Harvard Privacy Tools Project

Mean Sensitivity Proofs

June 22, 2021

Definition 1. *The sample mean of database X of size n is defined as*

$$f(X) = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

These are restricted-sensitivity proofs that only apply when N is known. The library makes use of the Resize component to guarantee this static property. If N is unknown, there is an argument on the DPMean component to estimate the mean by postprocessing plug-in estimates for the count and sum.

1 Neighboring Definition: Change One

1.1 ℓ_1-sensitivity

Theorem 1. *Say the space of datapoints $\mathcal{X}$ is bounded above by M and bounded below by m. Then $f(\cdot)$ has ℓ_1-sensitivity in the change-one model bounded above by*

$$\frac{M-m}{n}.$$

Proof. Say X and X' are neighboring databases which differ at data-point x_j, and let Δf indicate the ℓ_1-sensitivity of $f(\cdot)$. Then

$$\begin{aligned}
\Delta f &= \max_{X,X'} \left| f(X) - f(X)' \right| \\
&= \max_{X,X'} \frac{1}{n} \left| \left(\sum_{\{i \in [n] \mid i \neq j\}} x_i \right) + x_j - \left(\sum_{\{i \in [n] \mid i \neq j\}} x'_i \right) + x'_j \right| \\
&= \max_{X,X'} \frac{1}{n} \left| x_j - x'_j \right| \\
&\leq \frac{M-m}{n}.
\end{aligned}$$

□

1.2 ℓ_2-sensitivity

Theorem 2. *Say the space of datapoints $\mathcal{X}$ is bounded above by M and bounded below by m. Then $f(\cdot)$ has ℓ_2-sensitivity in the change-one model bounded above by*

$$\frac{M-m}{n}.$$

Proof. This follows the same logic as the above proof. □

2 Neighboring Definition: Add/Drop One

2.1 ℓ_1-sensitivity

Theorem 3. *Say the space of datapoints $\mathcal{X}$ is bounded above by M and bounded below by m. Then $f(\cdot)$ has ℓ_1-sensitivity in the add/drop-one model bounded above by*

$$\frac{M-m}{n}.$$

Proof. For notational ease, let n always refer to the size of database x. We must consider both adding and removing an element from x. First, consider adding a point:

Let $X' = X \cup \{x\}$. Without loss of generality, assume the point added is the $(n+1)^{\text{th}}$ element of database X'. Note that

$$\begin{aligned}
\left|f(X) - f(X)'\right| &= \left|\frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n+1}\sum_{i=1}^{n+1} x_i\right| \\
&= \left|\left(\frac{1}{n} - \frac{1}{n+1}\right)\sum_{i=1}^{n} x_i - \frac{x}{n+1}\right| \\
&= \frac{1}{n+1}\left|\frac{1}{n}\sum_{i=1}^{n} x_i - x\right| \\
&\leq \frac{|M-m|}{n+1}.
\end{aligned}$$

Second, consider removing a point:
Let $X' = X \backslash \{x\}$. Without loss of generality assume that the point subtracted is the n^{th} element of database X.

$$\begin{aligned}
\left|f(X) - f(X')\right| &= \left|\frac{1}{n-1}\sum_{i=1}^{n-1} x_i - \frac{1}{n}\sum_{i=1}^{n} x_i\right| \\
&= \left|\left(\frac{1}{n-1} - \frac{1}{n}\right)\sum_{i=1}^{n-1} x_i - \frac{x}{n}\right| \\
&= \frac{1}{n}\left|\frac{1}{n-1}\sum_{i=1}^{n-1} x_i - x\right| \\
&\leq \frac{|M-m|}{n}.
\end{aligned}$$

Then, since $\forall n > 0$,

$$\frac{1}{n+1} < \frac{1}{n},$$

the sensitivity of the mean in general is bound from above by

$$\frac{M-m}{n}.$$

$\square$

2.2 ℓ_2-sensitivity

Theorem 4. *Say the space of datapoints $\mathcal{X}$ is bounded above by M and bounded below by m. Then f has ℓ_2-sensitivity in the add/drop-one model bounded above by*

$$\frac{M-m}{n}.$$

Proof. This follows the same logic as the above proof. $\square$