Removing Disparate Impact on Model Accuracy in Differentially Private Stochastic Gradient Descent

0. Abstract

DPSGD

DPSGD 中，gradient clipping and random noise addition 不成比例地影响了 underrepresented and complex classes and subgroups.

Disparate impact

decrease more accuracy on these classes and subgroups vs. the non-private model.

如果 the original model is unfair (accuracy 在所有的 subgroups 中 不是相同的)，DPSGD 只会加剧这种 unfairness.

Contribution

- 研究了 the inequality in utility loss due to DP (比较 private model and non-private model)
- 分析了 the cost of privacy ， 并解释了 group sample size 是如何影响 group accuracy.
- 提出了 DPSGD-F, 实现 DP，满足 equal costs and good utility.

  自适应地根据 group clipping bias 调整 group 中 samples 的 contribution, 使得 DP 在 group accuracy 上没有 disparate impact.

1. Introduction

within-model

fairness: non-discriminative (只有 protected group)

> demographic parity 要求： a prediction 是对于 the protected attribute 是相互独立的。

unfairness: disproportionately impacts underrepresented subgroups.

cross-model

比较介于 the majority group and the protected group accuracy loss

unfairness: accuracy reduction is discriminatory against the protected group.

Paper study

由于 DP 带来的 the inequality in utility loss.

本文比较了在 private model 和 non-private model 之间的 prediction accuracy

DP : 保证了query 的结果不能被 attacker 利用，去提取某个 record 是否 present or absent.

cost of privacy: the utility loss between the private and non-private models, 即是 utility-privacy trade-off.

group sample sizes 的不同，导致了在 average group gradient norms的不同，进一步导致在 uniform clipping bound 情况下，不同的 group clipping biases.

> 更小的 group sample size 或者更大的group clipping bias 在实现相同的 DP 水平时候，会导致更多的 utility loss

DPSGGD-F

remove 在 utility loss 方面的 inequality

根据 the group clipping bias，调整一个 group 中 samples的 contribution

Result: differential privacy has no disparate impact on group accuracy in DPSGD-F

2. Related Works

2.1 Differential Privacy

- input perturbation

  local differential privacy

- output perturbation

  add noise to the model after the training procedure finishes.

- inner perturbation.

modify the learning algorithm

> adaptively bound the contributions of users
>
> adaptively clipping of the gradient
>
> adaptively clips to a value at a specified quantile
>
> adaptively injects noise into features
>
> adaptively allocates per-iteration privacy budget

2.2 Fairness-aware Machine Learning

许多研究通过 modify the training data 去 mitigating bias 并且实现 fairness.

> - 通过调整 learning process 或者 changing the predicted labels 来 mitigate discriminative bias in model prediction.
> - 使用 adversarial learning techniques 来实现在 classification and representation learning 中的 fairness.
> - adaptive sensitive reweighting 来识别 bias 的来源
> - agnostic learning 实现一个较好的 accuracy

这些方法不能直接应用于 DPSGD，因为 unbounded sensitivity，不能找到一个最优的 strategy.

2.3 Differential Privacy and Fairness

研究 privacy protection and fairness 之间的联系

- 实现 k-anonymity and fairness.
- 研究在 private model 增加 DP 外的 within-model fairness
- our work: prevent the disparate impact of the private model on model accuracy across different groups

3. Preliminary

> D 是有 n 个 tuples $x_1, x_2, \ldots, x_n$ 的 dataset
>
> x_i 包括 user i 在 d 个 unprotected attributes $A_1, A_2, \ldots, A_d$, protected attribute S and decision Y.
>
> D^k 代表有 S=k tuples 的子集

最优模型参数 $w* = argmin_w \frac{1}{n} \sum_{i=1}^{n} L_i(w)$

- non-private model 输出 a classifier $\eta(\alpha; w)$
- differentially private model 输出 a classifier $\tilde{\eta}(\alpha; \widetilde{w})$

  $\widetilde{w}$ 满足 DP 的同时，尽可能接近实际最优的 $w*$

3.1 Differential Privacy

Differential privacy

DP 保证了 query q 的输出对于 record 是否在 dataset 中是不敏感的

- 参数 ϵ 代表了 the privacy budget

  控制由 D 和 D'产生的分布不同的数量

- 参数 δ 是一个 broken probability.

  ϵ and δ 越小，意味着更强的隐私保证

Global sensitivity

$\triangle f = max_{D.D'} |q(D) - q(D'|$

表示，当数据集中的一条数据发生变化时，测量的可能发生的最大变化

Gaussian mechanism

给 model output 的每个 component 添加 Gaussian noise N(0,σ^2)

3.2 Differentially Private SGD

DPSGD

- 对于独立的 updates，使用 l_2 norm 的 clipping bound
- aggregates the clipped updates.
- adds Gaussian noise to the aggregate.

DPSGD 的 privacy leakage 是通过 (ϵ, δ) 测量的, 对于 privacy loss ϵ 计算一个 bound, 能够满足某种概率 δ

moment accounting mechanism: 计算一个 aggregate privacy bound. tighter bounds. (为 Gaussian mechanism 定制，提取 total privacy loss bound)

DPSGD 通过 truncates neural network 中的 gradients 来控制梯度和的灵敏度 （因为梯度的灵敏度和噪声的规模是无界的）

C 限制用户最大的 contribution （这会导致 bias 增大，但是也减小了添加的 noise 的数量）

Algorithm 1 DPSGD (Dataset D, loss function $\mathcal{L}_D(w)$, learning rate r, batch size b, noise scale σ, clipping bound C)

1: **for** $t \in [T]$ **do**
2: Randomly sample a batch B_t with $|B_t| = b$ from D
3: **for** each sample $x_i \in B_t$ **do**
4: $g_i = \nabla L_i(w_t)$
5: **for** each sample $x_i \in B_t$ **do**
6: $\bar{g}_i = g_i \times \min\left(1, \frac{C}{|g_i|}\right)$
7: $\tilde{G}_B = \frac{1}{b}\left(\sum_i \bar{g}_i + N(0, \sigma^2 C^2 \mathbf{I})\right)$
8: $\tilde{w}_{t+1} = \tilde{w}_t - r\tilde{G}_B$
9: Return $\tilde{w}_T$ and accumulated (ϵ, δ)

如何选择 *the truncation level?*

- too high, noise level 会很大，以至于结果的 utility 丧失
- too low, 大量的 gradients 会被 clipped.
- DPSGD 直接选择了 gradients 的中位数

4. Disparate impact on model accuracy

4.1 Preliminary Observation

- 对underrepresented 的 subgroups 训练，会产生更大的 gradients
- 随机噪声添加对underrepresented inputs有最大的影响
- DP 对于 underrepresented group 有 negative discrimination.

MNIST

- private model 下， minority group 有 larger utility loss.

> 在 well-represented classes (class 2)有 -0.0707 的 accuracy loss
>
> 在 underrepresented class (class 8)有显著性的较大达到 -0.6807 utility loss

- 与non-private SGD 相比，在 DPSGD 中 small sample size 减少了 the convergence rate 以及 the optimal utility of class 8
- 该model 远未收敛, clipping and noise addition 没有使其 move closer to the loss function 的最小值
- DP 减慢了 convergence，degrades the utility.

Adult and Dutch

Adult 是一个 unbalanced dataset, the female group is underrepresented.

即使male group 是 main group, 但是与 female group 相比，在 SGD 上有更低的 accuracy, 以及在 DPSGD 更多的 utility loss

---

Dutch dataset 是一个 balanced dataset

group sample size 对于 male 和 female 是相似的

然而对于 DPSGD 在 male group 上引入了更多的 negative discrimination.

male group 带来了更多的 accuracy loss

Dataset	MNIST			Adult			Dutch		
Group	Total	Class 2	Class 8	Total	M	F	Total	M	F
Sample size	54649	5958	500	45222	30527	14695	60420	30273	30147
SGD	0.9855	0.9903	0.9292	0.8099	0.7610	0.9117	0.7879	0.8013	0.7744
DPSGD vs. SGD	-0.1081	-0.0707	-0.6807	-0.0592	-0.0740	-0.0281	-0.1001	-0.1534	-0.0466

在 DPSGD 中，the average gradient norm , male group 要高很多

Table 2: The average loss and the average gradient norm w.r.t. groups at the last training epoch on the MNIST ($\epsilon = 6.55, \delta = 10^{-6}$), Adult ($\epsilon = 3.1, \delta = 10^{-6}$) and Dutch ($\epsilon = 2.66, \delta = 10^{-6}$) datasets

	Average loss						Average gradient norm					
Dataset	MNIST		Adult		Dutch		MNIST		Adult		Dutch	
Group	Class 2	Class 8	M	F	M	F	Class 2	Class 8	M	F	M	F
SGD	0.04	0.04	0.48	0.27	0.52	0.53	0.68	4.76	0.08	0.11	0.19	0.19
DPSGD	0.41	2.16	0.68	0.31	0.59	0.53	13.53	100.46	0.41	0.12	0.26	0.12

DP 带来的 inequality in utility loss, 可能不仅仅取决于 represented sample size, 还有 classification model. 实现DP 的机制等

observation： 有着更大的 utility loss 的 group，通常有着更大的 gradients 和 更差的 convergence.

eg.

- the underrepresented class 8 的 average gradient norm 超过 100，并且在 DPSGD 上较差的 utility.
- male group 比 female group 有更大的 average gradient norm

为了缓解 inequality in utility loss, 重要的是解决 larger gradients 和 worse convergence.

4.2 Cost of Privacy w.r.t. Each Group

utility loss : expected error of the estimated private gradient 来度量

B_t: a collection of b samples, $x_1, x_2, \ldots, x_b$

每个x_i 对应一个 sample, 生成 gradient g_i

- gradient before clipping: $G_B = \frac{1}{b}\sum_{i=1}^{b} g_i$
- gradient after clipping: $\overline{G}_B = \frac{1}{b}\sum_{i=1}^{b} \overline{g}_i$
- gradient after clipping and adding noise: $\widetilde{G}_B = \frac{1}{b}(\sum_{i=1}^{b} \tilde{g}_i + Lap(\frac{C}{\epsilon}))$
- The expected error of estimate $\widetilde{G}_B$:

  (a variance term: noise)

  (a bias term: contribution)

$$\mathbb{E}|\tilde{G}_B - G_B| \le \mathbb{E}|\tilde{G}_B - \bar{G}_B| + |\bar{G}_B - G_B| \le \frac{1}{b}\frac{C}{\epsilon} + \frac{1}{b}\sum_{i=1}^{b}\max(0, |g_i| - C).$$

expected error is tight

$$\mathbb{E}|\tilde{G}_B - G_B| \ge \frac{1}{2}\left[\frac{1}{b}\frac{C}{\epsilon} + \frac{1}{b}\sum_{i=1}^{b} max(0, |g_i| - C)\right].$$

---

batch of samples

batch samples B_t 来自 K 个 groups.

group k 的 sample size = b^k

$G_B^k = \frac{1}{b^k}\sum_{i=1}^{k} g_i^k$

$G_B = \frac{1}{b}\sum_{k=1}^{K} b^k G_B^k$

(1) DPSGD 通过 clipping bound C， 对每个 sample 的 gradient 进行 clipping.

$\overline{G}_B^k = \frac{1}{b^k}\sum_{i=1}^{b^k}\overline{g}_I^k = \frac{1}{b^k}\sum_{i=1}^{b^k} g_i^k \times min(1, \frac{C}{|g_i^k|})$

(2) DPSGD 在clipping 后的梯度和上添加 Laplace noise.

$\widetilde{G}_B^k = \frac{1}{b^k}(b^k\overline{G}_B^k + Lap(\frac{C}{\epsilon}))$

(3) The expected error of the estimate $\widetilde{G}_B^k$

$$\mathbb{E}|\tilde{G}_B^k - G_B^k| \le \mathbb{E}|\tilde{G}_B^k - \bar{G}_B^k| + |\bar{G}_B^k - G_B^k|$$

$$\le \frac{1}{b^k}\frac{C}{\epsilon} + \frac{1}{b^k}\sum_i^{b^k}\max(0, |g_i^k| - C) = \frac{1}{b^k}\frac{C}{\epsilon} + \frac{1}{b^k}\sum_i^{m^k}(|g_i^k| - C), \quad (1)$$

variance of the noise

bias due to contribution.

m^k 是在 group k 中被 clipped 的 examples 的数量

$m^k = |\{i : |g_i^k > C|\}|$

(4) The tight bound

$$\mathbb{E}|\tilde{G}_B^k - G_B^k| \ge \frac{1}{2}\left[\frac{1}{b^k}\frac{C}{\epsilon} + \frac{1}{b^k}\sum_i^{b^k}\max(0, |g_i^k| - C)\right].$$

分析：

group k 的 **utility loss**

- **bias** 来源于 contribution limit: $\frac{1}{b^k}\sum_i^{b^k} max(0, |g_i^k| - C)$

取决于 the size of gradients and the size of clipping bound.

- **variance** of noise: $\frac{1}{b^k}\frac{C}{\epsilon}$

取决于 the scale of the noise.

gradients 越大， the bias 越大

Before clipping

对于group 来说, gradients 越大， total gradient G_B 就有越大的 contribution.

After clipping

$\widetilde{G}_B$ 越接近于更小的 bias 或者更小的 gradients 的方向, 因为 clipping，导致 large gradients 的 contribution and convergence 都被减小

noise

noise 减慢了 convergence rate of the model.

the noise scales $\frac{C}{\epsilon}$ 以及 sensitivity of clipped gradients C 对于所有的 groups 都是一样的，都有相同的 DP 强度 ϵ。noise 的方向是随机的，对于每个 group 在期望上是相同的。

Overall

DPSGD, large gradients 的 group 有 larger cost of privacy.

- MNIST

larger sample size (the majority group) --> 对于 total gradients, 有 larger contribution --> faster and better convergence.

在之后， the minority group 有 larger gradients.

small sample size 是 large gradient norm 和 large utility loss 的主要原因

- Adult and Dutch

male group 的 average gradient norm 更大

因此，male group 的 contribution 通过 clipping 被限制，也就有了更大的 utility loss.

small group sample size and other factors --> large average gradient norm --> large cost of privacy.

DPSGD 中，clipping bound 是被每个 group uniformly 使用，而没有考虑 clipping bias 的不同，导致了在 noise addition 实现了不同的 utility-privacy trade-off，使得 underrepresented 的 group 导致了更大的 utility loss.

改进：为每个 group 实现不同水平的 privacy，来抵消不同的 cost of privacy

5. Removing Disparate Impact on Model Accuracy in DPSGD

目标：

- 实现 DP
- equality of utility loss
- good accuracy

5.1 Equal Costs of Differential Privacy

DP 导致了 accuracy loss

但是，不同的 groups 可能导致不同程度的 accuracy loss.

在 private model and non-private model 下的 accuracy reduction 使用 $\triangle^k$ 表示

new fairness: equal costs of differential privacy. 对于每个 group 的 utility loss 是一样的

即对于任意的i 和 j 满足 $\triangle^i(\tilde{\eta} - \eta) = \triangle^j(\tilde{\eta} - \eta)$.

5.2 Removal Algorithm

framework: adaptive sensitive clipping (每个 group k 都有自己的 clipping bound C^k)

对于更大的 clipping bias（large gradient），选择一个更大的 clipping bound。(large gradient 可能由于 group sample size 或者其他因素影响)

每个 group 的 contribution 与他们的 average gradient 成比例

Ideally：根据平均梯度norm的 private estimate来调整 clipping bound.

但是，clipping 之前的梯度是 unbounded, 导致不能得到 private estimate.

因此，要得到一个 average gradient 的近似估计

根据 m^k 来选择clipping bound C^k

使用 $\frac{m^k}{b^k}$ 代表 group k 中梯度值大于 C_0 的比例

$\frac{m^k}{b^k}$ 和 $\frac{m}{b}$ 的比例，可以近似代表 average gradient 的相对大小

group k 的 clipped gradient 的 sensitivity 是 $C^k = C_0 \times (1 + \frac{\tilde{m}^k/\tilde{b}^k}{\tilde{m}/b})$

total population 的 clipped sensitivity是 $max_k C^k$ (考虑的是最坏情况)

adding noise:

(1) large noise scale σ_1 (small privacy budget) 得到 private collection $\{\tilde{m}^k, \tilde{o}^k\}$

(2) small noise scale σ_2 去 perturb the gradients.

Algorithm 2 DPSGD-F (Dataset D, loss function $\mathcal{L}_D(w)$, learning rate r, batch size b, noise scales σ_1, σ_2, base clipping bound C_0)

for $t \in [T]$ **do**
  Randomly sample a batch B_t with $|B_t| = b$ from D
  for each sample $x_i \in B_t$ **do**
    $g_i = \nabla L_i(w_t)$
  for each group $k \in [K]$ **do**
    $m^k = \left|\left\{i : |g_i^k| > C_0\right\}\right|$
    $o^k = \left|\left\{i : |g_i^k| \leq C_0\right\}\right|$
  $\left\{\tilde{m}^k, \tilde{o}^k\right\}_{k\in[K]} = \left\{m^k, o^k\right\}_{k\in[K]} + N(0, \sigma_1^2 \mathbf{I})$
  $\tilde{m} = \sum_{k\in[K]} \tilde{m}^k$
  for each group $k \in [K]$ **do**
    $\tilde{b}^k = \tilde{m}^k + \tilde{o}^k$
    $C^k = C_0 \times \left(1 + \frac{\tilde{m}^k / \tilde{b}^k}{\tilde{m}/b}\right)$
  for each sample $x_i \in B_t$ **do**
    $\bar{g}_i = g_i \times \min\left(1, \frac{C^k}{|g_i|}\right)$
  $C = \max_k C^k$
  $\tilde{G}_B = \frac{1}{b}\left(\sum_i \bar{g}_i + N(0, \sigma_2^2 C^2 \mathbf{I})\right)$
  $\tilde{w}_{t+1} = \tilde{w}_t - r\tilde{G}_B$
Return $\tilde{w}_T$ and accumulated (ϵ, δ)

根据 contribution，C^k 与平均梯度大小成比例

MNIST

Algorithm 1 中，每个 group 实现相同程度的 privacy，但是 the underrepresented group 有更高的 privacy cost.

Algorithm 2 中，为 underrepresented group 选择更大的 clipping bound C

Adult/Dutch

smaller gradients 有更小的 privacy cost

Algorithm 2 为更小梯度的 group 实现了更高水平的 privacy.

5.3 Baseline

只考虑 sample size 的 contribution

基于 reweighting 技术提出 Algorithm 3

Algorithm 3 Naïve (Dataset D, loss function $\mathcal{L}_D(w)$, learning rate r, batch size b, noise scales σ_1, σ_2, base clipping bound C_0)

for $t \in [T]$ **do**
  Randomly sample a batch B_t with $|B_t| = b$ from D
  for each sample $x_i \in B_t$ **do**
    $g_i = \nabla L_i(w_t)$
  $\left\{\tilde{b}^k\right\}_{k\in[K]} = \left\{b^k\right\}_{k\in[K]} + N(0, \sigma_1^2 \mathbf{I})$
  for each group $k \in [K]$ **do**
    $\theta^k = 1 \times \frac{b/K}{\tilde{b}^k}$
  for each sample $x_i \in B_t$ **do**
    $\bar{g}_i = \theta^k \times g_i \times \min\left(1, \frac{C_0}{|g_i|}\right)$
  $C = C_0 \times \max_k \theta^k$
  $\tilde{G}_B = \frac{1}{b}\left(\sum_i \bar{g}_i + N(0, \sigma_2^2 C^2 \mathbf{I})\right)$
  $\tilde{w}_{t+1} = \tilde{w}_t - r\tilde{G}_B$
Return $\tilde{w}_T$ and accumulated (ϵ, δ)

reweighting

6. Experiments

6.1 Experiment Setup

Datasets

unbalanced **MNIST** dataset

> class 8 作为 underrepresented group (training samples 500)
>
> class 2 作为 well-represented group (training samples 6,000)

census **Adult and Dutch** dataset

> sex 作为 protected attribute
>
> unprotected attributes 使用 one-hot 编码
>
> 40 unprotected attributes for Adult and 35 unprotected attributes for Dutch

Model

MNIST dataset 使用 2 个卷积层 和 2 个线性层的 **neural network**

> learning rate r = 0.01
>
> batch size b = 256
>
> training epochs = 60

census datasets 使用 **logistic regression model**

> learning rate r = $1/\sqrt{T}$
>
> batch size b = 256
>
> training epochs = 20

Baseline

使用 moments accounting method 计算了 accumulated privacy budget ϵ

Metric

- 每个group 的 private SGD 和 non-private SGD 的模型准确率的减少量
- average loss and average gradient norm

6.2 MNIST Dataset

non-private SGD model

> 0.9292 accuracy on class 8
>
> 0.9903 accuracy on class 2

DPSGD model

> -0.6807 accuracy loss on class 8
>
> -0.0707 accuracy loss on class 2

Naive approach

> -0.1510 accuracy loss on class 8
>
> -0.1512 accuracy loss on class 2

DPSGD-F algorithm

> -0.0432 accuracy loss on class 8
>
> -0.0281 accuracy loss on class 2

Dataset	MNIST		
Group	Total	Class 2	Class 8
Sample size	54649	5958	500
SGD	0.9855	0.9903	0.9292
DPSGD vs. SGD	-0.1081	-0.0707	-0.6807
Naïve vs. SGD	-0.1500	-0.1512	-0.1510
DPSGD-F vs. SGD	-0.0293	-0.0281	-0.0432

model accuracy all classes on the MNIST

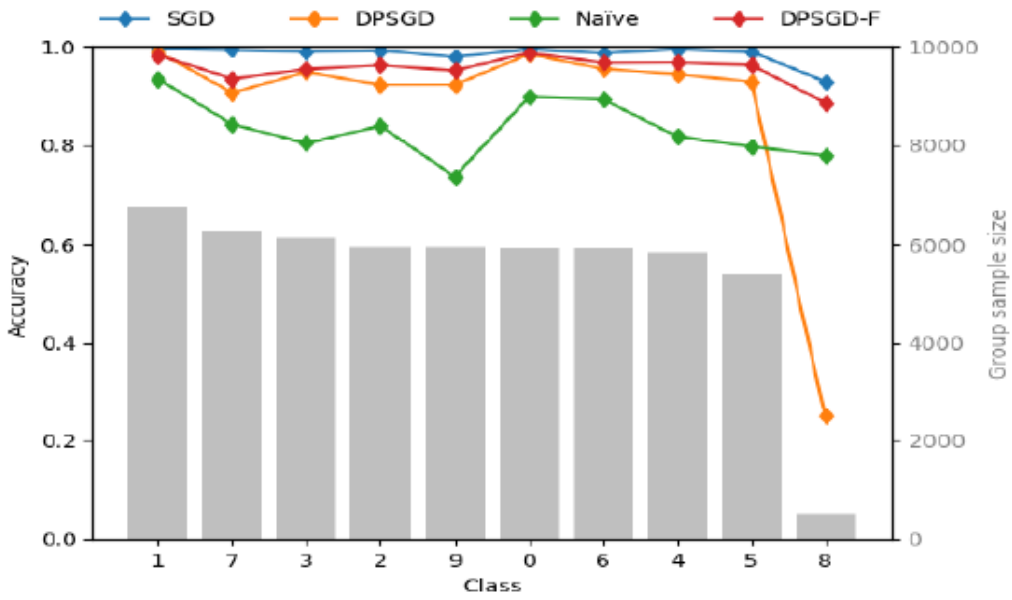


Figure 1: Model accuracy w.r.t. each class on MNIST

DPSGD 的 average gradient norm on class 8 > 100, average loss 2.16

DPSGD-F 的 average gradient norm on class 8 2.53, average loss 0.42

(接近于class 2)

	Average loss						Average gradient norm					
Dataset	MNIST		Adult		Dutch		MNIST		Adult		Dutch	
Group	Class 2	Class 8	M	F	M	F	Class 2	Class 8	M	F	M	F
SGD	0.04	0.04	0.48	0.27	0.52	0.53	0.68	4.76	0.08	0.11	0.19	0.19
DPSGD	0.41	2.16	0.68	0.31	0.59	0.53	13.53	100.46	0.41	0.12	0.26	0.12
Naïve	3.08	1.89	0.71	0.32	0.59	0.53	0.83	0.76	0.43	0.13	0.26	0.12
DPSGD-F	0.20	0.42	0.50	0.27	0.51	0.52	1.45	2.53	0.12	0.08	0.19	0.18

clipping bound changes

class 8 由于 underrepresented group sample size 有较大的 clipping bias, 因此有更高的 clipping bound，来提升对于 total gradient 的 sample contribution

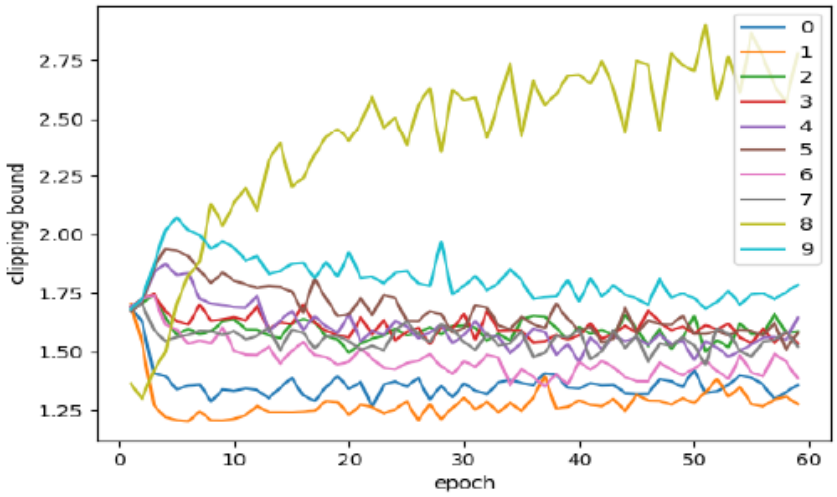


Figure 2: The clipping bound C^k w.r.t. each class over training epochs for DPSGD-F on the MNIST dataset

Different clipping bound

提升bound 可以提高 accuracy，但是在 loss accuracy 上面，仍然存在显著的 difference

Table 3: Model accuracy for different uniform clipping bound ($C = 1, 2, 3, 4, 5$) in DPSGD vs. adaptive clipping bound ($C_0 = 1$) in DPSGD-F on the MNIST dataset

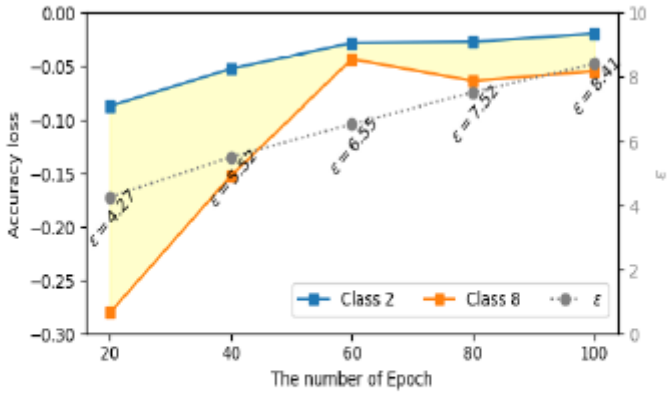
Group	Total	Class 2	Class 8
Sample size	54649	5958	500
SGD	0.9855	0.9903	0.9292
DPSGD ($C = 1$) vs. SGD	-0.1081	-0.0707	-0.6807
DPSGD ($C = 2$) vs. SGD	-0.0587	-0.0426	-0.3286
DPSGD ($C = 3$) vs. SGD	-0.0390	-0.0232	-0.2013
DPSGD ($C = 4$) vs. SGD	-0.0286	-0.0194	-0.1376
DPSGD ($C = 5$) vs. SGD	-0.0240	-0.0145	-0.1099
DPSGD-F ($C_0 = 1$) vs. SGD	-0.0293	-0.0281	-0.0432

Different ϵ

ϵ 受到两个因素影响， the number of epochs and the noise scale.

调整*epochs*

随着epochs的增加，积累的 ϵ 也在增加，class 2 and class 8 的 accuracy loss 的差距也在缩小，超过60 epochs，difference能够维持在门限值 τ

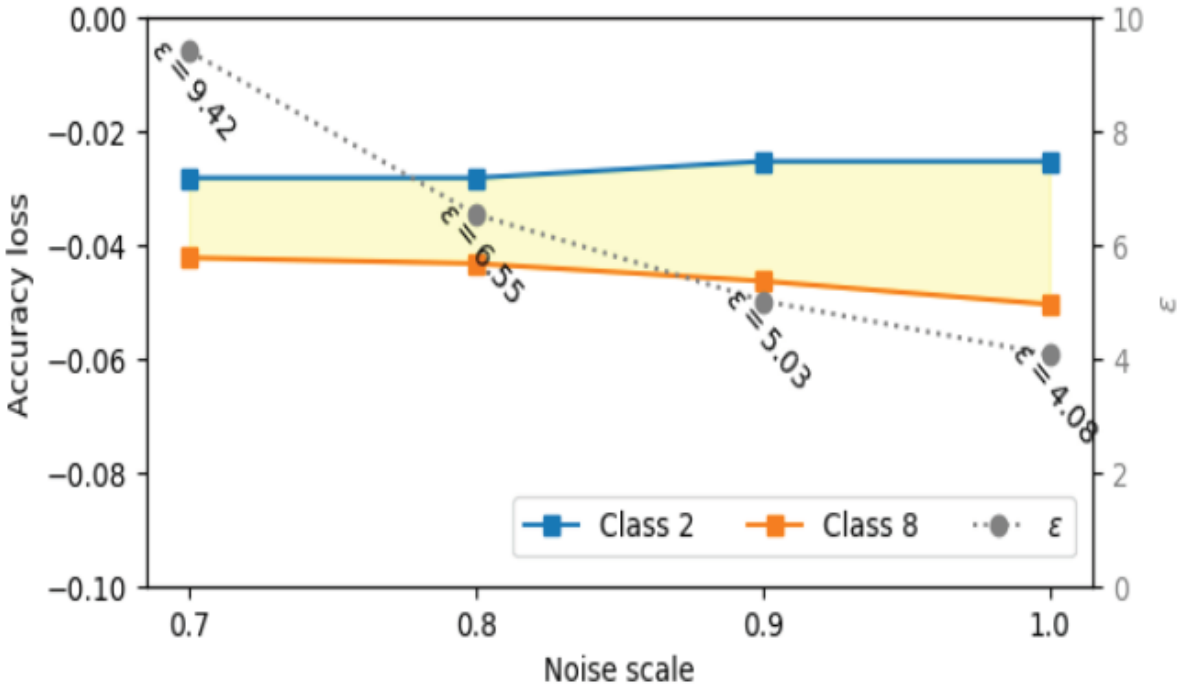


(a) Varying the number of epochs

调整*noise scale*

随着noise scale 的增加, accumulated ϵ 在减小, class 2 and class 8 之间的 accuracy loss 也在增加

因此可以通过增加 noise scale 来得到 stronger privacy。



(b) Varying noise scale

6.3 Adult and Dutch Datasets

Naive approach

在这种场景下，不起作用 （因为 group sample size 的作用不像MNIST 中的明显，还有其他的因素影响了gradient norm 以及 clipping bias）

DPSGD-F 可以在 male and female groups 中，实现类似的 accuracy

Dataset	MNIST			Adult			Dutch		
Group	Total	Class 2	Class 8	Total	M	F	Total	M	F
Sample size	54649	5958	500	45222	30527	14695	60420	30273	30147
SGD	0.9855	0.9903	0.9292	0.8099	0.7610	0.9117	0.7879	0.8013	0.7744
DPSGD vs. SGD	-0.1081	-0.0707	-0.6807	-0.0592	-0.0740	-0.0281	-0.1001	-0.1534	-0.0466
Naïve vs. SGD	-0.1500	-0.1512	-0.1510	-0.0593	-0.0742	-0.0281	-0.1004	-0.1549	-0.0458
DPSGD-F vs. SGD	-0.0293	-0.0281	-0.0432	-0.0254	-0.0298	-0.0161	-0.0130	-0.0160	-0.0099

Adult dataset

> male 的 gradient norm 是 SGD 的 5倍
>
> average lloss 也比SGD 中多 50%

DPSGD-F

> average gradient norm 以及 average loss 都与 SGD 中的类似

	Average loss						Average gradient norm					
Dataset	MNIST		Adult		Dutch		MNIST		Adult		Dutch	
Group	Class 2	Class 8	M	F	M	F	Class 2	Class 8	M	F	M	F
SGD	0.04	0.04	0.48	0.27	0.52	0.53	0.68	4.76	0.08	0.11	0.19	0.19
DPSGD	0.41	2.16	0.68	0.31	0.59	0.53	13.53	100.46	0.41	0.12	0.26	0.12
Naïve	3.08	1.89	0.71	0.32	0.59	0.53	0.83	0.76	0.43	0.13	0.26	0.12
DPSGD-F	0.20	0.42	0.50	0.27	0.51	0.52	1.45	2.53	0.12	0.08	0.19	0.18

7. Conclusion and Future Work

- Gradient clipping and random noise addition 是 DPSGD 的两个核心技术，不同程度地影响了 underrepresented groups.
- DPSGD-F 通过自适应调整samples 的 contribution，实现了相同的 cost-privacy
- Future work，从 group-wise 到 element-wise