1 Задание 1

Пусть

$$X = \{x_i\}_{i=1}^n, x_i \in 0; 1, \quad f(X) = \frac{1}{n}\sum_{i=1}^{n} x_i, A(X) = f(X) + \xi, \text{ где } \xi \sim \text{Lap}(\alpha).$$

Как следует выбрать α, чтобы обеспечить ε-дифференциальную приватность алгоритма A?

Плотность распределения Лапласа имеет вид:

$$f_\xi(x) = \frac{1}{2\alpha}\exp\left(-\frac{|x|}{\alpha}\right).$$

Найдем чувствительность функции $f(x)$:

$$|f(x) - f(x')| = \left|\frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n}\sum_{i=1}^{n} x_i'\right| = \frac{1}{n}|x_j - x_j'| \leqslant \frac{1}{n} = \Delta,$$

Теорема

Пусть $\xi = (z_1, \cdots, z_n)^T$, $z_j \sim \text{Lap}\left(\frac{\Delta}{\varepsilon}\right)$, $\Delta = \sup_{D,D'} \|f(D) - f(D')\|$. Тогда алгоритм $A(X) = f(X)+\xi$, где ξ - вектор из d независимых случайных величин, распределенных по Лапласу с параметром $\alpha = \Delta f/\varepsilon$, обеспечивает ε-дифференциальную приватность.

Таким образом, чтобы обеспечить ε-дифференциальную приватность алгоритма A, необходимо выбрать $\alpha = \frac{1}{n\varepsilon}$.

Можно ли подобрать α, если $\xi \sim U[-\alpha, \alpha]$?

Определение

Алгоритм A обеспечивает ε-дифференциальную приватность, если для любых соседних наборов D и D' и для любого множества выходных значений $E \subseteq Y$ выполняется следующее неравенство:

$$P(A(D) \in E) \leqslant e^\varepsilon P(A(D') \in E).$$

Рассмотрим два соседних набора данных X и X', которые отличаются только в одном элементе. Пусть X и X' отличаются в j-ом элементе, то есть $x_j \neq x_j'$. Тогда:

$$A(X) = f(X) + \xi$$
$$A(X') = f(X') + \xi$$

Следовательно

$$A(x') = f(x) - \Delta + \xi.$$

Рассмотрим множество значений $A(x)$ и $A(x')$:

$$A(x) \in [f(x) - \alpha, f(x) + \alpha],$$
$$A(x') \in [f(x) - \Delta - \alpha, f(x) - \Delta + \alpha].$$

Рассмотрим точку $X_0 = f(x) + \alpha - \frac{\Delta}{2}$, тогда:

$$X_0 \in [f(x) - \alpha, f(x) + \alpha], \text{ при } \frac{\Delta}{2} < \alpha,$$
$$X_0 \notin [f(x) - \Delta - \alpha, f(x) - \Delta + \alpha].$$

Следовательно

$$P(A(X) = X_0) = \frac{1}{2\alpha},$$
$$P(A(X') = X_0) = 0.$$

Таким образом, не существует такого α, чтобы выполнялось определение дифференциальной приватности.

2 Задание 2

Пусть датасеты $D = D_0$ и $D' = D_k$ отличаются на k элементов, а датасеты D_i и D_{i+1} отличаются на 1 элемент для всех $i \in \{0, \ldots, k-1\}$. Тогда, если алгоритм A обеспечивает ε-дифференциальную приватность, то для любых датасетов D_i и D_{i+1} выполняется следующее неравенство:

$$P(A(D_i) \in E) \leqslant e^{\varepsilon} P(A(D_{i+1}) \in E).$$

Последовательно применяя это неравенство для всех пар соседних датасетов от D_0 до D_k, получаем:

$$P(A(D_0) \in E) \leqslant e^{\varepsilon} P(A(D_1) \in E) \leqslant e^{2\varepsilon} P(A(D_2) \in E) \leqslant \ldots \leqslant e^{k\varepsilon} P(A(D_k) \in E).$$

Таким образом, для любых датасетов D и D', отличающихся на k элементов, выполняется следующее неравенство:

$$P(A(D) \in E) \leqslant e^{k\varepsilon} P(A(D') \in E).$$