $X_1, \ldots, X_m \sim D$ be i.i.d random variables. Then, there exist universal constants c, c' so that with probability $1 - \delta - \exp(-\Omega(\varepsilon m))$, there exists a set $S_{\text{good}} \subseteq [m]$ so that $|S| \geq (1-\varepsilon)m$ and:*

$$\|\widehat{\mu} - \mu\|_2 \lesssim \sqrt{\frac{d}{m\delta}} + \sqrt{\varepsilon} \tag{1}$$

$$\left\| \frac{1}{|S_{\text{good}}|} \sum_{i \in S_{\text{good}}} (X_i - \widehat{\mu})(X_i - \widehat{\mu})^\top \right\|_2 \lesssim \frac{d(\log d + \log 1/\delta)}{\varepsilon m}, \tag{2}$$

where $\widehat{\mu} = \frac{1}{|S_{\text{good}}|} \sum_{i \in S_{\text{good}}} X_i$.

The set S_{good} we will take is actually quite simple. For some constant $\alpha > 0$, define the event

$$E = \left\{ X : \|X - \mu\|_2 \leq \sqrt{\frac{d}{\alpha\varepsilon}} \right\},$$

and let $S_{\text{good}} = \{X_i : X_i \in E\}$.

(a) Show that with probability $1 - \exp(-\Omega(\varepsilon m))$, we have that $|S_{\text{good}}| \geq (1-\varepsilon)m$.

(b) Show that $\|\mathbb{E}[(X-\mu)\mathbb{I}_{X \in E}]\|_2 \lesssim \sqrt{\varepsilon}$. Conditioned on $|S_{\text{good}}| \geq (1-\varepsilon)m$, conclude that with probability $1 - \delta$, (1) holds.

(c) Conditioned on $|S_{\text{good}}| \geq (1-\varepsilon)m$, prove (2) holds with probability $1 - \delta$. The following matrix Chernoff bound will be useful, and you may use it without proof:

Fact 0.2. *Let $M_1, \ldots, M_n \in \mathbb{R}^{d\times d}$ be a sequence of independent random PSD matrices. Assume that $\|M_i\|_2 \leq L$ for all $i = 1, \ldots, n$ almost surely, and suppose that $\|\mathbb{E}\left[\sum_{i=1}^n M_i\right]\|_2 \leq n$. Then, for all $t \geq 2$, we have*

$$\Pr\left[\left\|\sum_{i=1}^n M_i\right\|_2 \geq tn\right] \leq d\exp\left(-\Omega(tn/L)\right) \ .$$

Problem 3: **Breakdown points.** The *breakdown point* of an estimator is the largest ε so that for any d, there is $n = f(d)$ so that given a set of ε-corrupted data of size n from a distribution D with bounded second moments (or more generally, from any given class of distributions), the estimator achieves bounded error as $d \to \infty$ with probability at least $9/10$ (this constant is arbitrary). In Lecture 5 we argued that the breakdown point of the filter was at least $\varepsilon = 0.134$.

Change the invariant preserved by the filtering algorithm for bounded second moments to

$$\alpha \sum_{i\in S_{\text{good}}} w_i\tau_i < \sum_{i\in S_{\text{bad}}} w_i\tau_i \ ,$$

for $\alpha \in [0, \infty)$.

(a) Demonstrate that by changing constants in the algorithm, we can maintain this more general invariant.

(b) By optimizing α, how large of a breakdown point can you achieve?

Problem 4: **From TV to Frobenius norm.** Let $\Sigma_2 \succ 0$. Prove that

$$d_{\text{TV}}(\mathcal{N}(0, I), \mathcal{N}(0, \Sigma_2)) \lesssim \|I - \Sigma_2\|_F \ .$$

Hint: Use Pinsker's inequality.

Problem 5: **Completing the picture for robustly learning Gaussians.** Let $\varepsilon > 0$ be sufficiently small, and let $\Sigma \succ 0$. Throughout this problem, suppose you have an polynomial-time estimator which, given ε-corrupted samples S from $\mathcal{N}(0, \Sigma)$, outputs $\widehat{\Sigma}$ so that $\left\|\Sigma - \widehat{\Sigma}\right\|_\Sigma \leq \delta$.

(a) Give an polynomial-time estimator which, given $\varepsilon/2$-corrupted samples from $\mathcal{N}(\mu, \Sigma)$ for μ and Σ both unknown, outputs $\widehat{\Sigma}$ so that $\left\|\Sigma - \widehat{\Sigma}\right\|_\Sigma \leq \delta$.

(b) Verify that the Gaussian filter presented in Lecture 7 still can achieve non-trivial recovery, if the covariance Σ of the Gaussian is unknown but satisfies $\|\Sigma - I\|_2 < \delta$. What is the final error you get, as a function of ε and δ?

(c) Using parts (a) and (b), give an polynomial-time algorithm which, given an ε-corrupted set of samples from $\mathcal{N}(\mu, \Sigma)$, outputs $\widehat{\mu}$ and $\widehat{\Sigma}$ so that

$$d_{\text{TV}}(\mathcal{N}(\mu, \Sigma), \mathcal{N}(\widehat{\mu}, \widehat{\Sigma})) \lesssim \delta + \sqrt{\varepsilon \log 1/\varepsilon} \ .$$

As a remark, the best efficiently achievable δ is $O(\varepsilon \log 1/\varepsilon)$, and so this yields an algorithm which achieves overall error $O(\varepsilon \log 1/\varepsilon)$.