Privacy in Statistics and Machine Learning — Spring 2021
In-class Exercises for Lecture 13 (Gradient Descent)
March 11 & 12, 2021

Adam Smith and Jonathan Ullman

Problems with marked with an asterisk (*) are more challenging or open-ended.

1. (Factorization example) Consider the following set of linear queries, expressed as a matrix:

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

   What is $\|\mathbf{F}\|_{1\to 2}$ and what is the (expected ℓ_2-norm) error of the gaussian mechanism for these queries? Find a factorization $\mathbf{RM} = \mathbf{F}$ with strictly lower error. [*Note*: there is a relatively simple factorization with lower error, but we did not work out what the *best* factorization is.]

2. (Composing factorizations) Suppose we already have a good factorization $\mathbf{RM} = \mathbf{F}$ for one set of queries $\mathbf{F}$ (e.g. threshold queries). Now suppose someone comes along with another set of queries $\mathbf{F}'$ such that $\mathbf{R}'\mathbf{F} = \mathbf{F}'$ (e.g. interval queries). Suppose we answer $\mathbf{F}'$ privately in the following way: run the factorization mechanism with $\mathbf{R}, \mathbf{M}$ to answer $\mathbf{F}$ and then transforming its answers using $\mathbf{R}'$. Express the error if this new mechanism in terms of appropriate quantities involving $\mathbf{R}, \mathbf{M}, \mathbf{R}'$.

3. (Binary tree as a factorization mechanism)

   (a) Express the binary tree mechanism from Lecture 9 as an instance of factorization. Specifically, for the domain $\mathcal{U} = \{1, \ldots, 8\}$, and the set of threshold queries $f_t(\mathbf{x}) = \frac{1}{n} \cdot \#\left\{j : x_j \le t\right\}$ for $t = 1, \ldots, 8$, write the matrix $\mathbf{F}$, the matrix $\mathbf{M}$ describing the set of queries in the binary tree, and the matrix $\mathbf{R}$ describing how to reconstruct the answers to the threshold queries.

   (b) In the binary tree mechanism, there are queries for which we can get two independent estimates. For example, the number of points in $\{1, ..., t\}$ can be estimated through (a) adding the noisy counts for $\log D$ intervals to the left of t, and (b) n minus the sum of noisy counts for $\log D$ intervals to the right of t.

   Combining two independent estimates with the same noise distribution reduces variance (say, by averaging). Will the matrix mechanism capture this kind of optimization automatically, or does it lie outside of the class of algorithms captured by the mechanism?

   (c) Express the binary tree mechanism and its error analysis for a general domain $\mathcal{U} = \left\{1, \ldots, 2^\ell\right\}$ as a factorization mechanism. [Notes: I suggest just getting the idea of what it looks like and how the relevant matrix norms scale with $|\mathcal{U}|$, since writing it with precise notation is going to be a mess. Also you shouldn't be concerned if your error bound isn't quite the same as it was in Lecture 9, because we're analyzing the mechanism for Gaussian noise instead of Laplace, and ℓ_2 error instead of the maximum error.]