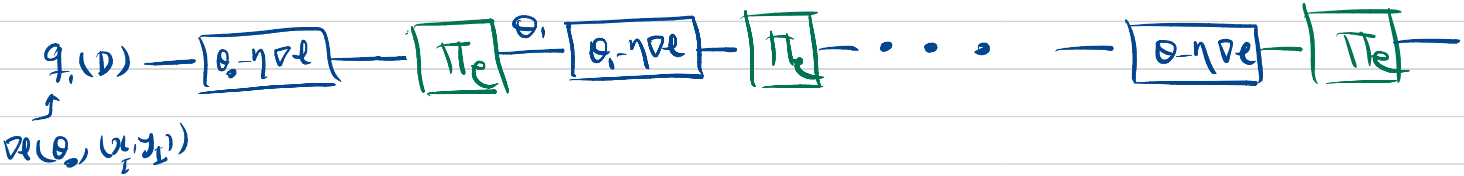
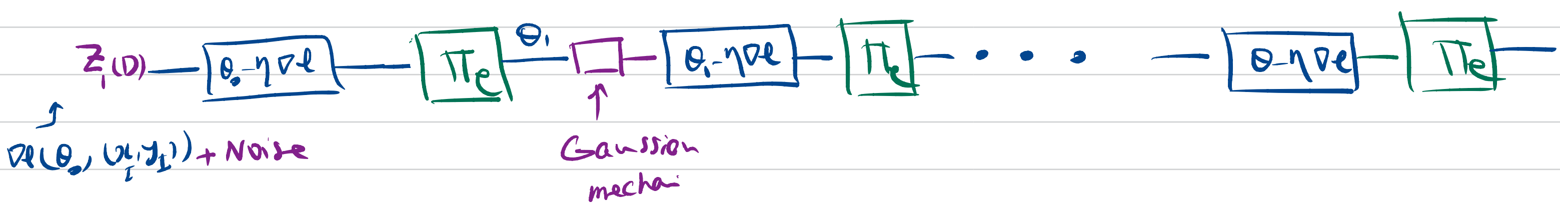
noisy

Private Projected SGD (PN-SGD)



To make projected SGD, differentially private, we need to make each iteration private. To do so, we need to pass the query response through a DP mechanism, say Gaussian mechanism.



Thus, each iteration proceeds as follows:

Private Projected GD

Input: Dataset $D = \{(x_i, y_i)\}_{i=1}^{n}$, loss function ℓ, parameter set $\mathcal{C}$ & η_t

1. Pick $\theta_0 \in \mathcal{C}$ arbitrarily
2. For $t=1$ to T
   - Select $I \in \{1, 2, \ldots, n\}$ uniformly at random
   - Set $\theta_t = \prod_{\mathcal{C}} \left(\theta_{t-1} - \eta_t \left[n \nabla \ell(\theta_{t-1}, (x_I, y_I)) + (N_{t-1}^{1}, \ldots, N_{t-1}^{d}) \right] \right)$

$N_t^i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$

Output $\theta_1, \theta_2, \ldots \theta_T$

At each iteration, query is $\nabla \ell(\theta, (x_I, y_I))$ which is an adaptive query as it depends on the previous iteration's output.

Each iteration becomes $(\varepsilon_0, \delta_0)$ - DP with $\varepsilon_0 = \frac{2nL}{\sigma} \sqrt{2 \log \frac{1}{\delta_0}}$

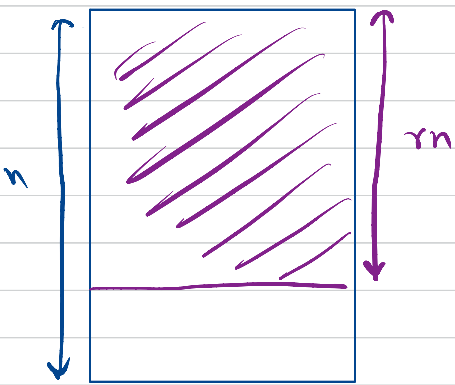
Note that : $\Delta_2^q \leq 2nL$ for the query $\nabla \ell(\theta, (x_I, y_I))$.

Therefore, advanced composition can be used to obtain the privacy guarantee. But, we can do better! At each iteration, we don't use the whole dataset! We only use One data record & not all the records. Thus, the privacy must be significantly better!

How to quantify this improvement:

* Privacy amplification by Sub_Sampling:

Let M be an (ε,δ)-DP mechanism. If it is run on a uniformly selected subset of dataset of size γn $(\gamma \leq 1)$, then it will provide $\left(\log(1+\gamma(e^{\varepsilon}-1)),\ \gamma\delta\right)$-DP.



$$q(D) \longrightarrow \boxed{M} \longrightarrow Z_D$$

$q(D)$ depends only the first part dataset

Z_D provides more privacy compared to the case when $q(D)$ depends on the whole dataset.

In the SGD: $\gamma = n$ (as the size of dataset $=1$)

Gaussian mechanism coupled with this sub_sampling is typically called: Sub_Sampled Gaussian mechanism.

Remark. we simplify the privacy amplification by sub sampling as follows:

$$\left(\underbrace{\log(1+\gamma(e^{\varepsilon}-1))}_{\lesssim 2\gamma\varepsilon},\ \gamma\delta\right)\text{-DP} \Rightarrow (2\gamma\varepsilon, \gamma\delta)\text{-DP}$$

There are tighter analysis for "privacy amplification" caused by sub_sampling.

To see the original proof (in terms of HS divergence), see "privacy amplification by sub-sampling: Tight analysis via Coupling" by Balle, Barthe, and Gaboardi, ICML 2018

Privacy analysis of PN-SGD:

Each iteration is an instance of sub-sampled Gaussian mechanism. To derive its privacy guarantee, we need to find the privacy guarantee of the corresponding Gaussian mechanism & then apply privacy amplification by sub-sampling.

ℓ_2-sensitivity

- Gaussian mechanism with noise $N(0, \sigma^2 I)$ is $\left(\frac{2nL}{\sigma} \sqrt{2 \log 1/\delta_0}, \delta_0 \right)$-DP because the query

  $n \nabla \ell(\theta, (x_i, y_i))$ has ℓ_2-sensitivity $= 2nL$ with L as Lipschitz constant.

- Applying "privacy amplification by subsampling" theorem, for $\gamma = \frac{1}{n}$ each iteration becomes $\left(\frac{2}{n} \frac{2nL}{\sigma} \sqrt{2 \log 1/\delta_0}, \frac{1}{n} \delta_0 \right)$ -DP.

•) Now we can advanced composition:

PN-SGD after $T = n^2$ number of iterations is

$$\left(\frac{4L}{\sigma} \sqrt{2 \log 1/\delta_0} \sqrt{2T \log 1/\delta'}, \frac{T}{n} \delta_0 + \delta' \right) \text{-DP} \quad \forall \delta' \in (0,1)$$

choosing $\delta' = n \delta_0$, we obtain

$$\frac{T}{n} \delta_0 + n\delta_0 \overset{T=n^2}{=} 2n\delta_0 =: \delta$$

so $\delta' = \delta/2$

& $\delta_0 = \delta/2n$

thus, the privacy guarantee becomes

$$\left(\frac{4L}{\sigma} n \sqrt{2 \log \frac{2n}{\delta}} \sqrt{2 \log 2/\delta} \,,\ \delta \right) - DP$$

So to have it be (ε, δ)-DP, we require

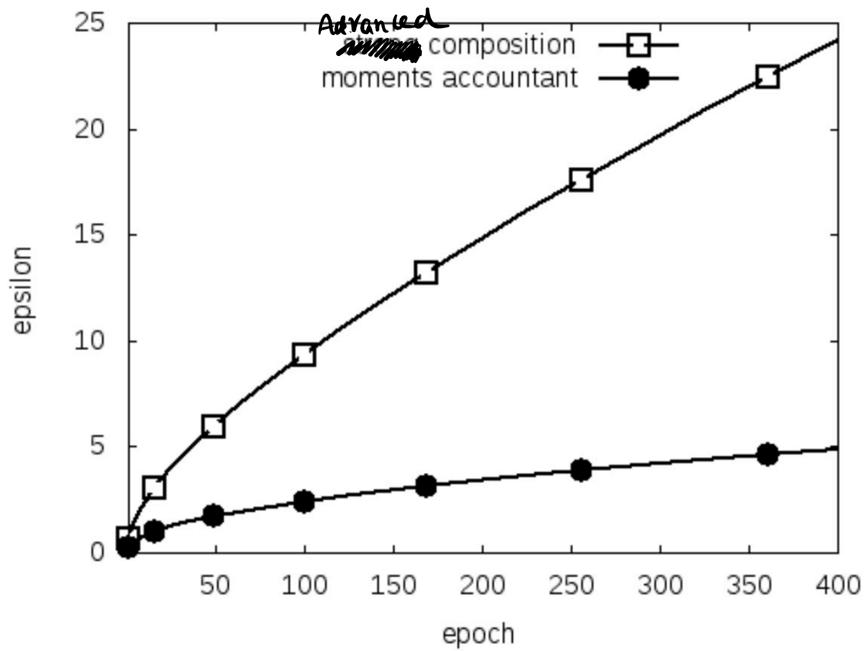
$$\boxed{\sigma^2 = \frac{64 n^2 L^2}{\varepsilon^2} \log \frac{2n}{\delta} \log 2/\delta}$$

Remarks:

1. A much better approach to study privacy guarantee of PN-SGD is to invoke Rényi DP, use linear composition of Rényi DP, & then convert back to (ε, δ)-DP. A technique known as

moments accountant. The difference is astonishing:

the resulting σ is $\sqrt{\frac{\log n}{\delta}}$ smaller!



[Taken from "Deep learning with differential Privacy"]

2- Our privacy analysis relies on the assumption that the loss function is Lipschitz. This is a restrictive assumption for deep learning. For example: MSE, squared hinge loss, or

To make sure sensitivity of $\nabla \ell$ is bounded even for non-Lipschitz loss, we usually <u>clip</u> the gradient:

$$\mathrm{clip}(\nabla \ell(\theta,(x_i,y_i)), C) := \begin{cases} C \dfrac{\nabla \ell}{\|\nabla \ell\|} & \text{if } \|\nabla \ell\|_2 \geq C \\ \nabla \ell & \text{if } \|\nabla \ell\|_2 < C \end{cases}$$

Thus, $\|\mathrm{clip}(\nabla \ell, C)\| \leq C \Rightarrow$ sensitivity of $\mathrm{clip}(\nabla \ell, C) = 2C$

Accuracy of PN-SGD:

Recall the folklore result: Given Oracle $G_t(\theta)$ with $\mathbb{E}[G_t(\theta)] = \nabla \mathcal{L}(\theta_t, D)$ & $\mathbb{E}[\| G_t(\theta)\|^2] \le G^2$, we have for

$$\theta_{t+1} = \prod_{\mathcal{C}} (\theta_t - \eta\, G_t(\theta_t)) : \qquad \mathbb{E}[\mathcal{L}(\theta_T, D)] - \mathcal{L}(\theta^*; D) \le \frac{\|\mathcal{C}\|_2 G \log T}{\sqrt{T}}.$$

Here, $$G_t(\theta_t) := n \nabla \ell(\theta_t, (x_I, y_I)) + \underbrace{(N_t^1, \ldots, N_t^d)}_{N_t}$$

Note that $$n\, \mathbb{E}[G_t(\theta_t)] = \nabla \mathcal{L}(\theta_t, D)$$

&
$$\begin{aligned} \mathbb{E}[\| G_t(\theta)\|^2] &= \mathbb{E}[\| n \nabla \ell(\theta, (x_I, y_I)) + N_t \|^2] \\ &= n^2 \mathbb{E}[\| \nabla \ell(\theta, (x_I, y_I))\|^2] + \underbrace{\mathbb{E}[n \langle \nabla \ell(\theta, (x_I, y_I)), N_t \rangle]}_{=0} \\ &\quad + \mathbb{E}[\|N_t\|^2] \end{aligned}$$

$$\leq n^2 L^2 + d\sigma^2$$

so $\quad G := \sqrt{n^2 L^2 + d\sigma^2}$

thus, PN_SGD satisfies:

$$\mathbb{E}\left[\mathcal{L}(\theta_T, D)\right] - \mathcal{L}(\theta^*, D) \leq \|e\|_2 \sqrt{n^2 L^2 + d\sigma^2}\ \frac{\log T}{\sqrt{T}}$$

replace σ & $T = n^2$

$$\lesssim O\left(\frac{\|e\|_2 L \sqrt{d \log 1/\delta}}{\varepsilon}\right)$$

Not decreasing in T

Local DP

All Privacy problems discussed so far were based on an assumption:

A private dataset is held by a <u>trusted</u> central entity (such as hospital or bank).

What if there can't be such central entity that's trusted by all individuals?

<u>Example</u>: we wish to find how many women in North Carolina have undergone abortion? We take a sample of women there & ask them. This is what statistician have been doing

for decades. The caveat here in North Carolina, it is a crime to do abortion. So nobody answer the quetion truthfully!

To address problem like this, we need to let each individual randomize their data before they release it publicly.

Consider the following the mechanism:

$$x \longrightarrow \boxed{\;M\;} \longrightarrow Z_x \sim M_x$$

$\uparrow$ individual's input

$\uparrow$ mechanism's output

Def. We say mechanism M is ε-locally DP (or ε-LDP for short)

if

$$E_{e^{\varepsilon}}(M_x \| M_{x'}) = 0 \qquad \forall\, x \,\&\, x' \text{ possible inputs}.$$

Example. Randomized-response mechanism:

$$\underset{\in\{0,1\}}{x} \longrightarrow \boxed{} \longrightarrow Z \qquad Z \sim \mathrm{Ber}(r) \text{ if } x=0 \quad \& \quad Z \sim \mathrm{Ber}(1-r) \text{ if } x=1$$

In fact, r indicates the lying probability.

* Show that this mechanism is ε-LDP if $r = \frac{1}{1+e^{\varepsilon}}$.

Theorem. A mechanism M is ε-LDP <u>if & only if</u>

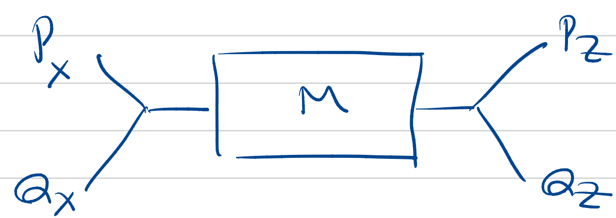
$$\eta_{e^{\varepsilon}}(M) = 0$$

Proof: HW4

In other word, an ε-LDP mechanism always satisfied:

$$E_{e^{\varepsilon}}(P_Z \| Q_Z) = 0 \quad \forall\, P_X \,\&\, Q_X$$

where P_Z & Q_Z are the output distributions of M when input distributions are P_X & Q_X.



Example: Prove that the following mechanism is ε-LDP :

$$P_{Z|X}(z|x) = \begin{cases} \dfrac{e^{\varepsilon}}{e^{\varepsilon}+k-1} & z = x \in \{1,2,\dots,k\} \\ \dfrac{1}{e^{\varepsilon}+k-1} & z \neq x \\ 0 & z \notin \{1,2,\dots,k\} \end{cases}$$

* In HW1, we already proved that $\eta_{e^{\varepsilon}}(P_{Z|X}) = 0$, showing that this mechanism is ε-LDP.

* this mechanism is known as k-ary randomized response, because both input & output take values in $\{1,2,\dots,k\}$.