ISYE 8813: Algorithmic Foundations of Ethical ML **Fall 2021**

Lectures 7-8: Sparse Vector

Lecturer: Juba Ziani

So far, in this class, we have seen how to answer different types of queries in a differentially private manner, under different notions of differential privacy. First, we have seen how to answer simple numerical queries, though the Laplace mechanism. Then, we have seen how to answer more complex, optimization queries (what is the hypothesis that maximizes some utility or minimizes some loss function) in a ε-DP manner. We also saw how to satisfy a slightly less stringent notion of privacy, through the Gaussian mechanism, which provides privacy with high probability but still leaves a very small probability δ that the privacy guarantees do not hold. Finally, we started seeing some of the essential building blocks to compose private mechanisms together, and to perform operations on the outcomes of such mechanism.

But what happens when we want to be able *not* to answer queries? What if we need to answer many queries in row, but we want to preserve our privacy budget? In cases in which we only care if the value of a query is important enough/above a certain threshold, and we have good reason to believe there are few queries that are above this threshold (so, we are in a *sparse* case), we may want not to answer every query that comes our way. rather, we want to select in a differentially private way *which* of these queries we are going to answer. That way, by answering only a small subset of the queries, we can try to preserve our privacy budget as much as possible.

Example 1. *In a machine learning setting, we may want to test a large number of features and try to understand their correlation with the labels. We may want to identify which features matter a lot, i.e. have at least a certain level of correlation with the label, so that we can limit our model to only take these "good" features into account.*

Another interesting application of this technique is for generalization in adaptive data analysis: one can use such a technique to detect, in a privacy-preserving manner, whether we are over-fitting to the data/not properly generalizing; when answering a large number of adaptive queries, we want to make sure that we do not overfit too often, otherwise the answer to our adaptively-chosen queries on the dataset may get further and further away from the truth. (If interested in this, note that this is one of the topics that you can read and write about later in this class).

This is what the *Sparse Vector* technique is going to help us with here. The mechanism works by taking in a long stream of queries, adding Laplace noise to each answer, comparing each (noisy) answer against a noisy (Laplace) threshold, and only outputting the answer to those queries with values above the threshold.

1 A building block for SparseVector: AboveThreshold

We first start with a simplified version of the mechanism: see Algorithm 1. AboveNoisyThreshold is a mechanism that keeps running queries until it detects that a query is above the desired threshold, then halts.

Algorithm 1: AboveNoisyThreshold($x, \{f_i\}, T, \epsilon$)

```
Input: database $x$, adaptively chosen stream of sensitivity-1 queries $\{f_i\}$, threshold $T$, privacy parameter $\epsilon$
Output: Stream of answers $\{a_i\} \in \{\bot, \top\}^*$
Let $\hat{T} = T + Lap(\frac{2}{\epsilon})$
for each query $f_i$ do
    Let $v_i = Lap(\frac{4}{\epsilon})$ if $f_i(x) + v_i \geq \hat{T}$ then
        output $a_i = \top$ Halt
    else
        output $a_i = \bot$
    end
end
```

Theorem 2. *AboveNoisyThreshold is $(\epsilon, 0)$-differentially private.*

Proof. See [1], p57-58. Slightly cleaner proof here:
`http://www.gautamkamath.com/CS860notes/lec9.pdf`. □

We also want to make sure that we are identifying the right query for this mechanism, that is above the desired threshold with high probability. For that, we need an accuracy guarantee. Note that we can't use our previous accuracy notions which say the answers provided by the mechanism are close to the true answers because AboveNoisyThreshold does not produce numeric answers. In turn, we need the new, following definition of accuracy:

Definition 3 (Accuracy)**.** *A mechanism that outputs a stream of answers $\{a_i\} \in \{\bot, \top\}^*$ to a stream of k queries $\{f_i\}$ is (α, β)-accurate with respect to a threshold T if, with probability at least $1 - \beta$, the mechanism does not halt before f_k, and*

$$\forall a_i = \top : f_i(x) \geq T - \alpha$$
$$\forall a_i = \bot : f_i(x) \leq T + \alpha$$

This definition requires that with high probability, the mechanism produces an approximately correct output for all k queries.

Theorem 4. *For any sequence of k sensitivity-1 queries $f_1, \ldots, f_k$ s.t. $|\{i < k : f_i(x) \geq T - \alpha\}| = 0$, then AboveNoisyThreshold is (α, β)-accurate for any $\beta > 0$ and*

$$\alpha = \frac{8(\ln(k) + \ln(\frac{2}{\beta}))}{\epsilon}.$$

Proof. See [1], p59. Slightly cleaner proof here:
`http://www.gautamkamath.com/CS860notes/lec9.pdf.` □

Note that this quantifier $|\{i < k : f_i(x) \geq T - \alpha\}| = 0$ requires that the only query close to being above threshold is possibly the last one. Without this condition, the algorithm would be required to halt before the k^{th} query with high probability, so it couldn't possibly satisfy the accuracy guarantee.

In terms of results, we can start seeing how AboveThreshold and SparseVector allow us to preserve our privacy budget: before, our privacy guarantee was a function of the number of queries k that we were running. The best we knew how to do was to have a privacy guarantee of $\sqrt{k}\varepsilon$ to answer k queries; so, we had to set $\varepsilon' = \frac{\varepsilon}{\sqrt{k}}$ to obtain ε-DP, which means our accuracy guarantee was evolving in $\alpha \sim \frac{\sqrt{k}}{\varepsilon}$; we could answer $k \sim \varepsilon^2\alpha^2$ queries with accuracy guarantee α. But now, the accuracy guarantees tells us that we can roughly answer an exponential number of queries $k \sim \exp(\varepsilon\alpha)$ before our mechanism halts with accuracy α, while still preserving ε (rather than $k\varepsilon$) differential privacy.

2 SparseVector Mechanism

The SparseVector mechanism takes as input a database x, an adaptively chosen stream of sensitivity-1 queries $\{f_i\}$, a threshold T, a total number of numeric answers c, and privacy parameters (ϵ, δ). It outputs a stream of answers $\{a_i\} \in (\mathbb{R} \cup \{\perp\})^*$.

Now let's see the full SparseVector algorithm (Algorithm 2).

Now let's take a closer look at what's going on with this mechanism. The middle part looks a whole lot like AboveNoisyThreshold (ANT). SparseVector (SV) works by repeated calls to ANT as a subroutine up to c times, until our counter reaches the pre-set limit c. Instead of halting after finding an above-threshold query, SV calls the Laplace Mechanism as a subroutine to output a noisy answer to that query.

Notice that we re-draw fresh noise for every call to the Laplace Mechanism and ANT, so SV is really just an adaptive composition of these two mechanisms. We allocate our overally privacy budget ϵ between these two mechanisms, where ϵ_1 is our ANT privacy budget and ϵ_2 is our Laplace Mechanism privacy budget.

Depending on whether our overall privacy goal is $(\epsilon, 0)$-differential privacy or (ϵ, δ)-differential privacy, we'll have to set parameters within these subroutines differently. If we want $(\epsilon, 0)$-differential privacy, we'll end up using Basic Composition[1], so we'll set our privacy parameters so they sum to ϵ. If we want (ϵ, δ)-differential privacy, then we can use Advanced Composition, and we'll set our parameters according to the Corollary that we saw last time.

[1] Note that we only proved Basic Composition for non-adaptive mechanisms, but the result also holds for adaptive mechanisms. The Laplace Mechanism is used adaptively based on the results of the ANT mechanism.

Algorithm 2: SparseVector$(x, \{f_i\}, T, c, \epsilon, \delta)$

```
SparseVector$(x, \{f_i\}, T, c, \epsilon, \delta)$:
Let $\epsilon_1 = \frac{8}{9}\epsilon$ and let $\epsilon_2 = \frac{2}{9}\epsilon$
if $\delta = 0$ then
    Let $\sigma(\epsilon) = \frac{2c}{\epsilon}$;
else
    Let $\sigma(\epsilon) = \frac{\sqrt{32c\ln(2/\delta)}}{\epsilon}$;
end
Let $\hat{T}_0 = T + \text{Lap}(\sigma(\epsilon_1))$
Let count = 0
for each query $f_i$ do
    Let $v_i = \text{Lap}(2\sigma(\epsilon_1))$ if $f_i(x) + v_i \geq \hat{T}_{\text{count}}$ then
        Output $a_i = f_i(x) + \text{Lap}(\sigma(\epsilon_2))$
        Update count = count + 1 and $\hat{T}_{\text{count}} = T + \text{Lap}(\sigma(\epsilon_1))$
    else
        Output $a_i = \perp$
    end
    if count $\geq c$ then
        Halt
    end
end
```

2.1 SparseVector Privacy

Theorem 5. *SparseVector is (ϵ, δ)-differentially private.*

Proof. Case $\delta = 0$:
We first consider the case where $\delta = 0$. SV consists of c runs of ANT, where each run is $(\frac{8}{9c}\epsilon, 0)$-differentially private, and c runs of the Laplace Mechanism, where each run is $(\frac{1}{9c}\epsilon, 0)$-differentially private. Then it will be straightforward to see through Basic Composition that SV is overall $(\epsilon, 0)$-differentially private. All that remains is to prove these claims about the subroutines.

Recall that ANT added $\text{Lap}(2/\epsilon)$ noise to the threshold and $\text{Lap}(4/\epsilon)$ noise to the query for overall ϵ-differential privacy. The subroutine in SV adds

$$\text{Lap}(\sigma(\epsilon_1)) = \text{Lap}\left(\frac{2c}{\epsilon_1}\right) = \text{Lap}\left(\frac{2c}{\frac{8}{9}\epsilon}\right) = \text{Lap}\left(\frac{2}{\frac{8}{9c}\epsilon}\right)$$

noise to the threshold and

$$\text{Lap}(2\sigma(\epsilon_1)) = \text{Lap}\left(\frac{4}{\frac{8}{9c}\epsilon}\right)$$

noise to the query. Therefore, each call to ANT is $(\epsilon', 0)$-differentially private for $\epsilon' = \frac{8}{9c}\epsilon$.

Recall that the Laplace Mechanism adds $\text{Lap}(\Delta f/\epsilon) = \text{Lap}(1/\epsilon)$ noise for our sensitivity-1 queries. The subroutine in SV adds

$$\text{Lap}(\sigma(\epsilon_2)) = \text{Lap}\left(\frac{2c}{\epsilon_2}\right) = \text{Lap}\left(\frac{2c}{\frac{2}{9}\epsilon}\right) = \text{Lap}\left(\frac{1}{\frac{1}{9c}\epsilon}\right)$$

noise. Therefore, each call to the Laplace Mechanism is $(\epsilon', 0)$-differentially private for $\epsilon' = \frac{1}{9c}\epsilon$.

Basic Composition gives that the overall privacy guarantee is:

$$c\left(\frac{8}{9c}\epsilon\right) + c\left(\frac{1}{9c}\epsilon\right) = \frac{8}{9}\epsilon + \frac{1}{9}\epsilon = \epsilon$$

so SV is $(\epsilon, 0)$-differentially private.

Case $\delta > 0$:

Now we address the case where $\delta > 0$. We will follow a similar structure, where we prove privacy guarantees of each subroutine, and then prove overall privacy through Advanced Composition this time.

Each run of ANT is $(\frac{8}{9\sqrt{8c\ln(2/\delta)}}\epsilon, 0)$-differentially private. Our subroutine adds

$$\text{Lap}(\sigma(\epsilon_1)) = \text{Lap}\left(\frac{\sqrt{32c\ln(2/\delta)}}{\epsilon_1}\right) = \text{Lap}\left(\frac{\sqrt{32c\ln(2/\delta)}}{\frac{8}{9}\epsilon}\right) = \text{Lap}\left(\frac{2}{\frac{8}{9\sqrt{8c\ln(2/\delta)}}\epsilon}\right)$$

noise to the threshold and

$$\text{Lap}(2\sigma(\epsilon_1)) = \text{Lap}\left(\frac{4}{\frac{8}{9\sqrt{8c\ln(2/\delta)}}\epsilon}\right)$$

noise to the answer. Therefore, each call to ANT is $(\epsilon', 0)$-differentially private for $\epsilon' = \frac{8}{9\sqrt{8c\ln(2/\delta)}}\epsilon$.

We have the following corollary of advanced composition (easy to check, but also can be found on p52 of [1]):

Corollary 6 (Advanced Composition)**.** *If $\mathcal{M} : \mathbb{N}^{|\mathcal{X}|} \to \mathcal{R}^k$ is a k-fold adaptive composition of $(\epsilon'/\sqrt{8k\ln(1/\delta')}, 0)$-differentially private mechanisms, then $\mathcal{M}$ is (ϵ', δ')-differentially private.*

Applying this result with $k = c, \epsilon' = \frac{8}{9}\epsilon, \delta' = \frac{\delta}{2}$, we see that these runs of ANT together are $(\frac{8}{9}\epsilon, \frac{\delta}{2})$-differentially private.

Each run of the Laplace Mechanism is $(\frac{1}{9\sqrt{8c\ln(2/\delta)}}\epsilon, 0)$-differentially private, and instantiating Corollary 6 with $\epsilon' = \frac{1}{9}\epsilon$ and $\delta' = \frac{\delta}{2}$ gives that these c runs of the Laplace Mechanism together are $(\frac{1}{9}\epsilon, \frac{\delta}{2})$-DP. Basic Composition of these two subroutines gives that SV is $(\frac{8}{9}\epsilon + \frac{1}{9}\epsilon, \frac{\delta}{2} + \frac{\delta}{2})$-differentially private, i.e., (ϵ, δ)-differentially private. □

2.2 SparseVector Accuracy

Before discussing the accuracy of SparseVector, let us recall the accuracy theorems for the two sub-routines used in the SparseVector algorithm: AboveNoisyThreshold and the Laplace Mechanism. Recall also that ANT outputs binary answers in $\{\bot, \top\}$, so the ANT accuracy corresponds to accurate comparison against a threshold.

Theorem 7 (ANT Accuracy)**.** *For any sequence of k sensitivity-1 queries $\{f_1, f_2, \dots f_k\}$ satisfying $|\{i < k : f_i(x) > T - \alpha\}| = 0$, then AboveNoisyThreshold is (α, β_{ANT})-accurate for*

$$\alpha = \frac{8(\log k + \log(2/\beta_{ANT}))}{\epsilon_{ANT}}.$$

Theorem 8 (Laplace Accuracy)**.** *Let $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}^k$ and let $\mathcal{M}_L(x, f, \epsilon_L)$ be the Laplace mechanism, then $\forall \beta_L \in [0, 1]$,*

$$\Pr\left[\|f(x) - M_L(x, f, \epsilon_L)\|_\infty \geq \log\left(\frac{k}{\beta_L}\right)\left(\frac{\Delta f}{\epsilon_L}\right)\right] \leq \beta.$$

To measure the accuracy of SparseVector, we have to modify our notion of accuracy from ANT to a setting where the algorithm also outputs numeric answers for some queries.

Definition 9 (Numeric Accuracy)**.** *A mechanism that outputs a stream of answers $\{a_i\} \in \mathbb{R} \cup \{\bot\}$ to a stream of k queries $\{f_i\}$ is (α, β)-accurate with respect to a threshold T if with probability at least $1 - \beta$, the mechanism does not halt before f_k, and*

$$\begin{aligned} &\forall a_i \in \mathbb{R} \quad |f_i(x) - a_i| < \alpha \textit{ and} \\ &\forall a_i = \bot, \quad f_i(x) \leq T + \alpha. \end{aligned}$$

The first condition above is the same additive accuracy notion that we use with numeric outputs, e.g., for the Laplace mechanism: the answer produced by the mechanism should be within an additive α of the true answer to the query on the database. The second condition comes from ANT accuracy, where we only output whether a query answer is above or below a threshold: if the mechanism produces $\bot$, then the true query value should not be more than α above the threshold.

Theorem 10 (Sparse Vector Accuracy)**.** *For any sequence of k sensitivity-1 queries $\{f_1, f_2, \dots f_k\}$ satisfying $|\{i{<}k \mid f_i(x) \geq T - \alpha\}| < c$, SparseVector is (α, β)-numeric accurate for*

$$\alpha = \begin{cases} \frac{9c\left(\log k + \log\left(\frac{4c}{\beta}\right)\right)}{\epsilon}, & \textit{if } \delta = 0 \\ \frac{9(\log k + \log(4c/\beta))\sqrt{8c\log(2/\delta)}}{\epsilon}, & \textit{if } \delta > 0 \end{cases}.$$

Note that the condition $|\{i{<}k \mid f_i(x) \geq T - \alpha\}| < c$ plays the same role as the analogous condition in ANT accuracy. Without this condition, the algorithm *should* provide numeric answers to more than c queries, which means that it *should* halt before the k-th query, which would violate the "no early halting" accuracy condition in Definition 9.

Proof. We will separately show that the two conditions required for accuracy are satisfied: (1) $f_i(x) \leq T + \alpha$ when $a_i = \perp$, and (2)$|f_i(x) - a_i| \leq \alpha$ when $a_i \in \mathbb{R}$.

Condition 1: $f_i(x) \leq T + \alpha$ when $a_i = \perp$.
If $\delta = 0$, instantiate Theorem 7 with $\beta_{ANT} = \frac{\beta}{2c}$ and $\epsilon_{ANT} = \frac{8}{9c}\epsilon$, to get that each of the c calls to ANT is $(\alpha, \frac{\beta}{2c})$-accurate for

$$\alpha_1 = \frac{9c\left(\log k + \log\left(\frac{4c}{\beta}\right)\right)}{\epsilon}.$$

If $\delta > 0$, instantiate Theorem 7 with $\beta_{ANT} = \beta/2c$ and $\epsilon_{ANT} = \frac{8}{9\sqrt{8c\log\left(\frac{2}{\delta}\right)}}\epsilon$, to get that each of the c runs of ANT is $(\alpha, \beta/2c)$-accurate for

$$\alpha_1 = \frac{9\left(\log k + \log\left(\frac{4c}{\beta}\right)\right)\sqrt{8c\log(2/\delta)}}{\epsilon}.$$

Condition 2: $|f_i(x) - a_i| \leq \alpha$ when $a_i \ \in \mathbb{R}$.
Instantiate Theorem 8 with $k = 1$, $\Delta f = 1$, $\beta_L = \beta/2c$. If $\delta = 0$, use $\epsilon_L = \frac{1}{9c}\epsilon$, and $\delta > 0$, then use $\epsilon_L = \frac{1}{9\sqrt{8c\log(2/\delta)}}\epsilon$. This ensures that $\Pr[|f_i(x) - a_i| \geq \alpha_2] < \frac{\beta}{2c}$ for each $a_i \in \mathbb{R}$, for

$$\alpha_2 = \begin{cases} 9c\log\left(\frac{2c}{\beta}\right), if \delta = 0 \\ 9\log\left(\frac{2c}{\beta}\right)\sqrt{8c\log(2/\delta)}\epsilon, \quad if \delta > 0 \end{cases}.$$

Note that $\alpha := \alpha_1 \geq \alpha_2$, for any value of δ_1. This can be seen by direct comparison on the two terms. Thus each of the two c-subroutines is $(\alpha, \frac{\beta}{2c})$-accurate. Taking a union bound over all $2c$ failure probabilities gives that SparseVector is (α, β)-accurate. □

References

[1] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, 2014.