Rate-Constrained Optimization Under Differential Privacy

Mohammad Yaghini[*], Tudor Ioan Cebere[*], Michael Menart, Aurélien Bellet, Nicolas Papernot

The Fourteenth International Conference on Learning Representations (ICLR 2026)

University of Toronto and Vector Institute

Rate constraints unify diverse ML challenges

Fairness

Hiring
Demographic Parity
Equal consideration across gender

Criminal Justice
Equalized Odds
Fair risk assessment across race

Rate constraints unify diverse ML challenges

Fairness

Hiring
Demographic Parity
Equal consideration across gender

Criminal Justice
Equalized Odds
Fair risk assessment across race

Healthcare

Medical Diagnosis
False Negative Rate $\leq \gamma$
Don't miss cancer cases

Drug Discovery
Precision $\geq 1 - \gamma$
Minimize false leads

Rate constraints unify diverse ML challenges

Fairness	**Hiring** *Demographic Parity* Equal consideration across gender	**Criminal Justice** *Equalized Odds* Fair risk assessment across race
Healthcare	**Medical Diagnosis** *False Negative Rate* $\leq \gamma$ Don't miss cancer cases	**Drug Discovery** *Precision* $\geq 1 - \gamma$ Minimize false leads
Business	**Customer Retention** *Churn Rate* $\leq \gamma$ Predict and prevent attrition	**Content Recommendation** *Coverage* $\geq 1 - \gamma$ Serve diverse user interests

Rate constraints unify diverse ML challenges

Fairness	**Hiring** *Demographic Parity* Equal consideration across gender	**Criminal Justice** *Equalized Odds* Fair risk assessment across race
Healthcare	**Medical Diagnosis** *False Negative Rate* $\leq \gamma$ Don't miss cancer cases	**Drug Discovery** *Precision* $\geq 1 - \gamma$ Minimize false leads
Business	**Customer Retention** *Churn Rate* $\leq \gamma$ Predict and prevent attrition	**Content Recommendation** *Coverage* $\geq 1 - \gamma$ Serve diverse user interests

General form:

$$\min_{\theta} \quad \text{training loss}(\theta)$$
$$\text{s.t.} \quad \underbrace{\Gamma(\theta)}_{\text{A rate function}} \leq \gamma$$

Rate constraints unify diverse ML challenges

Fairness	**Hiring** *Demographic Parity* Equal consideration across gender	**Criminal Justice** *Equalized Odds* Fair risk assessment across race
Healthcare	**Medical Diagnosis** *False Negative Rate* $\leq \gamma$ Don't miss cancer cases	**Drug Discovery** *Precision* $\geq 1 - \gamma$ Minimize false leads
Business	**Customer Retention** *Churn Rate* $\leq \gamma$ Predict and prevent attrition	**Content Recommendation** *Coverage* $\geq 1 - \gamma$ Serve diverse user interests

General form:

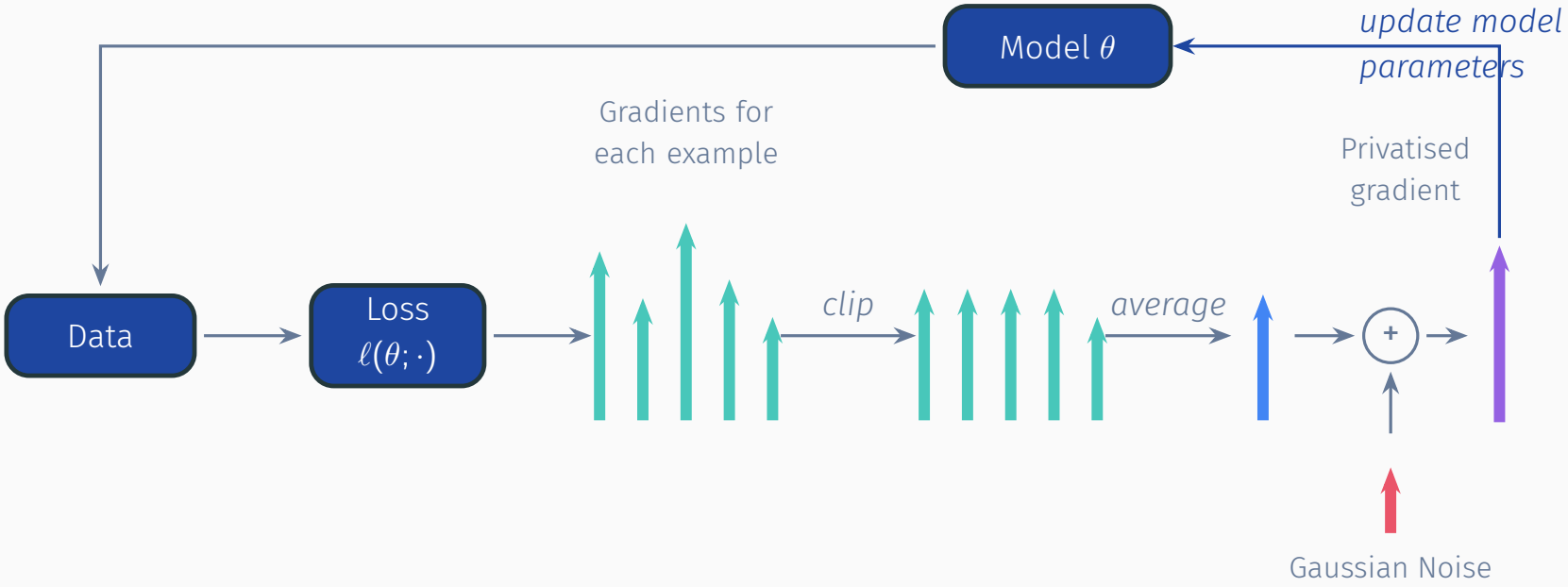
$$\min_{\theta} \quad \text{training loss}(\theta)$$
$$\text{s.t.} \quad \underbrace{\Gamma(\theta)}_{\text{A rate function}} \leq \gamma$$

Can we solve the general form with differential privacy?

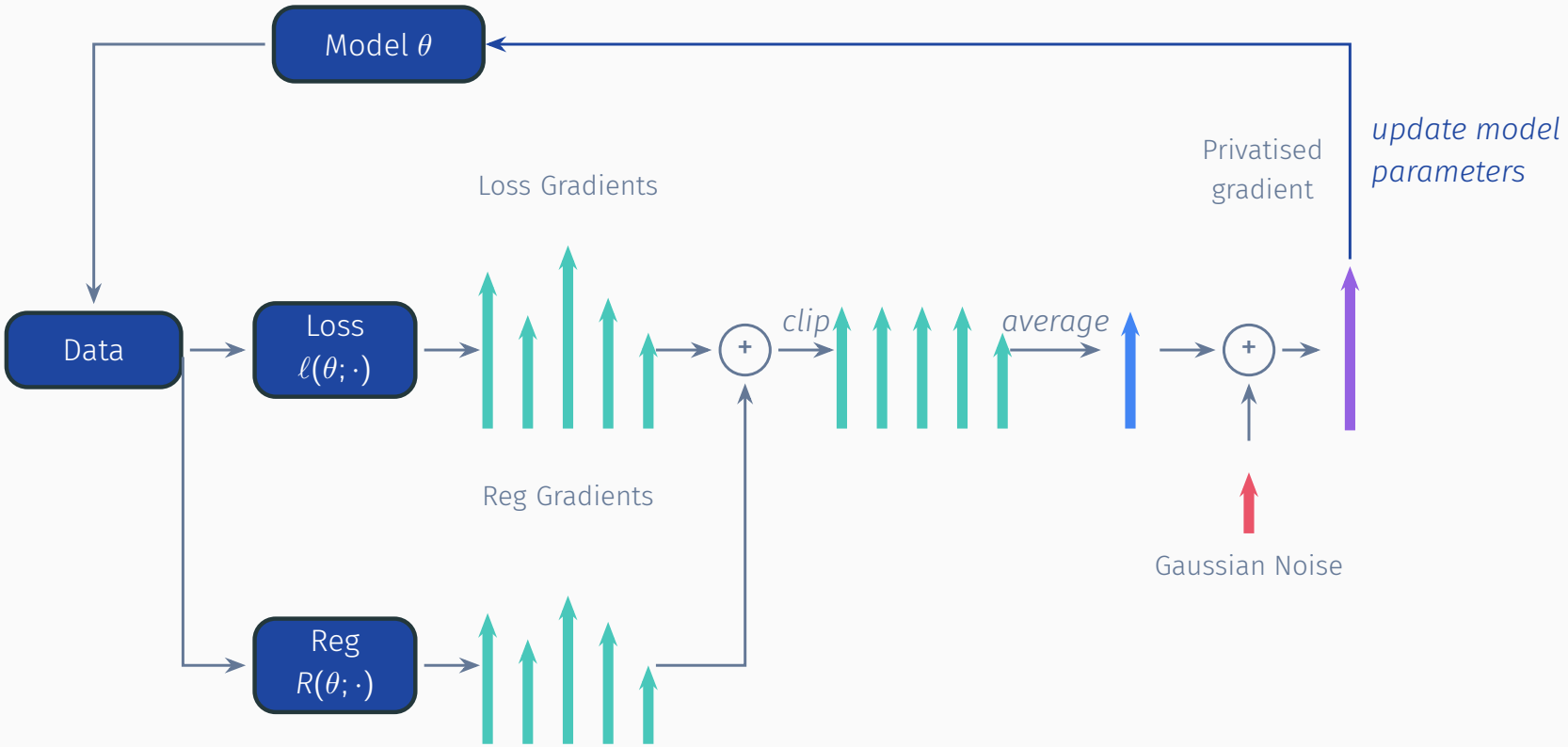
Private Training of Machine Learning Models

In Differential Privacy, we seek to **limit the detectability of the individual's contribution**.

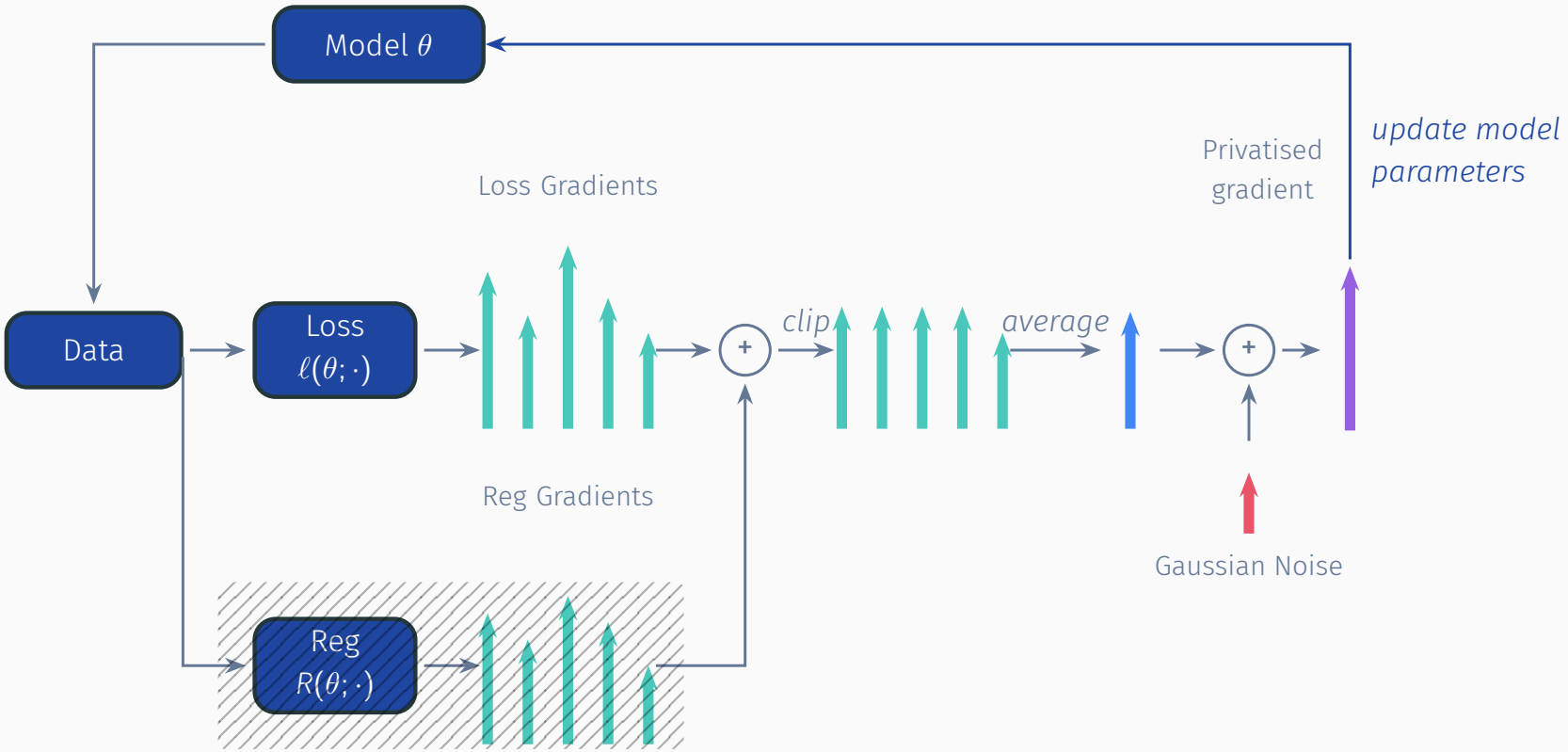
In training machine learning models, these contributions are gradients.



Making Constrained Optimization Private Is Not So Easy



Making Constrained Optimization Private Is Not So Easy



The Challenge: Arbitrary Constraints Break Per-Sample Privacy

Standard DP-SGD

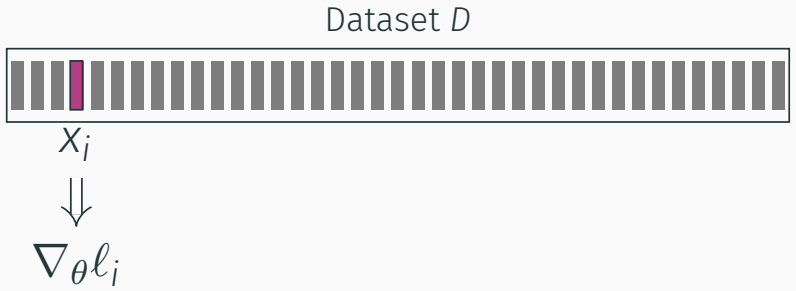
$$\min_{\theta} \frac{1}{n} \sum_{i} \ell(\theta; x_i) \qquad \text{(Loss)}$$

Dataset D

The Challenge: Arbitrary Constraints Break Per-Sample Privacy

Standard DP-SGD

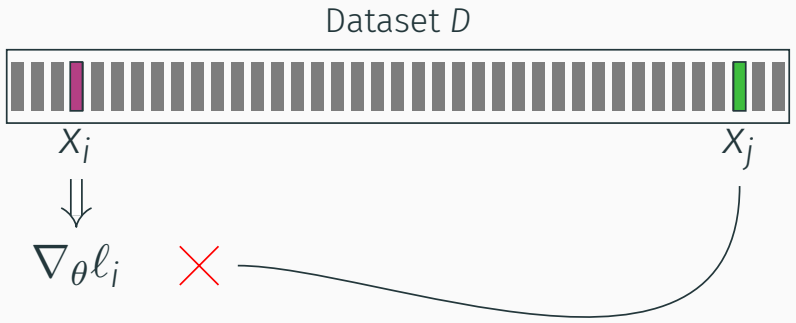
$$\min_{\theta} \frac{1}{n} \sum_{i} \ell(\theta; x_i) \qquad \text{(Loss)}$$



The Challenge: Arbitrary Constraints Break Per-Sample Privacy

Standard DP-SGD

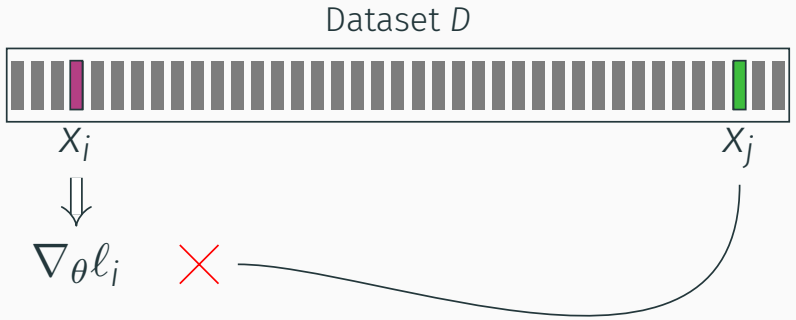
$$\min_{\theta} \frac{1}{n} \sum_{i} \ell(\theta; x_i) \qquad \text{(Loss)}$$



The Challenge: Arbitrary Constraints Break Per-Sample Privacy

Standard DP-SGD

$$\min_{\theta} \frac{1}{n} \sum_{i} \ell(\theta; x_i) \qquad \text{(Loss)}$$

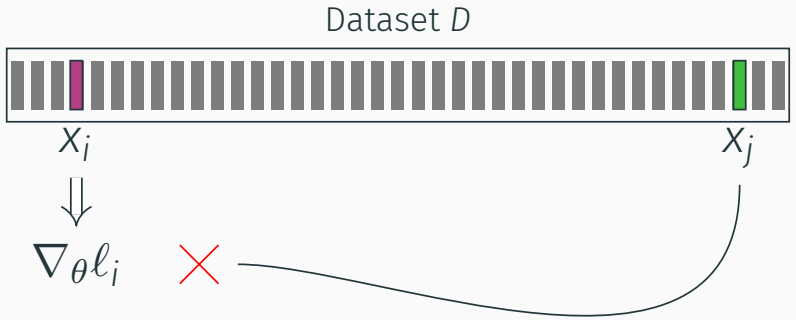


Each sample contributes **1 term**

The Challenge: Arbitrary Constraints Break Per-Sample Privacy

Standard DP-SGD

$$\min_{\theta} \frac{1}{n} \sum_{i} \ell(\theta; x_i) \qquad \text{(Loss)}$$

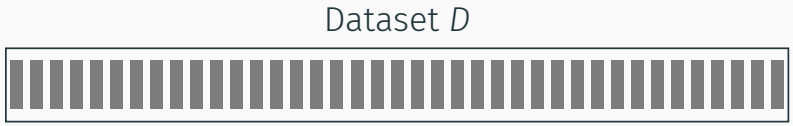


Each sample contributes **1 term**

With Arbitrary Constraints

$$\min_{\theta} \frac{1}{n} \sum_{i} \ell_i + \lambda(\Gamma(\theta, D) - \gamma)$$

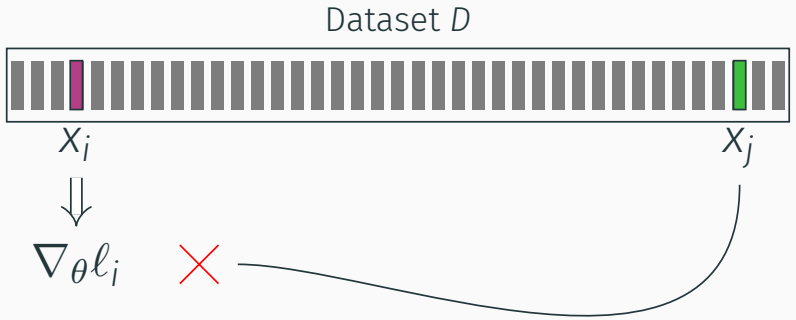
(Lagrangian)



The Challenge: Arbitrary Constraints Break Per-Sample Privacy

Standard DP-SGD

$$\min_\theta \frac{1}{n} \sum_i \ell(\theta; x_i) \qquad \text{(Loss)}$$

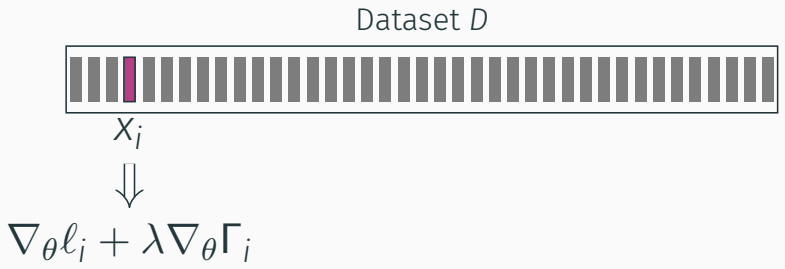


Each sample contributes **1 term**

With Arbitrary Constraints

$$\min_\theta \frac{1}{n} \sum_i \ell_i + \lambda(\Gamma(\theta, D) - \gamma)$$

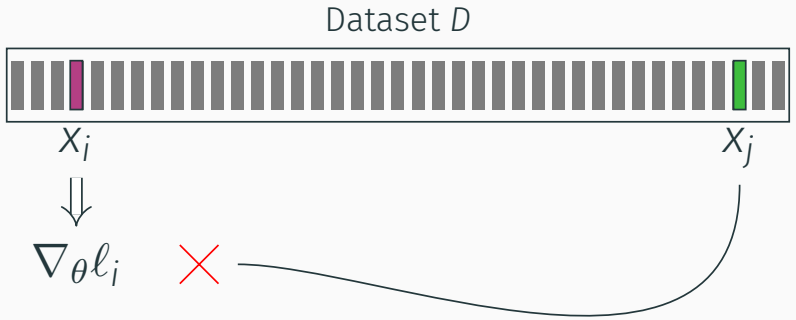
(Lagrangian)



The Challenge: Arbitrary Constraints Break Per-Sample Privacy

Standard DP-SGD

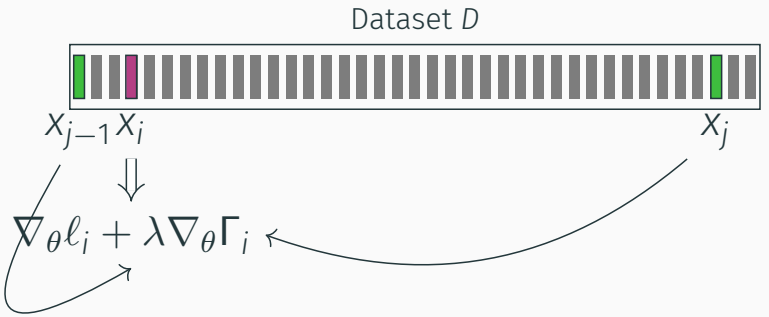
$$\min_\theta \frac{1}{n} \sum_i \ell(\theta; x_i) \qquad \text{(Loss)}$$



With Arbitrary Constraints

$$\min_\theta \frac{1}{n} \sum_i \ell_i + \lambda(\Gamma(\theta, D) - \gamma)$$

(Lagrangian)

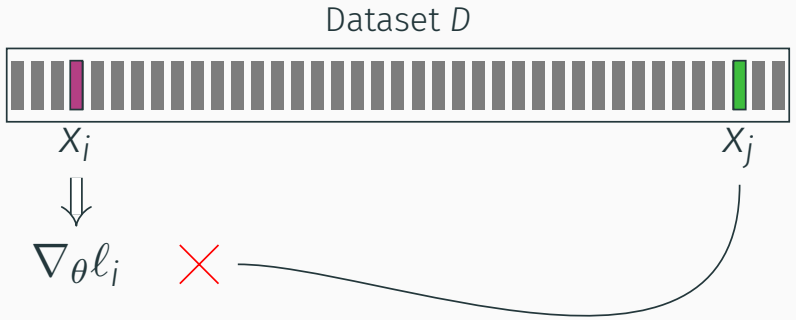


Each sample contributes **1 term**

The Challenge: Arbitrary Constraints Break Per-Sample Privacy

Standard DP-SGD

$$\min_\theta \frac{1}{n} \sum_i \ell(\theta; x_i) \qquad \text{(Loss)}$$

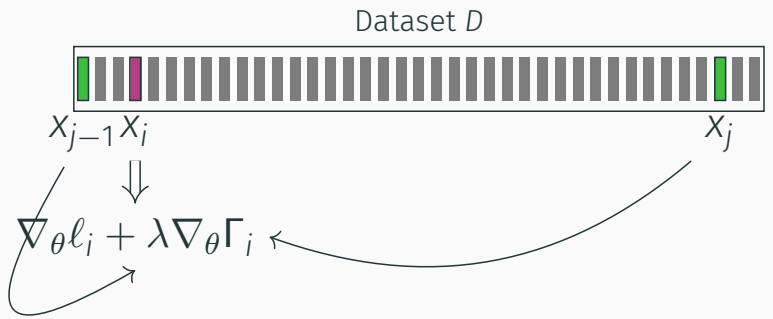


Each sample contributes **1 term**

With Arbitrary Constraints

$$\min_\theta \frac{1}{n} \sum_i \ell_i + \lambda(\Gamma(\theta, D) - \gamma)$$

(Lagrangian)

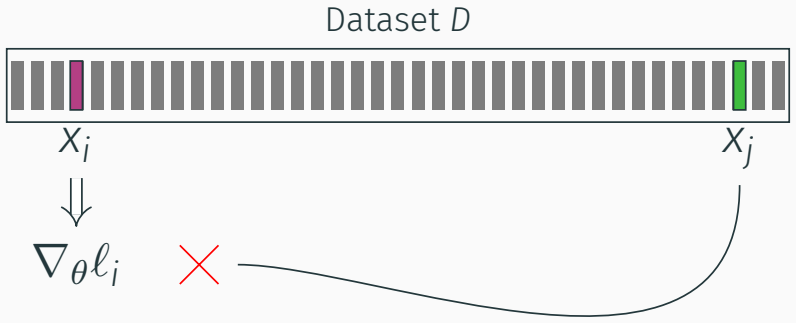


Each sample contributes up to $|D| + 1$ **terms**

The Challenge: Arbitrary Constraints Break Per-Sample Privacy

Standard DP-SGD

$$\min_\theta \frac{1}{n} \sum_i \ell(\theta; x_i) \qquad \text{(Loss)}$$

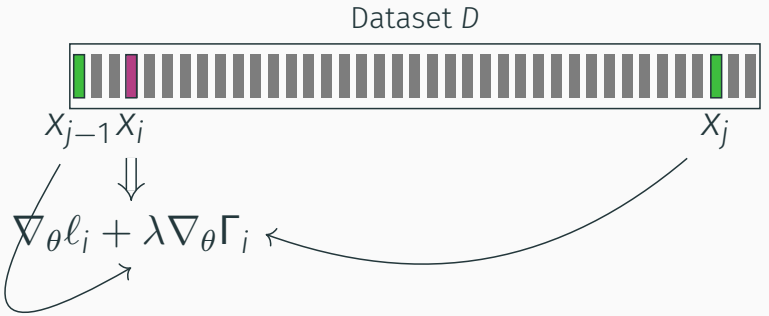


Each sample contributes **1 term**

With Arbitrary Constraints

$$\min_\theta \frac{1}{n} \sum_i \ell_i + \lambda(\Gamma(\theta, D) - \gamma)$$

(Lagrangian)



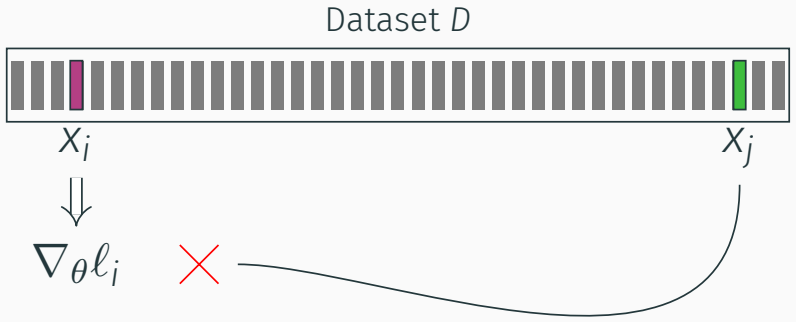
Each sample contributes up to $|D| + 1$ **terms**

Why does this matter? **Utility**!

The Challenge: Arbitrary Constraints Break Per-Sample Privacy

Standard DP-SGD

$$\min_\theta \frac{1}{n} \sum_i \ell(\theta; x_i) \qquad \text{(Loss)}$$

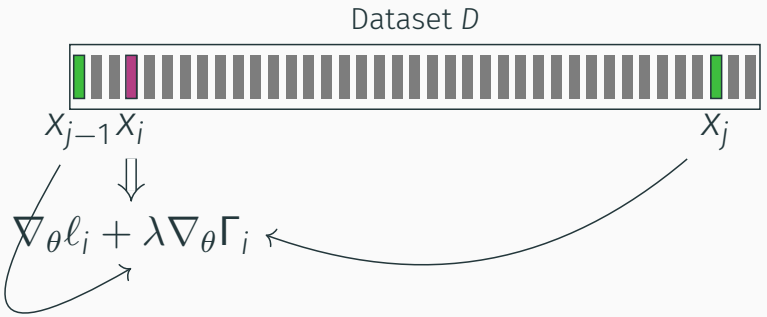


Each sample contributes **1 term**

With Arbitrary Constraints

$$\min_\theta \frac{1}{n} \sum_i \ell_i + \lambda(\Gamma(\theta, D) - \gamma)$$

(Lagrangian)



Each sample contributes up to $|D| + 1$ **terms**

Why does this matter? **Utility**! $\uparrow$ # per-sample contributions $\Rightarrow \uparrow$ noise added $\Rightarrow \downarrow$ utility

We Can Take Advantage of the Constraint Structure

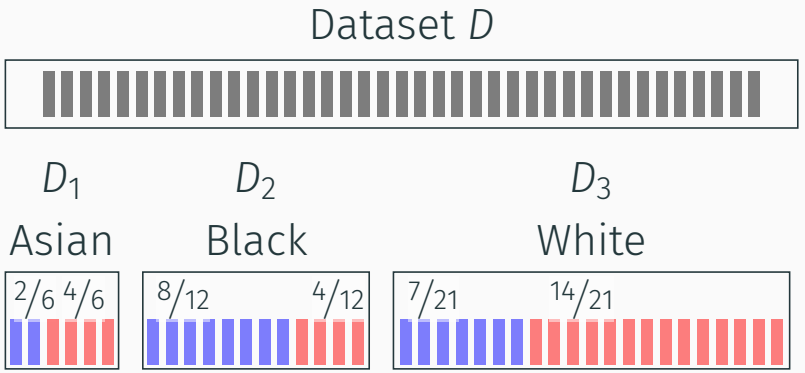
Demographic Parity:
prediction rate for different demographics should not be too different

$$\Gamma = P_k(D[Z = z]) - P_k(D[Z \neq z]) \leq \gamma$$

$k \in \{+\text{ve}, -\text{ve}\}$

$z \in \{\text{Asian}, \text{Black}, \text{White}\}$

$P_k(D[Z = z])$ is the prediction rate for class k over subpopulation z.



We Can Take Advantage of the Constraint Structure

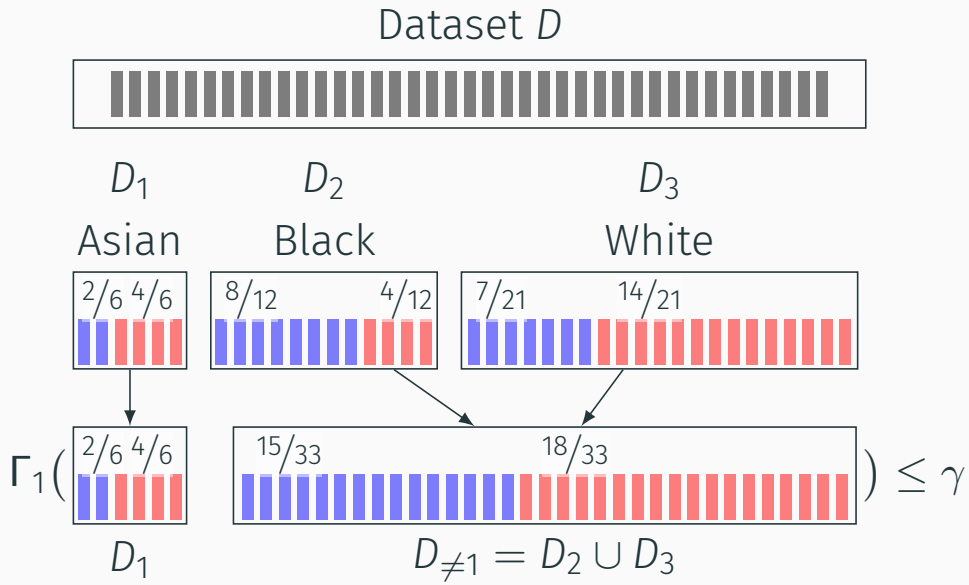
Demographic Parity:
prediction rate for different demographics should not be too different

$$\Gamma = P_k(D[Z = z]) - P_k(D[Z \neq z]) \leq \gamma$$

$$k \in \{+\text{ve}, -\text{ve}\}$$

$$z \in \{\text{Asian}, \text{Black}, \text{White}\}$$

$P_k(D[Z = z])$ is the prediction rate for class k over subpopulation z.



We Can Take Advantage of the Constraint Structure

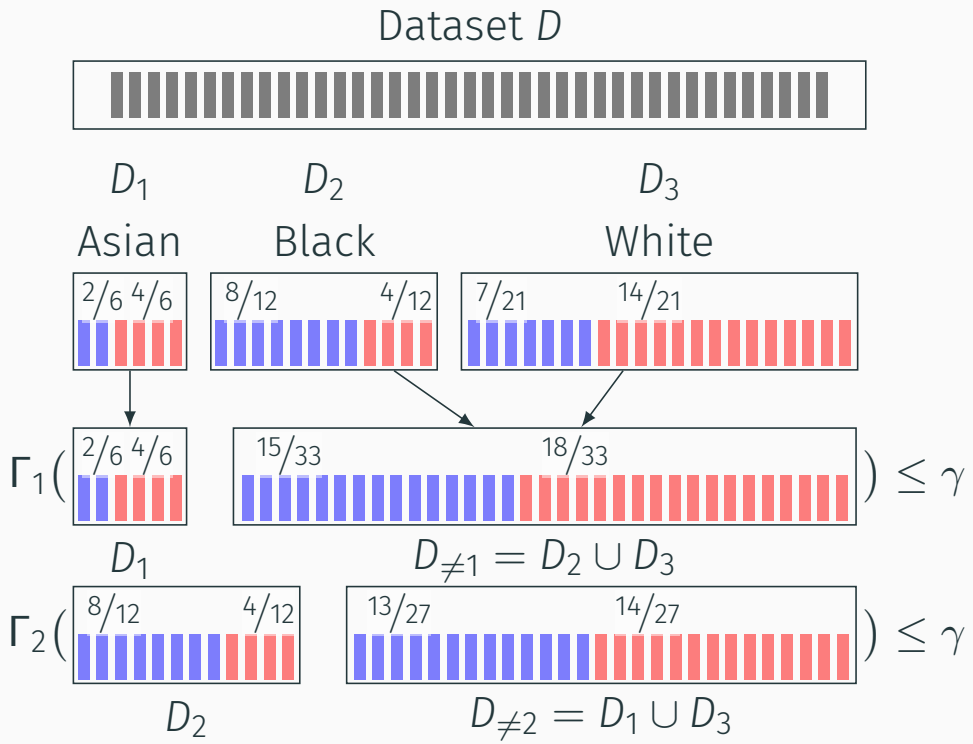
Demographic Parity:
prediction rate for different demographics should not be too different

$$\Gamma = P_k(D[Z = z]) - P_k(D[Z \neq z]) \leq \gamma$$

$$k \in \{+\text{ve}, -\text{ve}\}$$

$$z \in \{\text{Asian}, \text{Black}, \text{White}\}$$

$P_k(D[Z = z])$ is the prediction rate for class k over subpopulation z.



We Can Take Advantage of the Constraint Structure

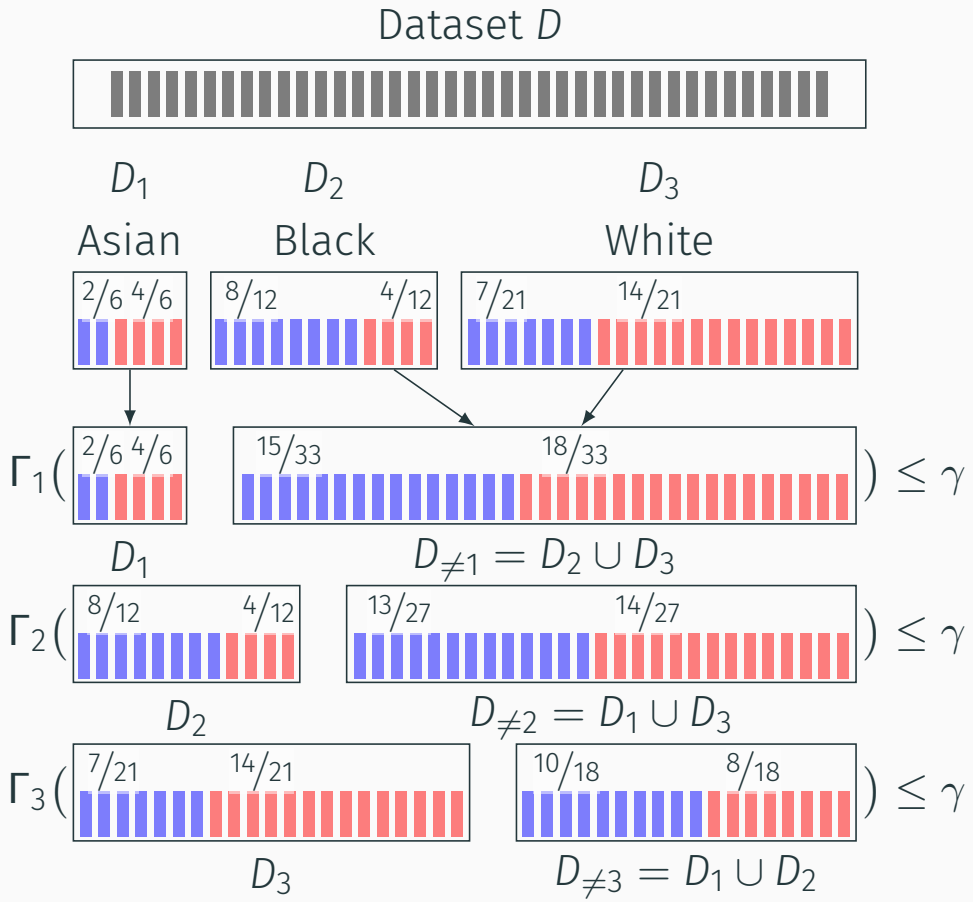
Demographic Parity:
prediction rate for different demographics should not be too different

$$\Gamma = P_k(D[Z = z]) - P_k(D[Z \neq z]) \leq \gamma$$

$$k \in \{+\text{ve}, -\text{ve}\}$$

$$z \in \{\text{Asian}, \text{Black}, \text{White}\}$$

$P_k(D[Z = z])$ is the prediction rate for class k over subpopulation z.



We Can Take Advantage of the Constraint Structure

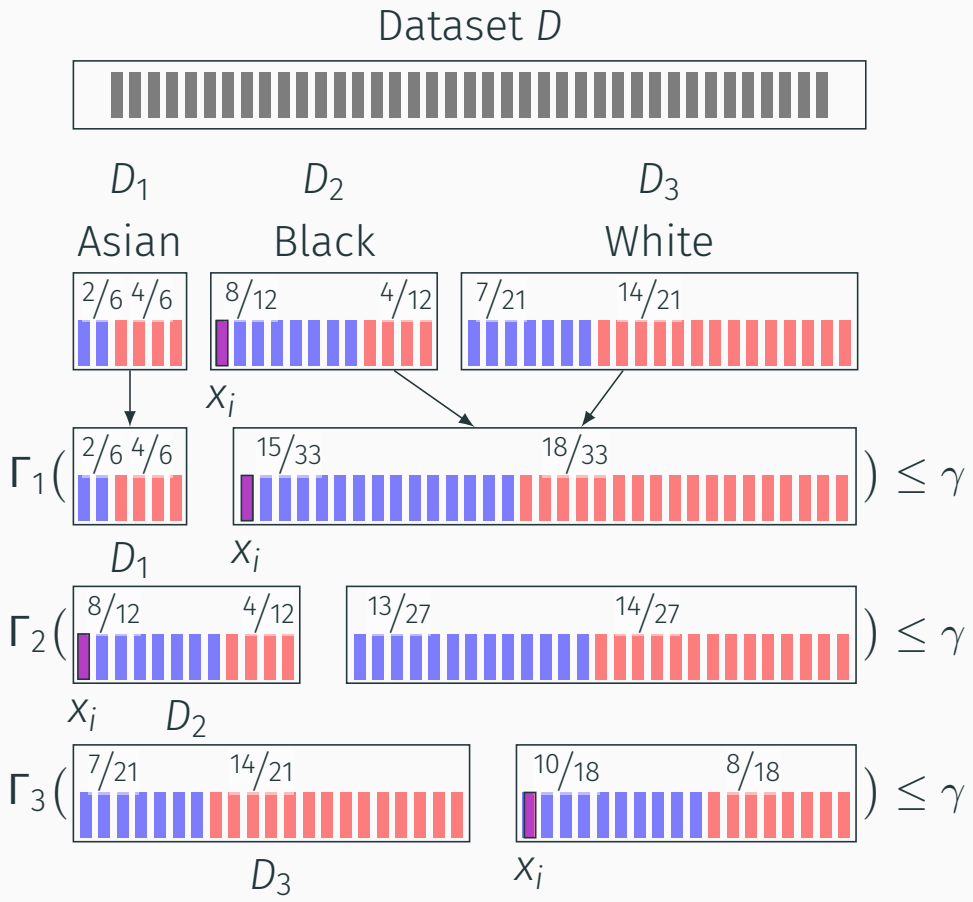
Demographic Parity:
prediction rate for different demographics should not be too different

$$\Gamma = P_k(D[Z = z]) - P_k(D[Z \neq z]) \leq \gamma$$

$$k \in \{+\text{ve}, -\text{ve}\}$$

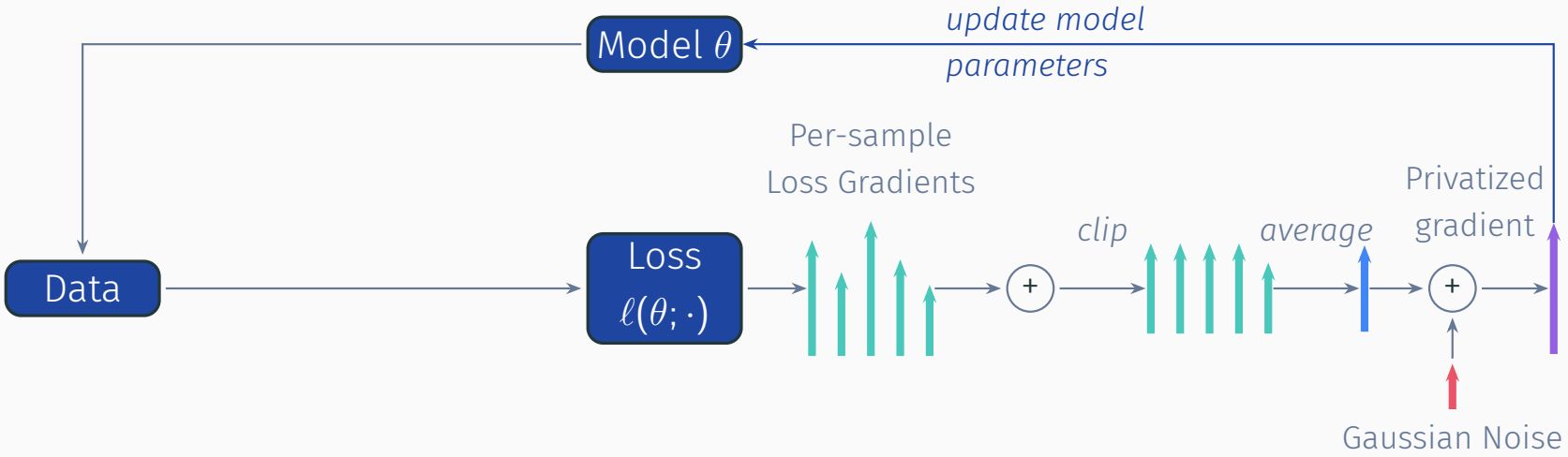
$$z \in \{\text{Asian}, \text{Black}, \text{White}\}$$

$P_k(D[Z = z])$ is the prediction rate for class k over subpopulation z.

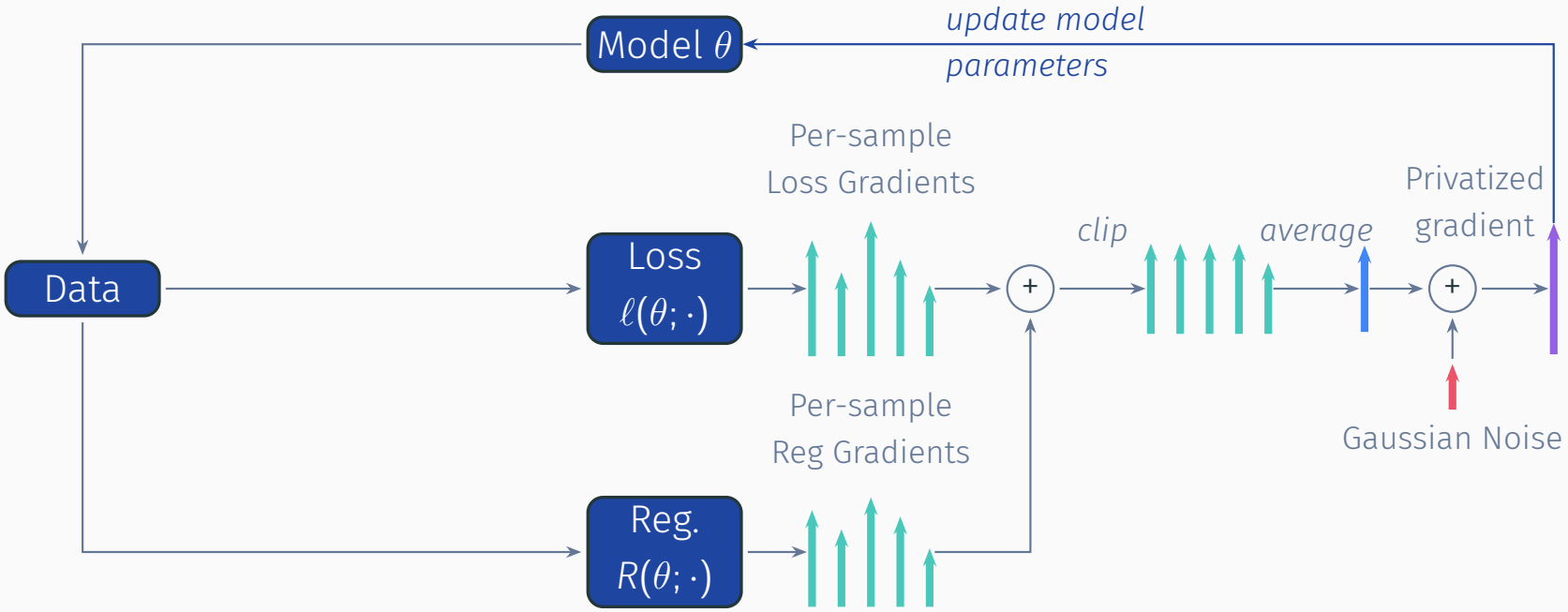


Every sample contributes exactly 3+1 terms to the loss.

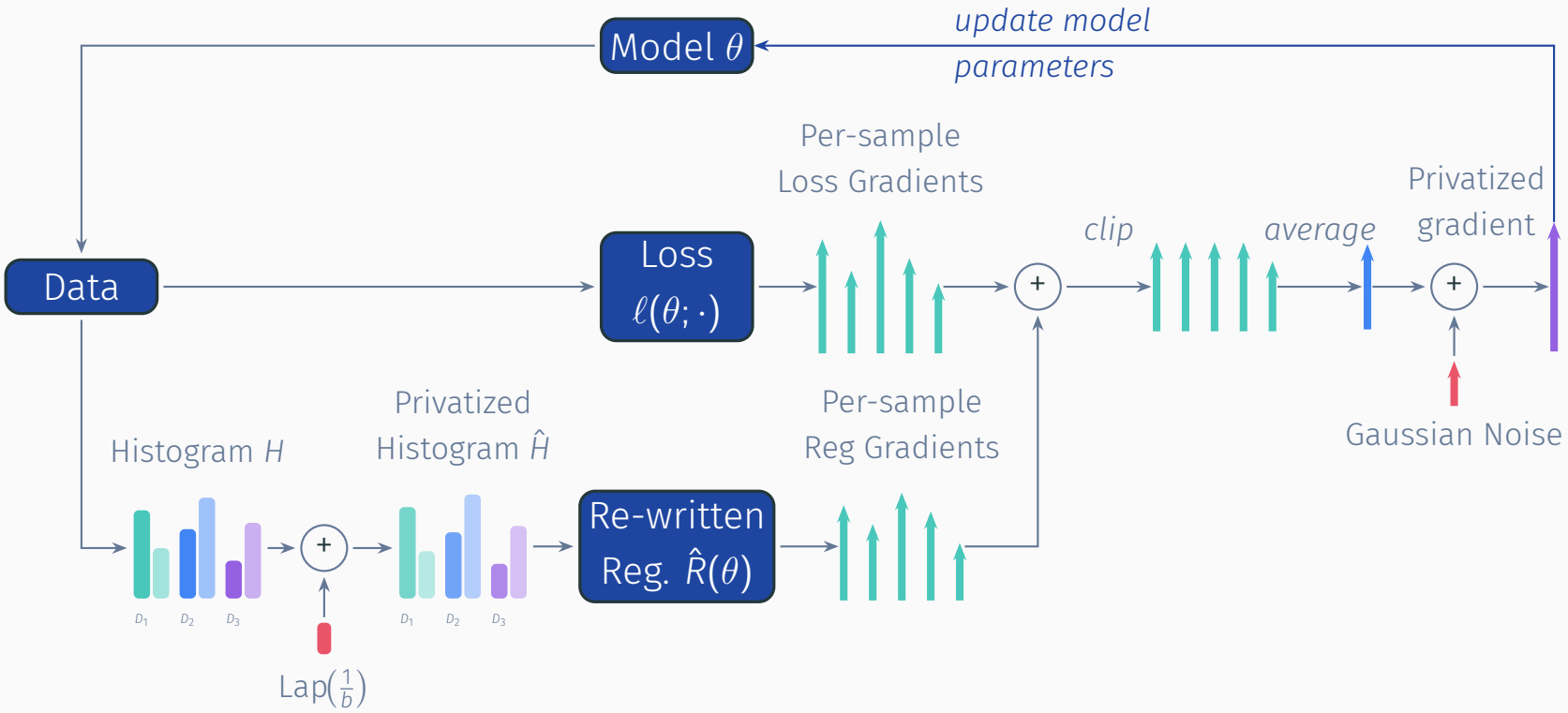
RaCO-DP: Private Rate-Constrained Optimization



RaCO-DP: Private Rate-Constrained Optimization



RaCO-DP: Private Rate-Constrained Optimization



The histogram allows us to derive per-sample gradients for the regularizer.

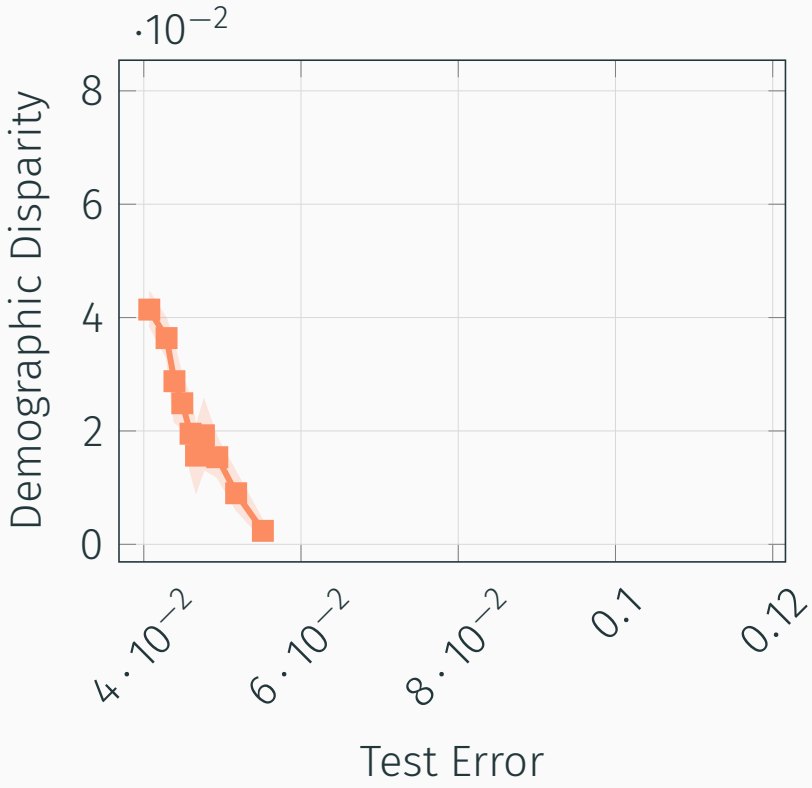


Figure 1: `Parkinsons` dataset. Logistic Regressions models trained with $\varepsilon = 1$

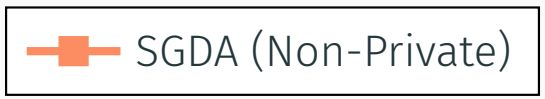
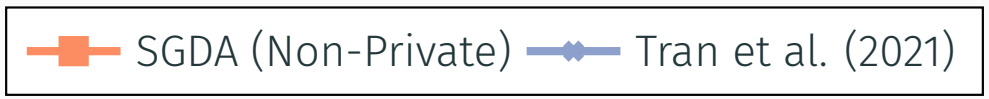


Figure 1: `Parkinsons` dataset. Logistic Regressions models trained with $\varepsilon = 1$



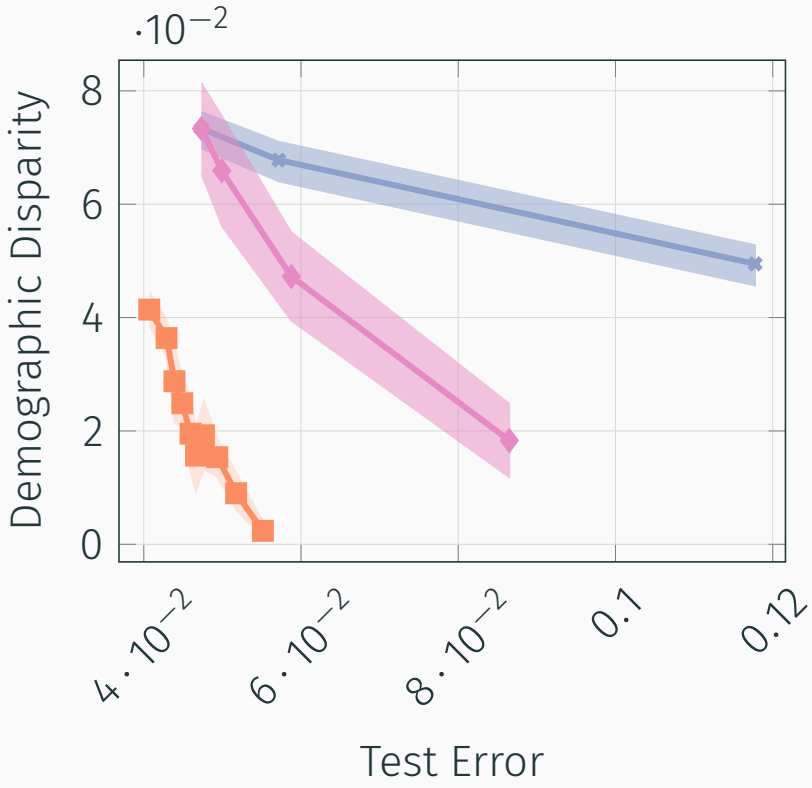


Figure 1: `Parkinsons` dataset. Logistic Regressions models trained with $\varepsilon = 1$

Figure 1: `Parkinsons` dataset. Logistic Regressions models trained with $\varepsilon = 1$

Figure 1: `Parkinsons` dataset. Logistic Regressions models trained with $\varepsilon = 1$

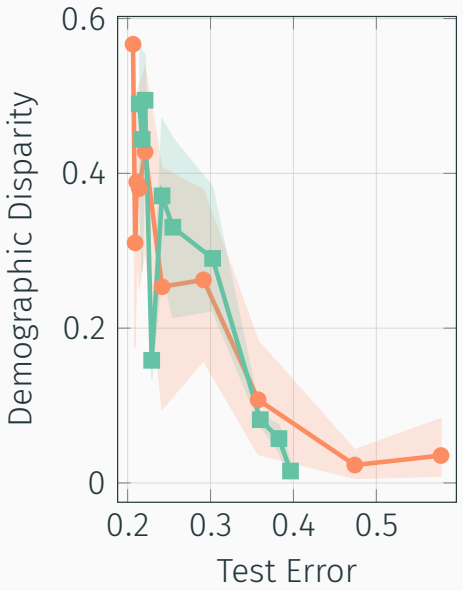


Figure 2: `ACSEmployment` dataset with 18 constraints.

Figure 1: `Parkinsons` dataset. Logistic Regressions models trained with $\varepsilon = 1$

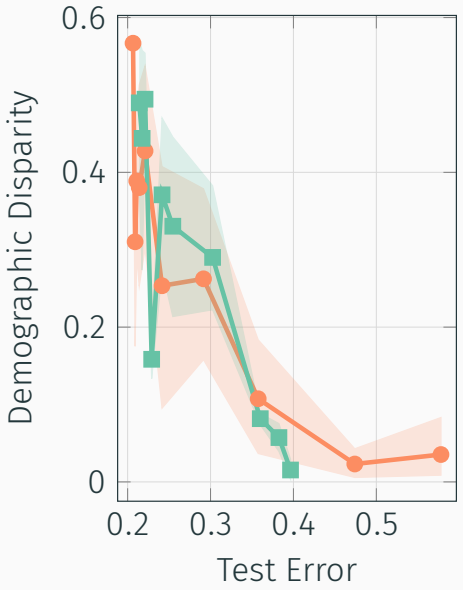


Figure 2: `ACSEmployment` dataset with 18 constraints.

SGDA (Non-Private) | Tran et al. (2021) | DP-FERMI (2023) | RaCO-DP

On tabular data, RaCO-DP Pareto dominates prior SOTA and nearly closes the optimality gap with non-private models

RaCO-DP Scales to Deep Learning Models

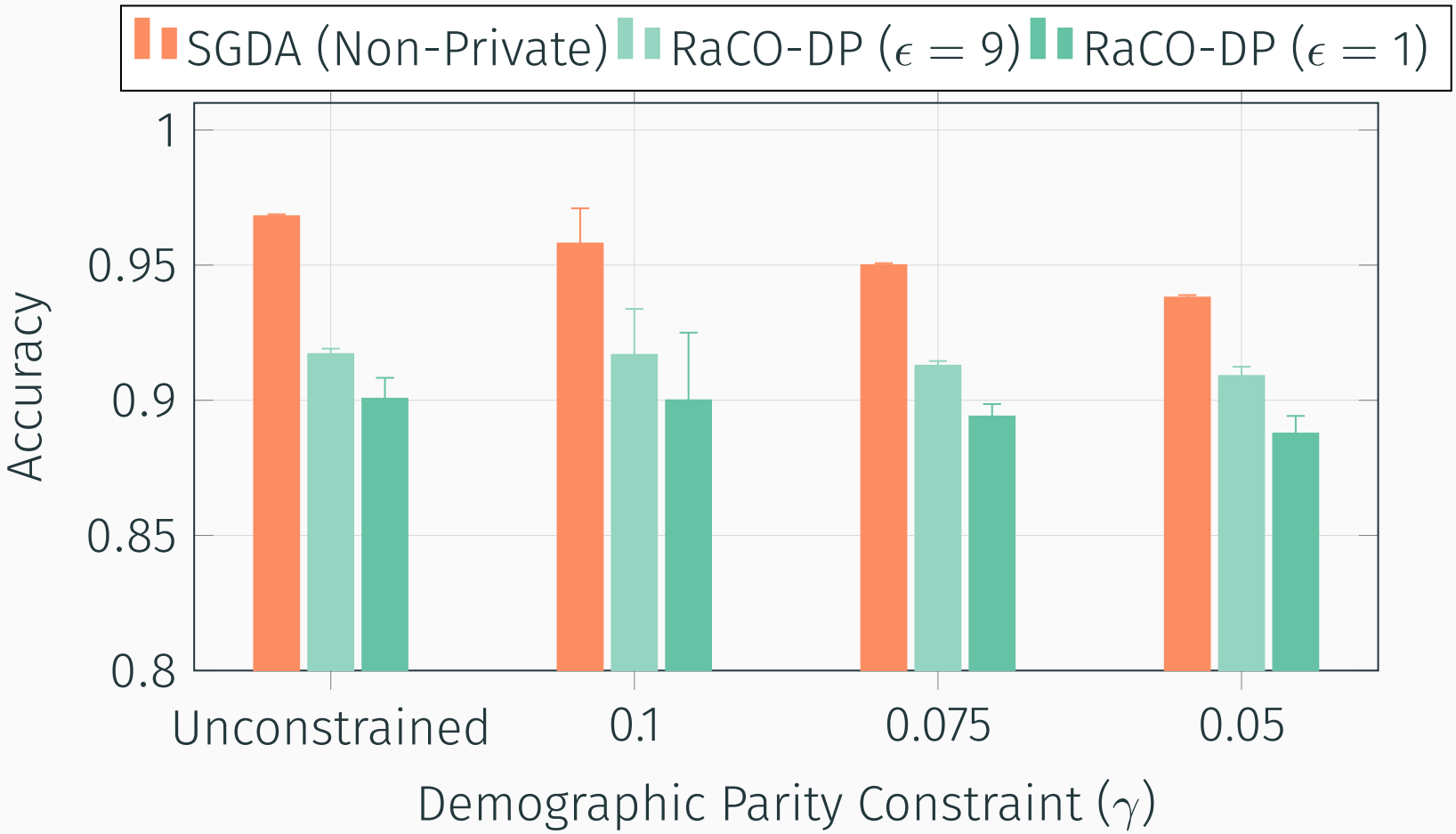


Figure 3: `CelebA` using a ResNet-16 model. Standard deviation error bars over 5 runs.

RaCO-DP Ensures Constraints Are Satisfied

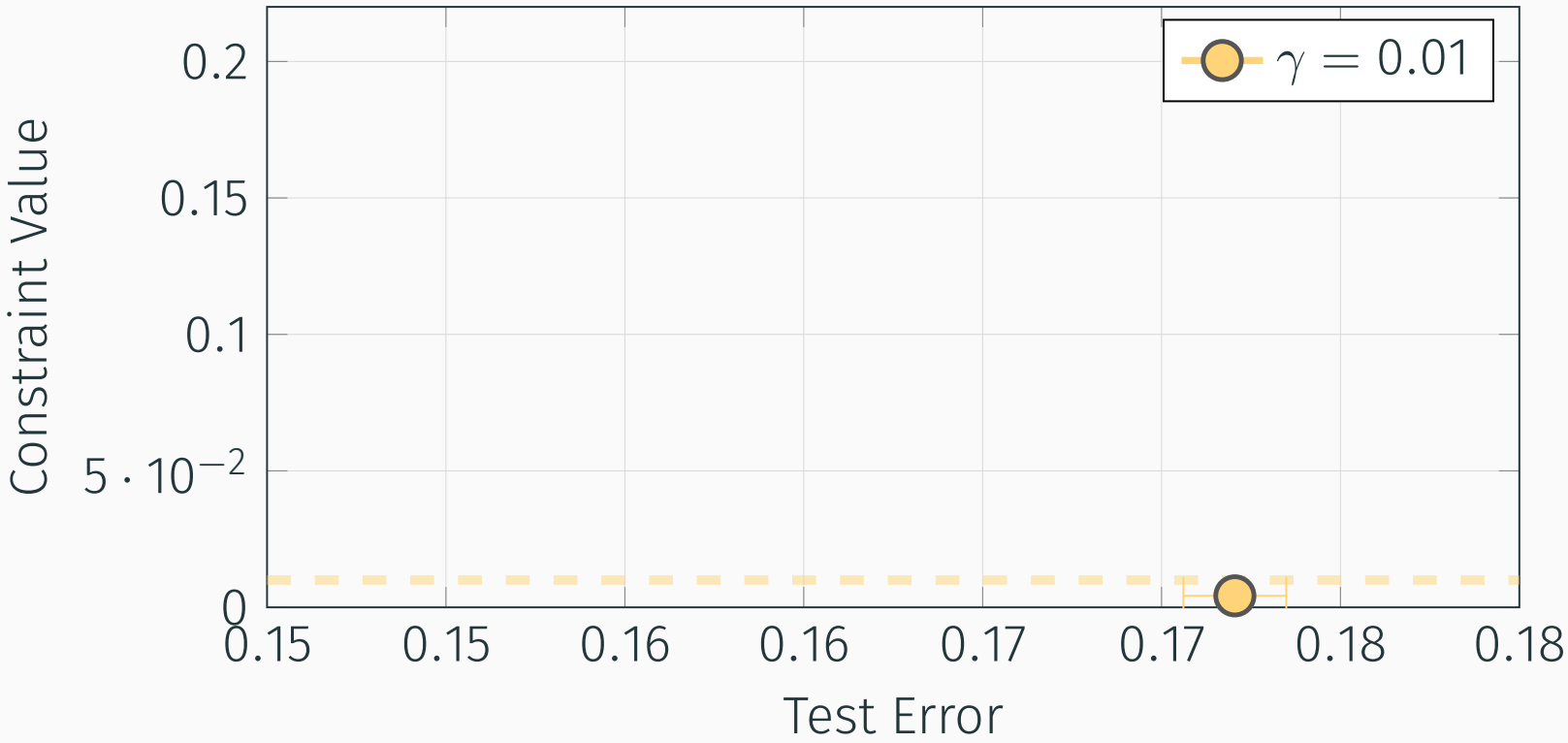


Figure 4: Satisfiability on `Adult`. LR models trained with $\varepsilon = 1$. Target values γ (dashed lines), averaged over 20 runs.

RaCO-DP Ensures Constraints Are Satisfied

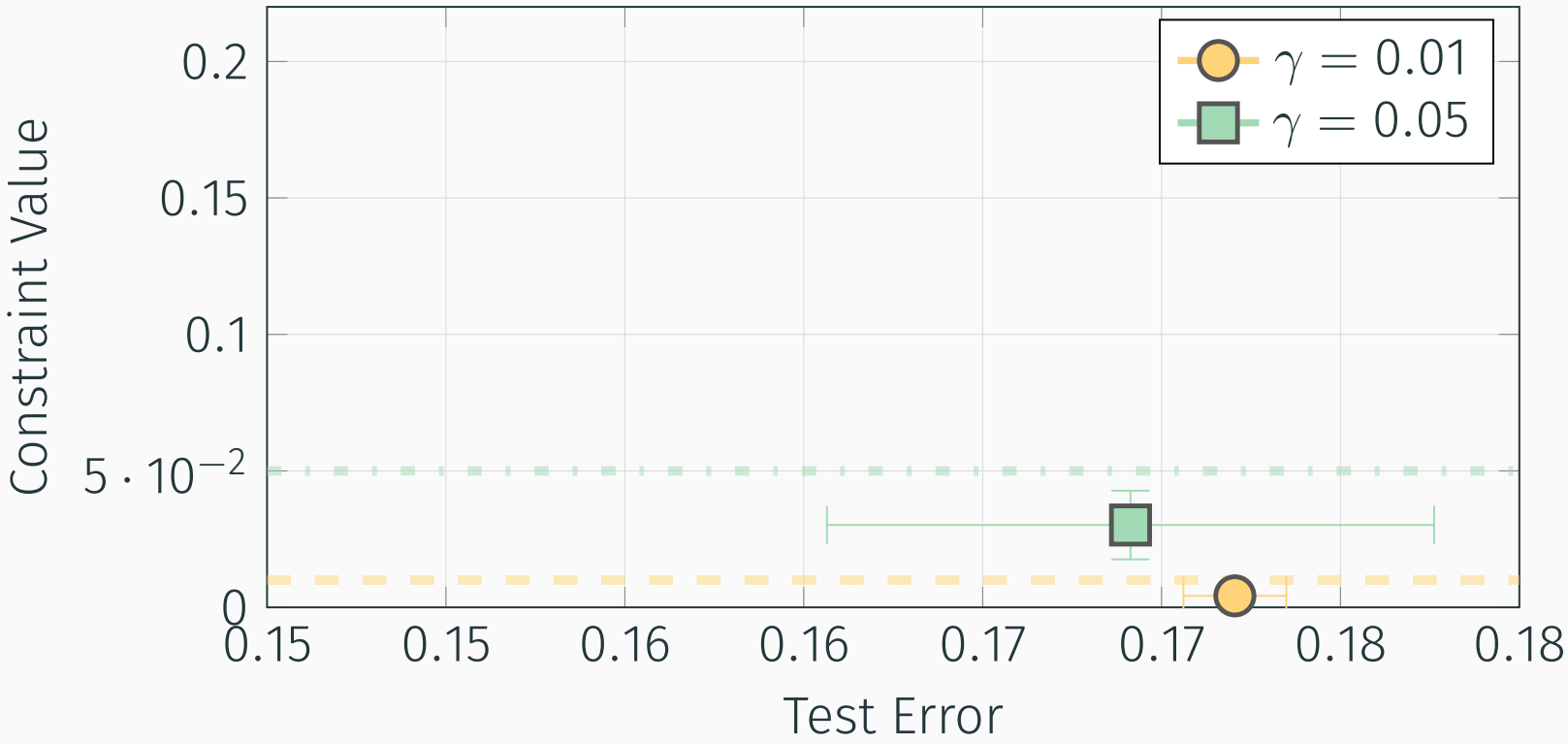


Figure 4: Satisfiability on `Adult`. LR models trained with $\varepsilon = 1$. Target values γ (dashed lines), averaged over 20 runs.

RaCO-DP Ensures Constraints Are Satisfied

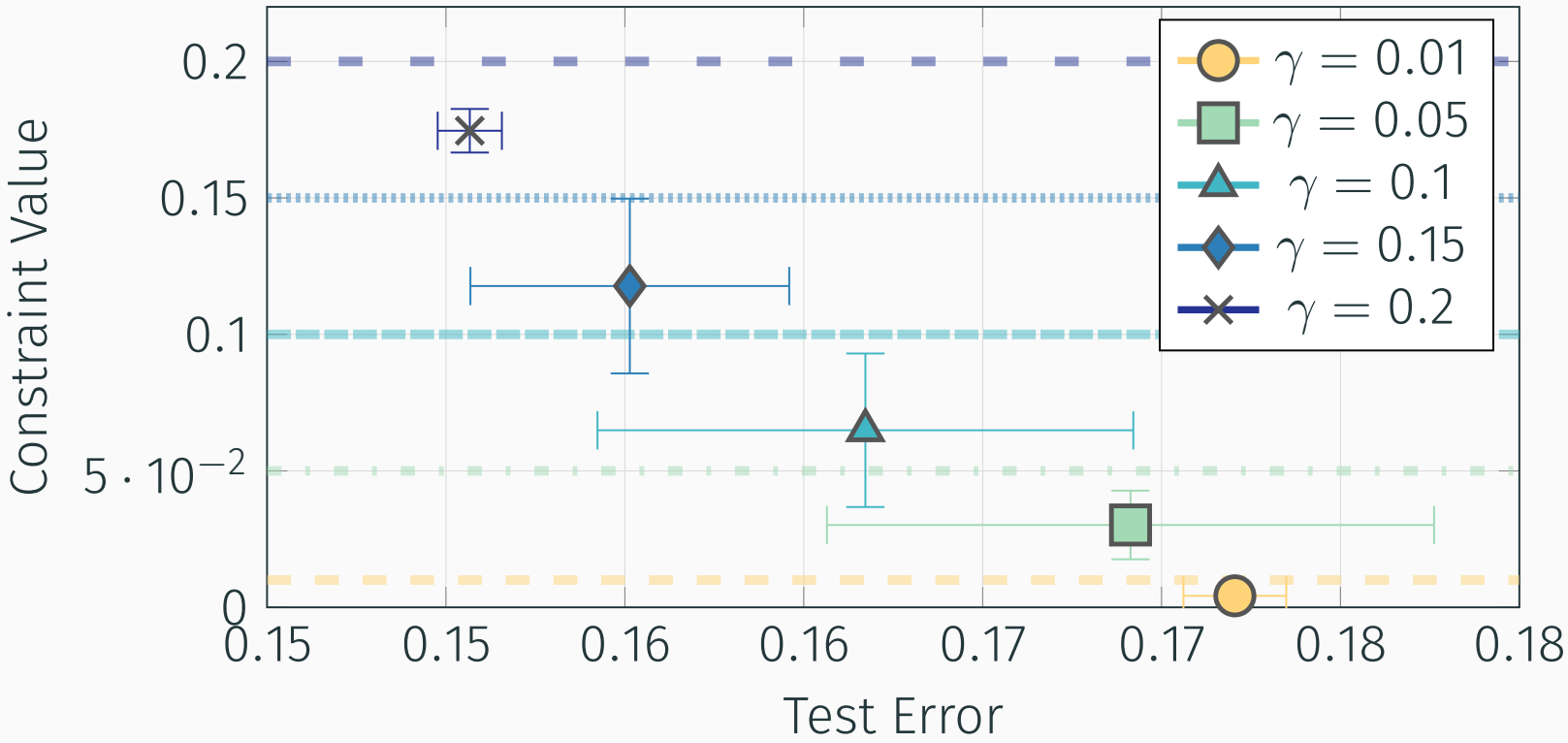


Figure 4: Satisfiability on `Adult`. LR models trained with $\varepsilon = 1$. Target values γ (dashed lines), averaged over 20 runs.