An Additional Definition

Jie Zhang, Yifan Dong, Li Yin, Zhiwu Li

Differential privacy is a privacy-preserving method with a rigorous mathematical definition, which offers a mechanism (or a function) that publishes aggregate information about a statistical database, where the private information in it is protected or restricted. In other words, if a dataset is considered as an input of a differential privacy mechanism, the addition or deletion of any one record (or element) in it does not affect the query result, i.e., an intruder cannot capture the private information with the slight modification of the dataset.

Definition 1 (Differential Privacy) Let ϵ be a positive real number and $\mathcal{K}_o$ be a randomized mechanism (function) that takes a dataset as input. Let $Im(\mathcal{K}_o)$ denote the image of $\mathcal{K}_o$. The mechanism $\mathcal{K}_o$ provides ϵ-differential privacy if for any two datasets $\mathcal{O}$ and $\mathcal{O}'$ that differ on a single element, for all $\mathcal{S} \subseteq Im(\mathcal{K}_o)$, it holds

$$P_r\left[\mathcal{K}_o(\mathcal{O}) \in \mathcal{S}\right] \leq \exp(\epsilon) P_r[\mathcal{K}_o(\mathcal{O}') \in \mathcal{S}],$$

where the value of $\mathcal{K}_o$ at a dataset $\mathcal{O}$ or $\mathcal{O}'$ is contained in the sample space, i.e., $Im(\mathcal{K}_o)$, with a probability decided by the randomness used in the mechanism, and $P_r[\mathcal{K}_o(\mathcal{O}) \in \mathcal{S}]$ is the probability of $\mathcal{K}_o(\mathcal{O}) \in \mathcal{S}$ representing that the output of $\mathcal{K}_o$ at $\mathcal{O}$ belongs to $\mathcal{S}$. $\diamondsuit$

In Definition 1, the value ϵ evaluates the performance of differential privacy. Namely, a smaller value of ϵ implies a finer difference between the probabilities of $\mathcal{K}_o(\mathcal{O}) \in \mathcal{S}$ and $\mathcal{K}_o(\mathcal{O}') \in \mathcal{S}$, i.e., the intruder is less likely to distinguish the two datasets. On the contrary, a larger ϵ means a lower degree of users' private information protection.