Homework 1

October 12, 2025

Instructions: Please submit a written solution to at least **one** problem below. You are welcome to submit more, but your grade will be the highest score for any single problem you submit.

Problem 1: **Completing the picture for univariate robust mean estimation.** Recall that in Lecture 1 we showed that the truncated mean recovered the mean of a distribution D with mean μ and variance σ^2 with additively corrupted samples. However, the instructor was very sloppy about constants, and moreover, the algorithm was presented only for additive corruption. Here you'll work out the complete picture.

(a) (Breakdown point.) It is a bit of a pain to formally define breakdown point, but intuitively, it just captures the largest fraction of outliers that an estimator can tolerate before its error becomes unbounded large. Formally, an estimator for this problem is a collection of (potentially randomized) functions $\{A_n\}_{n=1}^{\infty}$, where $A_n : \mathbb{R}^n \to \mathbb{R}$ is a function which takes a dataset S of size n and outputs a candidate mean $A_n(S)$ for this dataset. One definition of the *breakdown point* of this estimator relative to some class of distributions $\mathcal{D}$ can be defined to be the largest $\varepsilon_0 > 0$ so that for all $\varepsilon < \varepsilon_0$, we have that

$$\sup_{D \in \mathcal{D}} \lim_{n \to \infty} \mathop{\mathbb{E}}_{S \sim_{n,\varepsilon} D} [|A(S) - \mu_D|] < \infty \,, \tag{1}$$

where we let $S \sim_{n,\varepsilon} D$ denote that S is an ε-corrupted set of samples from D, and we let μ_D denote the mean of D. You may recall that in class, we (informally) argued that the breakdown point of the median for the class of Gaussian distributions is $1/2$.

Devise a refinement of the truncated mean for robust mean estimation in this setting that achieves breakdown point $1/2$ for the class of distributions with variance at most σ^2.

(b) Obtain tight rates (up to sub-constant factors) for the best error achievable by any algorithm for learning the mean of a distribution with variance at most σ^2, with breakdown point $1/2$. That is, for some constant $C > 0$, demonstrate some algorithm (which may be similar to the one above) that, for all $\varepsilon < 1/2$ and $\delta > 0$, outputs $\widehat{\mu}$ so that with probability $1 - \delta$, we have that

$$|\widehat{\mu} - \mu| \le (1 + o_n(1)) C\sigma\sqrt{\varepsilon} + f(n, 1/\delta) \,,$$

so that $f(n, 1/\delta) \to 0$ as $n \to \infty$, for any fixed δ, and moreover, demonstrate a lower bound, stating that any algorithm, even with unboundedly many samples, must incur error $|\widehat{\mu} - \mu| \ge (1 - o(1)) C\sigma\sqrt{\varepsilon}$.

Problem 2: **Statistical bounds for high-dimensional Gaussians**. In this problem, we will work out a fairly tight bound for the statistical distance between two high-dimensional Gaussians. Throughout this problem, let $\Sigma_1, \Sigma_2 \succ 0$ be two positive definite $d \times d$-sized covariance matrices, and $\mu_1, \mu_2 \in \mathbb{R}^d$.

(a) First, prove that

$$\mathrm{d_{TV}}(\mathcal{N}(\mu_1, I), \mathcal{N}(\mu_2, I)) \lesssim \min\left(\|\mu_1 - \mu_2\|_2, 1\right) \,.$$

(b) Next, prove that

$$\mathrm{d_{TV}}(\mathcal{N}(0,I),\mathcal{N}(0,\Sigma_2)) \lesssim \min\left(\|I-\Sigma_2\|_F, 1\right) .$$

Hint: Use Pinsker's inequality.

(c) From this, derive a general form for the TV distance between $\mathcal{N}(0,\Sigma_1)$ and $\mathcal{N}(0,\Sigma_2)$ in terms of Mahalanobis distance. Crucially, this bound should yield non-vacuous results (i.e. TV distances which are $\ll 1$) even when Σ_1 and Σ_2 could be very ill-conditioned.

(d) Now, using these sub-problems, derive a bound for the total variation distance between $\mathcal{N}(\mu_1,\Sigma_1)$ and $\mathcal{N}(\mu_2,\Sigma_2)$.

Hint: Note that there are two ways that a mean shift could cause a large TV distance between two Gaussians: either as in part (a) where the distance is legitimately large in some direction witnessed by *both* covariances, or, where the distance is in some direction *not* witnessed by at least one covariance matrix.

(e) **Extra credit:** For imaginary extra-credit points, prove that your bound is tight up to constant factors.

Problem 3: **Population level spectral signatures.** In this problem you will prove Lemma 4.1 from Lecture 4. 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 $X_1,\dots,X_m \sim D$ 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 $S_{\mathrm{good}} \subseteq [m]$ so that $|S| \geq (1-\varepsilon)m$ and:*

$$\|\widehat{\mu}-\mu\|_2 \lesssim \sqrt{\frac{d}{m\delta}} + \sqrt{\varepsilon} \tag{2}$$

$$\left\| \frac{1}{|S_{\mathrm{good}}|} \sum_{i\in S_{\mathrm{good}}} (X_i-\widehat{\mu})(X_i-\widehat{\mu})^\top \right\|_2 \lesssim \frac{d(\log d + \log 1/\delta)}{\varepsilon m} , \tag{3}$$

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

The following matrix Chernoff bound will be useful, and you may use it without proof:

Fact 0.2. *Let $M_1,\dots,M_n \in \mathbb{R}^{d\times d}$ be a sequence of independent random PSD matrices. Assume that $\|M_i\|_2 \leq L$ for all $i=1,\dots,n$ 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*

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

Hint: The naive way to apply this matrix Chernoff bound would be to attempt to take $M_i = (X_i-\widehat{\mu})(X_i-\widehat{\mu})^\top$. However, this is tricky in two ways: first, it a necessary condition to apply this matrix Chernoff bound is an *deterministic* bound on the M_i, and second, $\widehat{\mu}$ makes these random matrices dependent.

Problem 4: **Implementation of the filter**

Write (but don't submit) code that implements the spectral filter for bounded second moment distributions, as discussed in class. Recall that a key subroutine for this spectral filter is a downweighting or removal scheme, where given non-negative scores $\tau_1,\dots,\tau_n$, and potentially weights $w_1,\dots,w_n$, either removes points from the dataset or downweights the weights based on the scores. Write three variants of the downweighting scheme:

(i) The weighted downweighting scheme as described in class:

$$w_i \leftarrow \left(1 - \frac{\tau_i}{\max_i \tau_i}\right) w_i \ .$$

(ii) The independent subsampling scheme: for each point, throw it out of the dataset with probability $p_i \propto \frac{\tau_i}{\max_i \tau_i}$.

(iii) A randomized threshold subsampling scheme: choose a threshold $T \in [0, \max_i \tau_i]$ uniformly at random, and throw away all points so that $\tau_i \geq T$.

In the rest of the problem you will compare their performance against each other.

(a) Try running these schemes on synthetically generated data. Find an inlier distribution D with covariance $\Sigma \preceq I$ and a ε-corruption scheme for this distribution so that empirically, all three of these algorithms pay $O(\sqrt{\varepsilon})$ error when given ε-corrupted samples from D. What differences do you notice in the behavior of these algorithms?

(b) Try implementing another scheme, in which you design some *deterministic* threshold T, and throw away all points so that $\tau_i \geq T$. Compare the performance of the error of this scheme.

(c) Find a corruption scheme which forces all three of the downweighting schemes to run for as many iterations of the filter as possible. How many iterations as a function of the dimension are necessary?

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 a 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) Redo the analysis of the the Gaussian filter to demonstrate that it can still 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?

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

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

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