Count Sensitivity Proofs

March 17, 2020

Definition 1. *Let $\mathcal{X}$ be the universe of possible rows (individuals) and let $I : \mathcal{X} \to \{0, 1\}$ be a predicate on rows. Let $x \in \mathcal{X}^n$ be a dataset. Then a count over x is defined as*

$$q(x) = \sum_{i=1}^{n} I(x_i).$$

Definition 2. *Let $q_1, \dots, q_k$ be a series of counts with predicates $I_1, \dots, I_k$. These counts are disjoint for every row in the database, only one of them evaluates to 1. In other words, they are disjoint if $\forall x_i \in \mathcal{X}$,*

$$\sum_{j=1}^{k} I_j(x_i) = 1.$$

1 Neighboring Definition: Change One

1.1 ℓ_1-sensitivity

Theorem 1. *A single count query has sensitivity 1. A series of k disjoint counts has sensitivity 2.*

Proof. Let q be a count query with predicate I, and let x' be equal to x with point x_i changed to x_i'. Then $I(x_i)$ will change by at most 1, and since no other x_j with $i \neq j$ is changed, this term in q is the only thing affected. Thus, the sensitivity of a single query is bounded by 1.
Consider k disjoint count on the same databases x and x'. Since they are disjoint, only one of the k counts is influenced by x_i, and only one of the counts is affected by x_i'. Since each query can be affected by at most 1 by a single data point by the above logic, in total there will be a change of at most 2. □

1.2 ℓ_2-sensitivity

Theorem 2. *A single count query has sensitivity 1. A series of k disjoint counts has sensitivity 4.*

Proof. From the proof of Theorem 1, the difference between counts on two neighboring databases is at most 1. Squaring this gives the same value. Similarly, the difference between the results of k disjoint counts on two neighboring databases is 2, and squaring this gives 4. $\square$

2 Neighboring Definition: Add/Drop One

2.1 ℓ_1-sensitivity

Theorem 3. *A single count query has sensitivity 1. A series of k disjoint counts also has sensitivity 1.*

Proof. Let q be a count query with predicate I, and let x' be equal to database x with point x_i removed. Then

$$\begin{aligned} \left|q(x) - q(x')\right| &= \left|\sum_{j=1}^{n} I(x_j) - \sum_{j\neq i}^{n} I(x_j)\right| \\ &= \left|\sum_{j\neq i} I(x_j) + I(x_i) - \sum_{j\neq i} I(x_j)\right| \\ &\leq 1. \end{aligned}$$

A nearly identical argument holds for adding a point. Consider a series of k disjoint count queries $\mathbf{q} = \{q_1, \ldots, q_k\}$ and consider database x' equal to database x with point x_i removed. Note that only a single one of the k queries will be affected by the change from x to x', so

$$\begin{aligned} \left|\mathbf{q}(x) - \mathbf{q}(x')\right| &= \left|\left(q_1(x) - q_1(x')\right) + \ldots + \left(q_k(x) - q_k(x')\right)\right| \\ &\leq 1. \end{aligned}$$

The same argument holds for x' equal to x with a single point added. $\square$

2.2 ℓ_2-sensitivity

Theorem 4. *A single count query has sensitivity 1. A series of k disjoint counts also has sensitivity 1.*

Proof. Squaring the sensitivity bounds from Theorem 3 gives 1 as an upper bound on the ℓ_2 sensitivity for both a single count and a series of k disjoint counts. $\square$