ISYE/CS 8803: Foundations of Data Privacy **Fall 2017**

Lecture 6: Sparse Vector

Lecturer: Rachel Cummings *Scribe: Kevin Lai September 13, 2017*

1 Review of AboveNoisyThreshold

Last time, we covered the AboveNoisyThreshold (ANT) algorithm, which took in a database x, a stream of sensitivity 1 queries, a threshold T, and privacy parameter ϵ. It would report back whether the noisy answer to each query was above or below a noisy version of the threshold, and it halted the first time it found an above-threshold query.

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

AboveNoisyThreshold$(x, \{f\}, T, \epsilon)$:
Input: Database x, sensitivity 1 queries $\{f\}$, threshold T, privacy parameter ϵ
Output: $\{\perp, \top\}$
Let $\hat{T} = T + \text{Lap}(2/\epsilon)$
for each query f_i **do**
  Let $v_i = \text{Lap}(4/\epsilon)$
  if $f_i(x) + v_i \geq \hat{T}$ **then**
    Output $a_i = \top$
    Halt
  else
    Output $a_i = \perp$
  end if
end for

We previously showed that ANT is $(\epsilon, 0)$-differentially private and (α, β)-accurate, meaning that with probability at least $1 - \beta$, all $\{\perp, \top\}$ outputs were correct up to an additive α.

Today, we'll extend this to the Sparse Vector Mechanism, which can handle multiple above-threshold queries without halting, and it can produce numerical answers to those queries.

2 SparseVector Mechanism

The SparseVector mechanism was developed in [DNR+09] and refined in [HR10]. It 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 $\{a_i\} \in (\mathbb{R} \cup \{\perp\})^*$.

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 $\sigma(\epsilon) = \frac{32c\ln(2/\delta)}{\epsilon}$
end if
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
    if count $\geq c$ **then**
        Halt
    end if
end for

Now let's 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 $(\epsilon, 0)$-DP or (ϵ, δ)-DP, we'll have to set parameters within these subroutines differently. If we want $(\epsilon, 0)$-DP, 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)-DP, then we can use Advanced Composition, and we'll set our parameters according to the Corollary that we saw last time.

Theorem 1. *[DNR+09] SparseVector is (ϵ, δ)-DP.*

[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.

Proof. <u>Case $\delta = 0$:</u>
We first consider the case where $\delta = 0$. SV consists of c runs of ANT, where each run is $(\frac{8}{9c}\epsilon, 0)$-DP and c runs of the Laplace Mechanism, where each run is $(\frac{1}{9c}\epsilon, 0)$-DP.

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

$$\text{Lap}(\sigma(\epsilon_1)) = \text{Lap}\left(\frac{2c}{\epsilon_1}\right) = \text{Lap}\left(\frac{2c}{\frac{8}{9}\epsilon}\right) = \text{Lap}\left(\frac{2}{\frac{8}{9c}\epsilon}\right) \tag{1}$$

noise to the threshold and

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

noise to the query. Therefore, each call to ANT is $(\epsilon', 0)$-DP 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

$$\text{Lap}(\sigma(\epsilon_2)) = \text{Lap}\left(\frac{2c}{\epsilon_2}\right) = \text{Lap}\left(\frac{2c}{\frac{2}{9}\epsilon}\right) = \text{Lap}\left(\frac{1}{\frac{1}{9c}\epsilon}\right) \tag{3}$$

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

Basic Composition gives that the overall privacy guarantee is:

$$c\left(\frac{8}{9c}\epsilon\right) + c\left(\frac{1}{9c}\epsilon\right) = \frac{8}{9}\epsilon + \frac{1}{9}\epsilon = \epsilon \tag{4}$$

so SV is $(\epsilon, 0)$-DP.
<u>Case $\delta > 0$:</u>
Now we address the case where $\delta > 0$. Each run of ANT is $(\frac{8}{9\sqrt{8c\ln(2/\delta)}}\epsilon, 0)$-DP. Our subroutine adds

$$\text{Lap}(\sigma(\epsilon_1)) = \text{Lap}\left(\frac{\sqrt{32c\ln(2/\delta)}}{\epsilon_1}\right) = \text{Lap}\left(\frac{\sqrt{32c\ln(2/\delta)}}{\frac{8}{9}\epsilon}\right) = \text{Lap}\left(\frac{2}{\frac{8}{9\sqrt{8c\ln(2/\delta)}}\epsilon}\right) \tag{5}$$

noise to the threshold and

$$\text{Lap}(2\sigma(\epsilon_1)) = \text{Lap}\left(\frac{4}{\frac{8}{9\sqrt{8c\ln(2/\delta)}}\epsilon}\right) \tag{6}$$

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

Recall the corollary of advanced composition from last time.

Corollary 2 (Advanced Composition)**.** *If $\mathcal{M} : \mathbb{N}^{|\chi|} \to \mathcal{R}^k$ is a k-fold adaptive composition of $(\epsilon'/\sqrt{8k\ln(1/\delta')}, 0)$-DP mechanisms, then $\mathcal{M}$ is (ϵ', δ')-DP.*

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

Each run of the Laplace Mechanism is $(\frac{1}{9\sqrt{8c\ln(2/\delta)}}\epsilon, 0)$-DP, and instantiating the corollary 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 $(\frac{8}{9}\epsilon + \frac{1}{9}\epsilon, \frac{\delta}{2} + \frac{\delta}{2})$-DP i.e. (ϵ, δ)-DP. $\square$

References

[DNR+09] Cynthia Dwork, Moni Naor, Omer Reingold, Guy N. Rothblum, and Salil Vadhan. On the complexity of differentially private data release: Efficient algorithms and hardness results. In *Proceedings of the Forty-first Annual ACM Symposium on Theory of Computing*, STOC '09, pages 381–390, New York, NY, USA, 2009. ACM.

[HR10] Moritz Hardt and Guy N. Rothblum. A multiplicative weights mechanism for privacy-preserving data analysis. In *Proceedings of the 2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, FOCS '10, pages 61–70, Washington, DC, USA, 2010. IEEE Computer Society.