Concentrated Differentially Private Gradient Descent with Adaptive per-Iteration Privacy Budget

Paper by Jaewoo Lee & Daniel Kifer 2018

Differential Privacy

Lemma 3.5 ([4]). *Suppose two mechanisms satisfy ρ_1-zCDP and ρ_2-zCDP, then their composition satisfies $(\rho_1 + \rho_2)$-zCDP.*

Lemma 3.6 ([4]). *The Gaussian mechanism, which returns $q(D) + N(0, \sigma^2)$ satisfies $\Delta_2(q)^2/(2\sigma^2)$-zCDP.*

Lemma 3.7 ([4]). *If $\mathcal{M}$ satisfies ϵ-differential privacy, them $\mathcal{M}$ satisfies $(\frac{1}{2}\epsilon^2)$-zCDP.*

Lemma 3.8 ([4]). *If $\mathcal{M}$ is a mechanism that provides ρ-zCDP, then $\mathcal{M}$ is $(\rho + 2\sqrt{\rho \log(1/\delta)}, \delta)$-DP for any $\delta > 0$.*

$(\epsilon_{tot}, \delta_{tot})$-DP.

$\rho + 2\sqrt{\rho \log(1/\delta_{tot})} \le \epsilon_{tot}$

Gradient Descent

- Objective: Find $\mathbf{w}^* \in \mathbb{R}^p$ minimizes an objective function f

- Start with an initial guess $\mathbf{w}_0$
- Generate a sequence of iterates, updates have a form

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \alpha_t(\nabla f(\mathbf{w}_t))$$

Differentially Private Gradient Descent

- Objective: Find $\mathbf{w}^* \in \mathbb{R}^p$ minimizes an objective function f

$(\epsilon, 0)$-DP T $\frac{\epsilon}{T}$

- Start with an initial guess $\mathbf{w}_0$
- Generate a sequence of iterates, updates typically have a form

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \alpha_t(\nabla f(\mathbf{w}_t) + Y_t)$$

budget ϵ split evenly.
$\epsilon = \epsilon_1 + \dots + \epsilon_T = T \cdot \epsilon_1$

where Y_t is an appropriately scaled noise variable (e.g., Laplace or Gaussian)

T too small, algorithm stops before optimum

- Problems: pre-specified number of iterations; fixed allocation of private budget

T too large, budget ϵ_t is small, noise large

Differentially Private Gradient Descent with Adaptive per-Iteration Privacy Budget

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \alpha_t(\nabla f(\mathbf{w}_t) + Y_t)$$

- Solution:

ϵ_t

Use part of the privacy budget allocated to step *t* to compute the noisy gradient $S_t = \nabla f(\mathbf{w}_t) + Y_t.$

Use the remaining part of the privacy budget to select the best step size α_t

- If the selected step size α is not 0, update

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \alpha \tilde{S}_t$$

- If α is 0, it is likely that the noise was so large that the noisy gradient is not a descent direction
  - Triggers an increase in share of the privacy budget assigned to subsequent steps
  - Measures the noisy gradient again $\epsilon_{t+1} - \epsilon_t$, merge the result with our previous noisy gradient

How to Select the Best Step Size α_t for

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \alpha_t(\nabla f(\mathbf{w}_t) + Y_t)$$

Idea: Start with a predefined set of step sizes Φ , use the differentially private noisy min algorithm to approximately find the $\alpha \in \Phi$, which $f(\mathbf{w}_t - \alpha \tilde{S}_t)$ is smallest

- NosiyMax Algrithm

- NosiyMax Algrithm

Algorithm 1: NoisyMax($\Omega, \Delta_1(f), \epsilon$)

Input: Ω: a set of candidates, $\Delta_1(f)$: sensitivity of f, ϵ: privacy budget for pure differential privacy

1 $\widetilde{\Omega} = \{\tilde{v}_i = v + \mathsf{Lap}(\Delta_1(f)/\epsilon) : v \in \Omega, i \in [|\Omega|]\}$

2 **return** $\arg\max_{j \in [|\Omega|]} \tilde{v}_j$

---

an $(\epsilon, 0)$-DP algorithm / $\frac{\epsilon^2}{2}$-zCDP

Let $\psi = \{w_1, \cdots, w_s\}$ be a set of points in $\mathbb{R}^p$

$f: \mathbb{R}^p \to \mathbb{R}$ be a function

want to find $w_i \in \psi$ with max $f(w_i; D)$

$\to i = \arg\max_{j \in [s]} \{f(w_j) + \mathrm{Lap}(\Delta f/\epsilon)\}$.

set $-f(w_j)$ noisyMin

$\psi = \{w_1, \cdots, w_s\}$.

$[s]$

$w_i \in \psi$ with max $f(w_i; D)$

$i = \arg\max_{j \in [s]} \{\underline{f(w_j)} + \mathrm{Lap}(\Delta f/\epsilon)\}$.

$-f(w_j)$

How to Merge the Result of the Noisy Gradients

Lemma 3.6 $q(D) + N(0, \sigma^2)$ satisfies $\frac{\Delta_2(q)^2}{2\sigma^2}$-zCDP

- Gradient Averaging Algorithm
  - current noisy gradient under zCDP

    ρ_t-zCDP

    L2 sensitivity

    $$S_t = \nabla f(\mathbf{w}_t) + N(\mathbf{0}, \frac{\Delta_2(\nabla f)^2}{2\rho_t})$$

  - another independent measurement using $\rho_{t+1} - \rho_t$ privacy budget

    $(\rho_{t+1} - \rho_t)$-zCDP

    $$S'_t = \nabla f(\mathbf{w}_t) + N(\mathbf{0}, \frac{\Delta_2(\nabla f)^2}{2(\rho_{t+1}-\rho_t)})$$

  - combine S_t and S'_t in the following way Lemma 3.5 ρ_{t+1}-zCDP

    $$\hat{S}_t = \frac{\rho_t S_t + (\rho_{t+1} - \rho_t) S'_t}{\rho_t + (\rho_{t+1} - \rho_t)}$$

    - calculations show that

$$E[\hat{S}_t] = \nabla f(\mathbf{w}_t)$$

$$Var(\hat{S}_t) = \left(\rho_t^2 \frac{\Delta_2(\nabla f)^2}{2\rho_t} + \frac{\Delta_2(\nabla f)^2}{2(\rho_{t+1} - \rho_t)}(\rho_{t+1} - \rho_t)^2\right) / \rho_{t+1}^2 = \frac{\Delta_2(\nabla f)^2}{2\rho_{t+1}}$$

- Gradient Averaging Algorithm

ρ_t ρ_{t+1} S_{t+1} S_t Clip val

Function $\textsc{GradAvg}(\rho_{\text{old}}, \rho_H, \mathbf{g}, \tilde{\mathbf{g}}, C_{\text{grad}})$:

$$\tilde{\mathbf{g}}_2 \leftarrow \mathbf{g} + N(\mathbf{0}, (\frac{C_{\text{grad}}^2}{2(\rho_H - \rho_{\text{old}})})\mathbf{I})$$

$$\tilde{S} \leftarrow \frac{\rho_{\text{old}}\tilde{g} + (\rho_H - \rho_{\text{old}})\tilde{g}_2}{\rho_H}$$

return $\tilde{S}$

end

while $\rho > 0$ **do**

$i \leftarrow 0$

$\mathbf{g}_t \leftarrow \sum_{i=1}^{n} \left(\nabla \ell(\mathbf{w}_t; d_i) / \max(1, \frac{\|\nabla \ell(\mathbf{w}_t)\|_2}{C_{\text{grad}}}) \right)$

$\tilde{\mathbf{g}}_t \leftarrow \mathbf{g}_t + N(0, (C_{\text{grad}}^2 / 2\rho_{\text{ng}}) \mathbf{I})$ ρ_{ng} - zCDP Lemma 3.6

$\rho \leftarrow \rho - \rho_{\text{ng}}$

$\tilde{\mathbf{g}}_t \leftarrow \tilde{\mathbf{g}}_t / \|\tilde{\mathbf{g}}_t\|_2$

while $i = 0$ **do**

$\Omega = \{ f(\mathbf{w}_t - \alpha \tilde{\mathbf{g}}_t) : \alpha \in \Phi \}$

$\rho \leftarrow \rho - \rho_{\text{nmax}}$ ρ_{nmax} - zCDP

$i \leftarrow$ NoisyMax$(-\Omega, C_{\text{obj}}, \sqrt{2\rho_{\text{nmax}}})$

if $i > 0$ **then**

if $\rho > 0$ **then** $\mathbf{w}_{t+1} \leftarrow \mathbf{w}_t - \alpha_i \tilde{\mathbf{g}}_t$

else

$\rho_{\text{old}} \leftarrow \rho_{\text{ng}}$

$\rho_{\text{ng}} \leftarrow (1+\gamma)\rho_{\text{ng}}$

$\tilde{\mathbf{g}}_t \leftarrow$ GradAvg$(\rho_{\text{old}}, \rho_{\text{ng}}, \mathbf{g}_t, \tilde{\mathbf{g}}_t, C_{\text{grad}})$

$\rho \leftarrow \rho - (\rho_{\text{ng}} - \rho_{\text{old}})$

end

end

$t \leftarrow t + 1$

end

return $\mathbf{w}_t$

A General Framework: DP-AGD Algorithm

Input database: $D = \{d_1, \ldots, d_n\}$

$d_i = \{x_i, y_i\}$

$x_i \in \mathbb{R}^p$

ERM problem: $\underset{\mathbf{w} \in C}{\text{minimize}} \; f(\mathbf{w}; D) := \frac{1}{n} \sum_{i=1}^{n} \ell(\mathbf{w}; d_i)$

- Private gradient approximation

$$g_t = \sum_{i=1}^{n} \frac{\nabla \ell(w_t; d_i)}{\max(1, \frac{\|\nabla \ell(w_t)\|_2}{C_{grad}})} = \begin{cases} \sum \nabla \ell(w_t; d_i) & \|\nabla \ell\|_2 < C_{grad} \\ \sum_{i=1}^{n} \frac{\nabla \ell(w_t; d_i)}{\|\nabla \ell(w_t)\|_2} \cdot C_{grad} & \|\nabla \ell\|_2 > C_{grad} \end{cases}$$

add Gaussian noise with $\sigma^2 = \frac{C_{grad}^2}{2\rho_{ng}}$ L_2 sens. bound by C_{grad}

- Step size selection

difficulty:
no known a priori bound on ℓ
→ apply grad. clipping tech.
clip val. > C_{obj}. then take the summation

- Adaptive noise reduction

- $\tilde{g}_t$ is a bad direction
→ increase ρ_{ng} by a factor of $(1+\gamma)$
→ gradAvg to get $\tilde{g}_t$ → check by NoisyMax( )

A General Framework: DP-AGD Algorithm

ρ_{nmax} ρ_{ng} ε_{tot} δ_{tot} C_{grad} C_{obj} γ D

while $\rho > 0$ **do**

  $i \leftarrow 0$

  $\mathbf{g}_t \leftarrow \sum_{i=1}^{n} \left(\nabla \ell(\mathbf{w}_t; d_i) / \max(1, \frac{\|\nabla \ell(\mathbf{w}_t)\|_2}{C_{\text{grad}}}) \right)$ $\Rightarrow \rho$ – zCDP ($\varepsilon_{tot}, \delta_{tot}$) – DP

  $\tilde{\mathbf{g}}_t \leftarrow \mathbf{g}_t + N(0, (C^2_{\text{grad}}/2\rho_{\text{ng}})\mathbf{I})$ ρ_{ng} – zCDP

  $\rho \leftarrow \rho - \rho_{\text{ng}}$

  $\tilde{\mathbf{g}}_t \leftarrow \tilde{\mathbf{g}}_t / \|\tilde{\mathbf{g}}_t\|_2$

  while $i = 0$ **do**

    $\Omega = \{f(\mathbf{w}_t - \alpha \tilde{\mathbf{g}}_t) : \alpha \in \Phi\}$

    $\rho \leftarrow \rho - \rho_{\text{nmax}}$ ρ_{nmax} – zCDP

    $i \leftarrow$ NoisyMax$(-\Omega, C_{\text{obj}}, \sqrt{2\rho_{\text{nmax}}})$

    if $i > 0$ **then**

      if $\rho > 0$ **then** $\mathbf{w}_{t+1} \leftarrow \mathbf{w}_t - \alpha_i \tilde{\mathbf{g}}_t$

    else

      $\rho_{\text{old}} \leftarrow \rho_{\text{ng}}$

      $\rho_{\text{ng}} \leftarrow (1 + \gamma)\rho_{\text{ng}}$

      $\tilde{\mathbf{g}}_t \leftarrow$ GradAvg$(\rho_{\text{old}}, \rho_{\text{ng}}, \mathbf{g}_t, \tilde{\mathbf{g}}_t, C_{\text{grad}})$ $\alpha = 0$

      $\rho \leftarrow \rho - (\rho_{\text{ng}} - \rho_{\text{old}})$

    end

  end

  $t \leftarrow t + 1$

end

return $\mathbf{w}_t$

Input database: $D = \{d_1, \ldots, d_n\}$

$\varepsilon_{tot} \geq \rho + 2\sqrt{\rho \log(1/\delta_{tot})}$

ERM problem: $\underset{\mathbf{w} \in C}{\text{minimize}}\ f(\mathbf{w}; D) := \frac{1}{n} \sum_{i=1}^{n} \ell(\mathbf{w}; d_i)$

- Composition (ε, δ)-DP

conversion by lemma 3.7, 3.8

$\varepsilon_{tot} \geq \rho + 2\sqrt{\rho \log(1/\delta_{tot})}$.

total privacy budget ρ for zCDP

dynamically compute & deduct the amount of required privacy budget

- Adjusting step sizes

① the var. in private grad est. need to be control

② it is possible $\tilde{g}_t$ is a descent direction but NoisyMax fails to choose a step size since all step sizes in candidate set are large

$\alpha_{max} = (1+\eta) \max(\alpha_t, \alpha_{t-1}, \ldots, \alpha_{t-T+1})$. at every T iter

Splits small

algs might not have enough

~~$\epsilon_{max} = \epsilon_{avg}$ Large~~

~~less noise~~