Lecture 1

1-20-26

first version of definition:

Def An algorithm $A: \mathcal{X}^n \to \mathcal{Y}$ is (ε, δ)-DP if $\forall x, x' \in \mathcal{X}^n$ such that x & x' differ in one entry and events $E \subseteq \mathcal{Y}$,

$$\Pr[A(x) \in E] \leq e^{\varepsilon} \Pr[A(x') \in E] + \delta.$$

Here: $\mathcal{X}$ = space of data points

$\mathcal{X}^n$ = datasets of n examples

$\mathcal{Y}$ = output space for algorithm

$\varepsilon, \delta \geq 0$

Another way to state definition.

Def (Adjacency) Two datasets $x=(x_1,\dots,x_n)$ and $x'=(x_1',\dots,x_n')$ are <u>adjacent</u>, written $x\sim x'$, if $\exists\, i^* \in [n]$ such that $\forall i \neq i^*$, $x_i = x_i'$.

Def ((ε,δ)-indistinguishability) Two distributions p, q are <u>(ε,δ)-indistinguishable</u>, written $p \approx_{(\varepsilon,\delta)} q$, if $\forall$ events E

$$p(E) \le e^{\varepsilon} q(E) + \delta$$
$$q(E) \le e^{\varepsilon} p(E) + \delta.$$

Def Algorithm A is <u>(ε,δ)-DP</u> if $\forall x \sim x'$,

$$A(x) \approx_{(\varepsilon,\delta)} A(x').$$

With different notions of adjacency and/or different notions of "closeness of distributions", get different flavors of DP.

$\varepsilon, \delta \Rightarrow$ approx

$\delta = 0 \Rightarrow$ pure

DP is worst-case over datasets.