Advanced Topics in Machine Learning 2024

Rasmus Pagh

Home Assignment 1

Deadline: 23:59, Thursday September 19, 2024

The assignments must be submitted individually – each student must write and submit a personal solution but we do not prevent you from discussing ***high-level*** *ideas in small groups. If you use any LLM tool such as ChatGPT, please specify the* ***purpose*** *and* ***manner*** *in which you have utilized it.*

We are interested in how *you solved the problems, and not just in the final answers. Please explain your solutions and ideas as clearly as you can.*

Late Penalty and multiple Submissions *Late submissions will incur a penalty of 10% of the total marks for every hour of delay (rounded up) with a maximum allowed delay of 5 hours after which the submission server will close. If you submit multiple submissions, only the last submission will be considered relevant both for grading answers as well as late penalty.*

Submission format: *Please upload your answers in a single* `.pdf` *file. If you have created code please include at least the key parts in the PDF as well as a link to the full code (on Github, Colab, or similar).*

1 Reconstruction attacks (60 points)

In this problem we consider a sensitive dataset $x \in \{-1,1\}^n$ (think of a sensitive binary attribute for each of n people). We consider the bounded setting where neighboring n-dimensional datasets differ in one coordinate. A mechanism is available that supports statistical queries on x. Specifically, for a query $q \in \{-1,1\}^n$ the mechanism computes the dot product $\langle x, q\rangle = \sum_{i=1}^n x_i q_i$ and releases $\langle x, q\rangle + z$ where $z \sim \mathrm{Lap}(\lambda)$ is Laplace-distributed noise added to protect the privacy of individial entries of x.

(a) Argue that any single query is ε-differentially private with $\varepsilon = 2/\lambda$.

(b) Perform a reconstruction attack by making queries to a remote database, using the code template. The remote database contains datasets of size $n = 100$, uses $\lambda = 10$, and rounds results to the nearest integer in $[-n, n]$. It is suggested to use the attack of Dinur and Nissim discussed in the lecture, in the version that minimizes the ℓ_1 distance to the observed noisy query results, but it is allowed to use other equally effective techniques. The maximum number of queries made before submitting a (partial) reconstruction should be limited to $t = 2n$. (It is possible to make more queries, but only submissions after at most t queries to a given dataset are considered valid.) State your linear program, provide your code along with comments/explanation, and report on the percentage of correct entries by including a screenshot from the output of your code. (It is possible to achieve about 80% correct or better.)

2 Differential privacy (40 points)

We have seen how *randomized response* can be used to make binary data differentially private. Now suppose that data is not from $\{0,1\}$ but comes from the set $\mathcal{X} = \{A, C, G, T\}$. Suppose we encode values in $\mathcal{X}$ as 2-bit strings, i.e., consider $\mathcal{X} = \{00, 01, 10, 11\}$. Consider a mechanism $\mathcal{M}(x)$ that for input $x \in \mathcal{X}$ uses the *randomized response* algorithm from *Dwork and Roth* (sec. 3) independently on *each bit*.

(c) Argue that $\mathcal{M}$ satisfies $(\ln 9)$-differential privacy.

(d) Propose a different mechanism $\mathcal{M}'$ that still satisfies $(\ln 9)$-differential privacy and satisfies the inequality $\Pr[\mathcal{M}'(x) = x] > \Pr[\mathcal{M}(x) = x]$, i.e., outputs x with a higher probability than $\mathcal{M}$.

Good luck!

Rasmus