数据隐私加密方法与伦理实现 HW2

王嵘晟　PB1711614

1.

1.1

使用 Laplace 差分隐私，由于查询针对的是数据库中数据的前 n 项和，将前 n 项和与前 n-1 项和做减法就可以得到第 n 个数据的值。选择 $(\epsilon, 0)$ 差分隐私。

$$M_L(x, f(\cdot), \epsilon) = f(x) + (Y_1 + ... + Y_k)$$

所以对于给定的 $\epsilon = 0.1$，查询的 $\Delta f = 5$，所以 $Y_i\ iid\ Lap(50) = \frac{1}{100}exp\{-\frac{|x|}{50}\}$ 所以：

$$M_L(x, f(\cdot), \epsilon) = \sum_{i=1}^{5} x_i + (Y_1 + ... + Y_5)$$

其中 $Y_i\ iid\ \frac{1}{100}exp\{-\frac{|x|}{50}\}$

1.2

同样使用 $(\epsilon, 0)$ 差分隐私，用 Laplace 加噪音，由于 $\epsilon = 0.1$，$\Delta f = 2$ 所以 $Y_i\ iid\ Lap(20) = \frac{1}{40}exp\{-\frac{|x|}{20}\}$。
所以

$$M_L(x, f(\cdot), \epsilon) = max_{i\in[1,5]}x_i + (Y_1 + ... + Y_5)$$

其中 $Y_i\ iid\ \frac{1}{40}exp\{-\frac{|x|}{20}\}$

2.

2.1

根据给定的条件：

$$\Delta_t \le \begin{cases} \Delta_{t-1} + 2L\eta_t, & if\ t = 1 + jm,\ j \in \mathbb{N} \\ \Delta_{t-1}, & otherwise, \end{cases} \tag{1}$$

可得 $\Delta_1 \le 2L\eta_1$ 当 $1 \le t < 1 + m$ 时，$\Delta_t \le 2L\eta_1$，$\Delta_{1+m} \le 2L(\eta_1 + \eta_{1+m})$ 以此类推可得，当 T=km 时：

$$\Delta_T \le Delta_{1+(k-1)m} \le 2L \sum_{j=0}^{k-1} \eta_{1+jm}$$

2.2

根据给定的条件：

$$\Delta_t \le \begin{cases} (1 - n\gamma)\Delta_{t-1} + 2L\eta, & if\ t = 1 + jm,\ j \in \mathbb{N} \\ (1 - n\gamma)\Delta_{t-1}, & otherwise, \end{cases} \tag{2}$$

可得 $\Delta_1 \le 2L\eta$，当 $1 \le t < 1 + m$ 时，$\Delta_t \le (1 - n\gamma)^{t-1} 2L\eta$。以此类推，可得当 $T = km$ 时：

$$\Delta_T \le (1 - n\gamma) Delta_{T-1} \le 2L\eta \sum_{j=0}^{k-1} (1 - n\gamma)^{(k-j)m-1}$$

3.

3.1

根据定理：

Theorem A.1. Let $\varepsilon \in (0, 1)$ be arbitrary. For $c^2 > 2\ln(1.25/\delta)$, the Gaussian Mechanism with parameter $\sigma \ge c\Delta_2 f/\varepsilon$ is (ε, δ)-differentially private.

SGD 算法的更新操作，每个更新的数值都在 [0,1] 之间，令 $c = \delta\epsilon$，可得每次更新都是 $(\epsilon, \delta) - DP$

3.2

根据 composition theorem，由于 T=10000, 所以对于每个 DP 来说，$(\epsilon, \delta) = (1.25\times10^{-4}, 10^{-9})$。
所以 $\sigma \geq \sqrt{2ln(1.25 \times 10^9)/(1.25 \times 10^{-4})^2} = 51779.73$

3.3

根据 3.20，k=T=10000，$k\delta + \delta' = 0.10001$，$\epsilon' = \sqrt{2kln(\frac{1}{\delta'})\epsilon} + k\epsilon(e^\epsilon - 1) = 31369.21$ 所以
$\delta \geq \sqrt{2ln(1.25 \times 10^5)/31369.21^2} = 50311.17$

4.

$$t_i^* = \begin{cases} t_i & w.p.\ p = \frac{e^\epsilon}{1+e^\epsilon} \\ \overline{t_i} & w.p.\ 1-p = \frac{1}{1+e^\epsilon} \end{cases}$$

所以 $\sum_{x\in\hat{X}} x$ 的期望为 $\frac{1}{1+e^\epsilon}(1-p) = \frac{1}{2+2e^{0.2}}$，$\sum_{x\in X} x$ 的期望为 $\frac{1}{1+e^\epsilon}p = \frac{1}{2+2e^{0.2}}$
$p = 0.1$ 时 $\sum_{x\in\hat{X}} x$ 的期望为 $\frac{1}{1+e^\epsilon}(1-p) = 0.9\frac{1}{1+e^{0.2}}$，$\sum_{x\in X} x$ 的期望为 $\frac{1}{1+e^\epsilon}p = 0.1\frac{1}{1+1e^{0.2}}$
$p = 0.9$ 时 $\sum_{x\in\hat{X}} x$ 的期望为 $\frac{1}{1+e^\epsilon}(1-p) = 0.1\frac{1}{1+e^{0.2}}$，$\sum_{x\in X} x$ 的期望为 $\frac{1}{1+e^\epsilon}p = 0.9\frac{1}{1+1e^{0.2}}$
观察得：连个期望也符合参数为 p 的伯努利分布

5.

5.1