Distributed Learning Without Distress: Privacy-Preserving Empirical Risk Minimization

Bargav Jayaraman[1], Lingxiao Wang[2], David Evans[1] and Quanquan Gu[2]

[1] Department of Computer Science, University of Virginia, Charlottesville, Virginia USA
[2] Department of Computer Science, UCLA, Los Angeles, California USA

Accepted at NeurIPS 2018, Montreal Canada

Background on Empirical Risk Minimization

Given the following convex objective function:

$$J(\theta) = \frac{1}{n} \sum_{i=1}^{n} \ell(\theta, x_i, y_i) + \lambda N(\theta)$$

Find θ that minimizes the objective function:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \; J(\theta)$$

Background on Empirical Risk Minimization

Given the following convex objective function:

$$J(\theta) = \frac{1}{n} \sum_{i=1}^{n} \underbrace{\ell(\theta, x_i, y_i)}_{\text{convex loss function}} + \lambda N(\theta)$$

Logistic loss

$(\log(1+e^{-x_i^T\theta y_i}))$

Quadratic loss

$(\frac{1}{2}(x_i^T\theta - y_i)^2)$

Find θ that minimizes the objective function:

$$\hat{\theta} = \operatorname*{argmin}_{\theta} J(\theta)$$

Background on Empirical Risk Minimization

Given the following convex objective function:

$$J(\theta) = \frac{1}{n}\sum_{i=1}^{n} \underbrace{\ell(\theta, x_i, y_i)}_{\text{convex loss function}} + \lambda \underbrace{N(\theta)}_{\text{Regularization term}}$$

Regularization term
- L1 norm ($|\theta|_1$)
- L2 norm ($\frac{1}{2}\|\theta\|_2^2$)

convex loss function
- Logistic loss ($\log(1+e^{-x_i^T\theta y_i})$)
- Quadratic loss ($\frac{1}{2}(x_i^T\theta - y_i)^2$)

Find θ that minimizes the objective function:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\ J(\theta)$$

Background on Differential Privacy

A randomized mechanism M is (ϵ, δ)-DP if for two neighbouring datasets D and D'

$$\frac{\Pr[M(D) \in S]}{\Pr[M(D') \in S]} \leq e^{\epsilon} + \delta$$

Background on Differential Privacy

A randomized mechanism M is (ϵ, δ)-DP if for two neighbouring datasets D and D'

$$\frac{Pr[M(D) \in S]}{Pr[M(D') \in S]} \leq e^{\epsilon} + \delta$$

Given that sensitivity of M is:

$$\Delta M = \max_{D, D'} \| M(D) - M(D') \|$$

We can ensure ϵ-DP if we sample Laplace noise:

$$Lap(b) \quad , \text{ where } \quad b = \frac{\Delta M}{\epsilon}$$

Background on Differential Privacy

A randomized mechanism M is (ϵ, δ)-DP if for two neighbouring datasets D and D'

$$\frac{Pr[M(D) \in S]}{Pr[M(D') \in S]} \leq e^{\epsilon} + \delta$$

Given that sensitivity of M is:

$$\Delta M = \max_{D, D'} \| M(D) - M(D') \|$$

We can ensure ϵ-DP if we sample Laplace noise:

$$Lap(b) \quad , \text{ where } \quad b = \frac{\Delta M}{\epsilon}$$

Example: Logistic Regression

If $D = (X, Y)$ such that $\|x_i\| \leq 1$ and $y_i \in \{-1, 1\}$

If Logistic Regression model M minimizes the following objective function:

$$J(\theta) = \frac{1}{n} \sum_{i=1}^{n} \log(1 + e^{-x_i^T \theta y_i}) + \frac{\lambda}{2} \|\theta\|_2^2$$

Background on Differential Privacy

A randomized mechanism M is (ϵ, δ)-DP if for two neighbouring datasets D and D'

$$\frac{Pr[M(D) \in S]}{Pr[M(D') \in S]} \leq e^{\epsilon} + \delta$$

Given that sensitivity of M is:

$$\Delta M = \max_{D, D'} \| M(D) - M(D') \|$$

We can ensure ϵ-DP if we sample Laplace noise:

$$Lap(b) \quad , \text{ where } \quad b = \frac{\Delta M}{\epsilon}$$

Example: Logistic Regression

If $D = (X, Y)$ such that $\|X_i\| \leq 1$ and $Y_i \in \{-1, 1\}$

If Logistic Regression model M minimizes the following objective function:

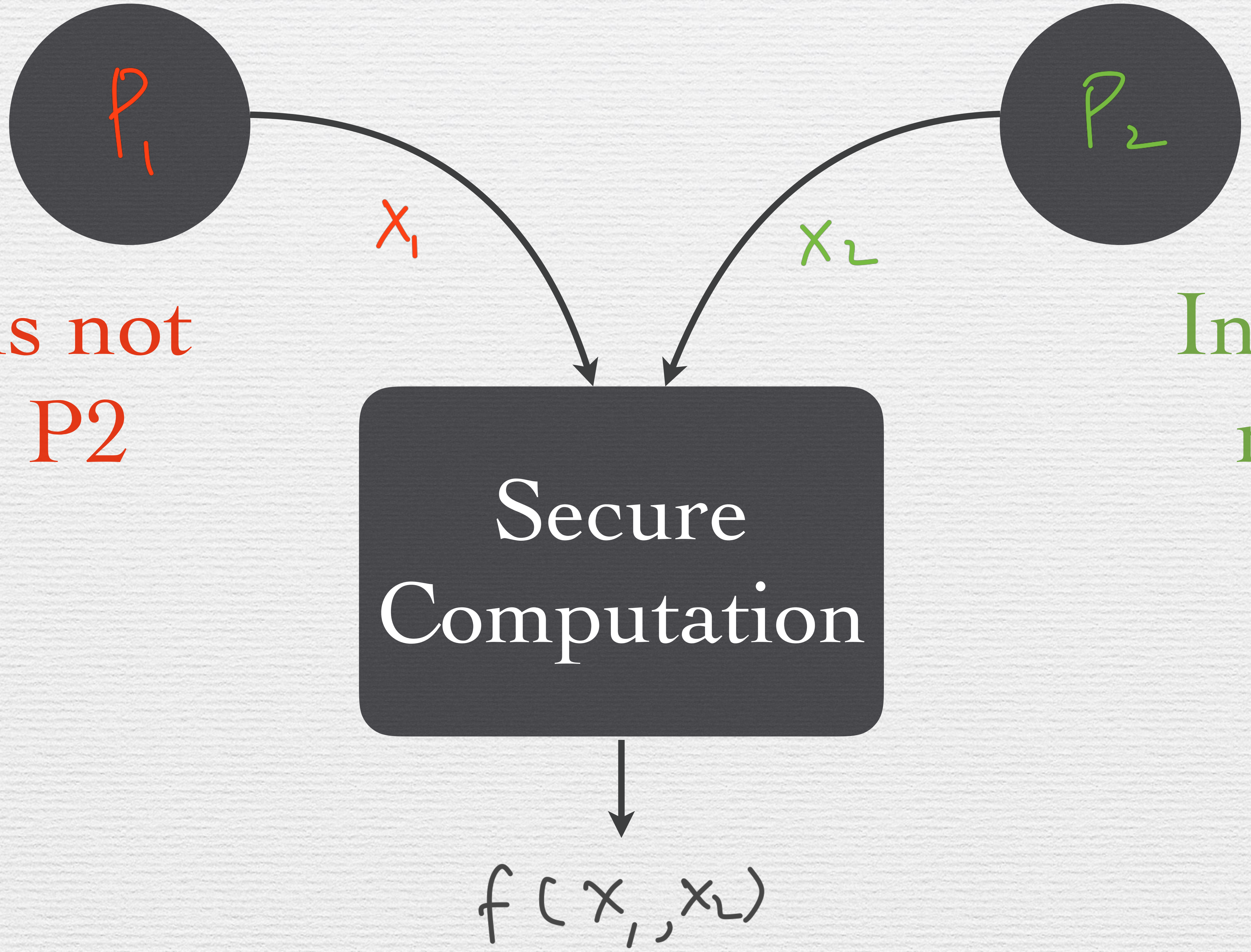
$$J(\theta) = \frac{1}{n} \sum_{i=1}^{n} \log(1 + e^{-X_i^T \theta Y_i}) + \frac{\lambda}{2} \|\theta\|_2^2$$

$$\text{then} \quad \Delta M = \frac{2}{n\lambda}$$

$$\therefore M \text{ is } \epsilon\text{-DP if } \theta \leftarrow \theta^* + Lap\left(\frac{2}{n\lambda\epsilon}\right)$$

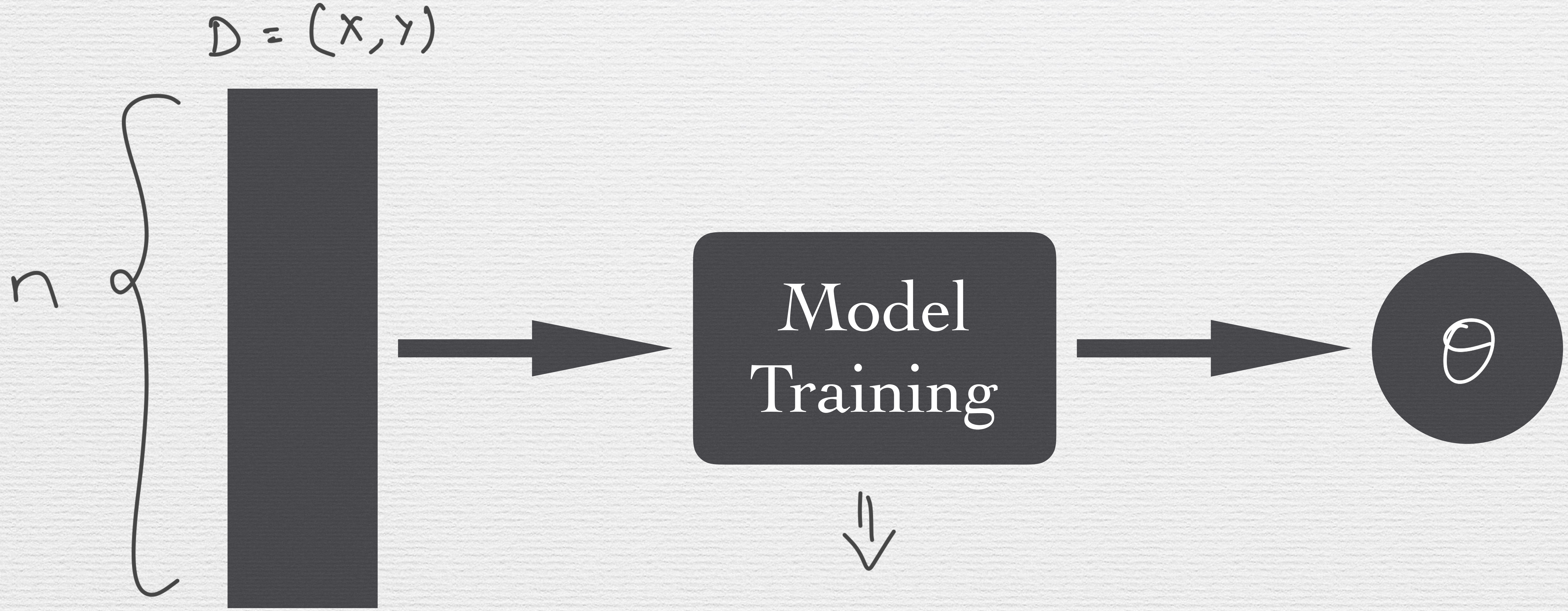
Background on Multi-Party Computation

Input of P1 is not revealed to P2



Input of P2 is not revealed to P1

Differential Private Solutions for Single Party Setting



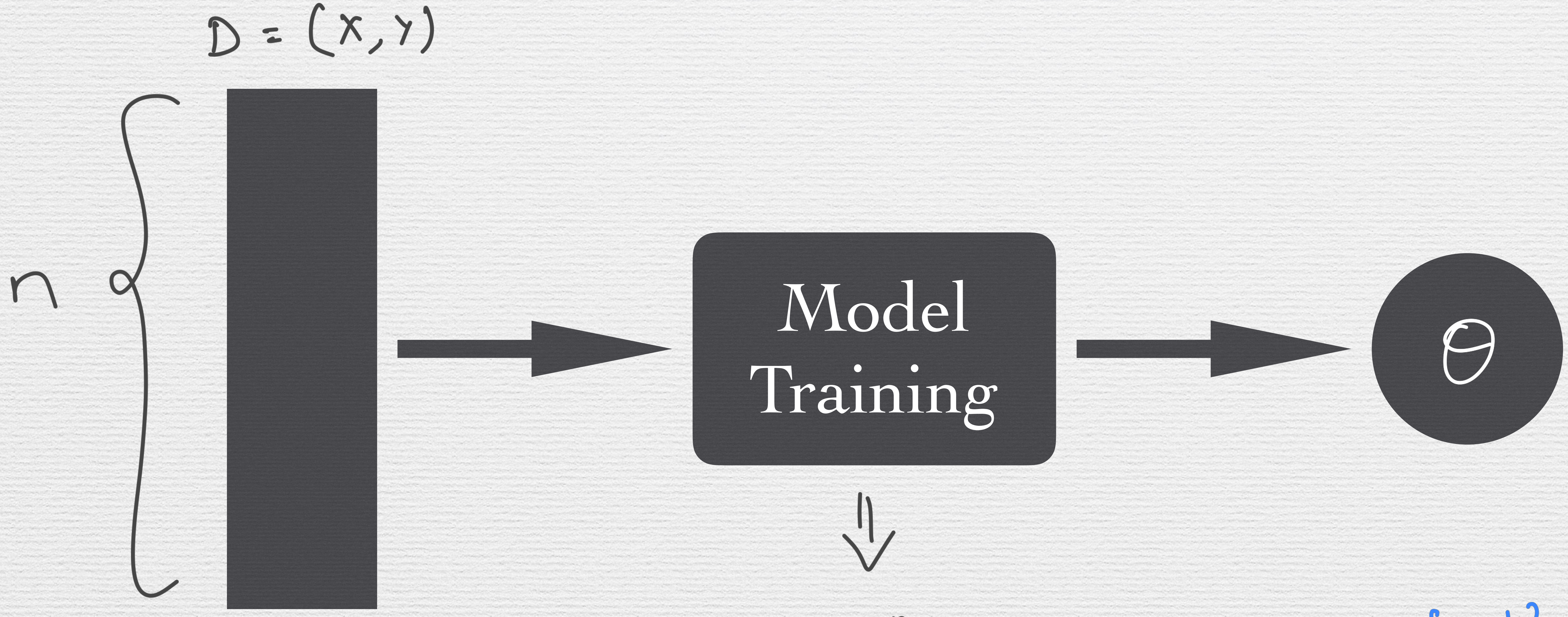
$$J(\theta) = \frac{1}{n} \sum_{i=1}^{n} \ell(\theta, x_i, y_i) + \lambda N(\theta)$$

for t in T:

$\quad \theta \leftarrow \theta - \eta \nabla J(\theta)$

return θ

Differential Private Solutions for Single Party Setting



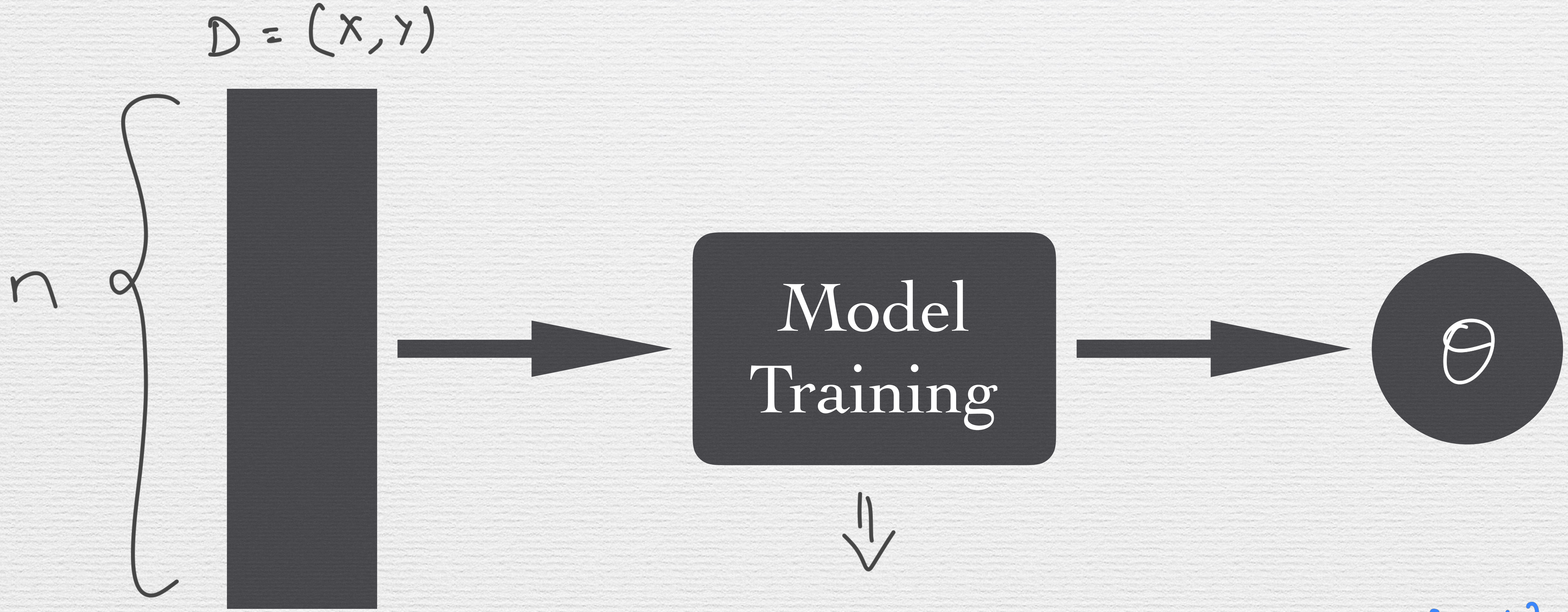
$$\Downarrow$$

$$J(\theta) = \frac{1}{n}\sum_{i=1}^{n} \ell(\theta, x_i, y_i) + \lambda N(\theta) + \beta\{\propto \tfrac{1}{n}\}$$

Chaudhuri et al. (2011)
Objective Perturbation

for t in T:
$\quad \theta \leftarrow \theta - \eta \nabla J(\theta)$
return θ

Differential Private Solutions for Single Party Setting



$$J(\theta) = \frac{1}{n}\sum_{i=1}^{n} \ell(\theta, x_i, y_i) + \lambda N(\theta) + \beta\left\{\propto \frac{1}{n}\right\}$$

Chaudhuri et al. (2011)
Objective Perturbation

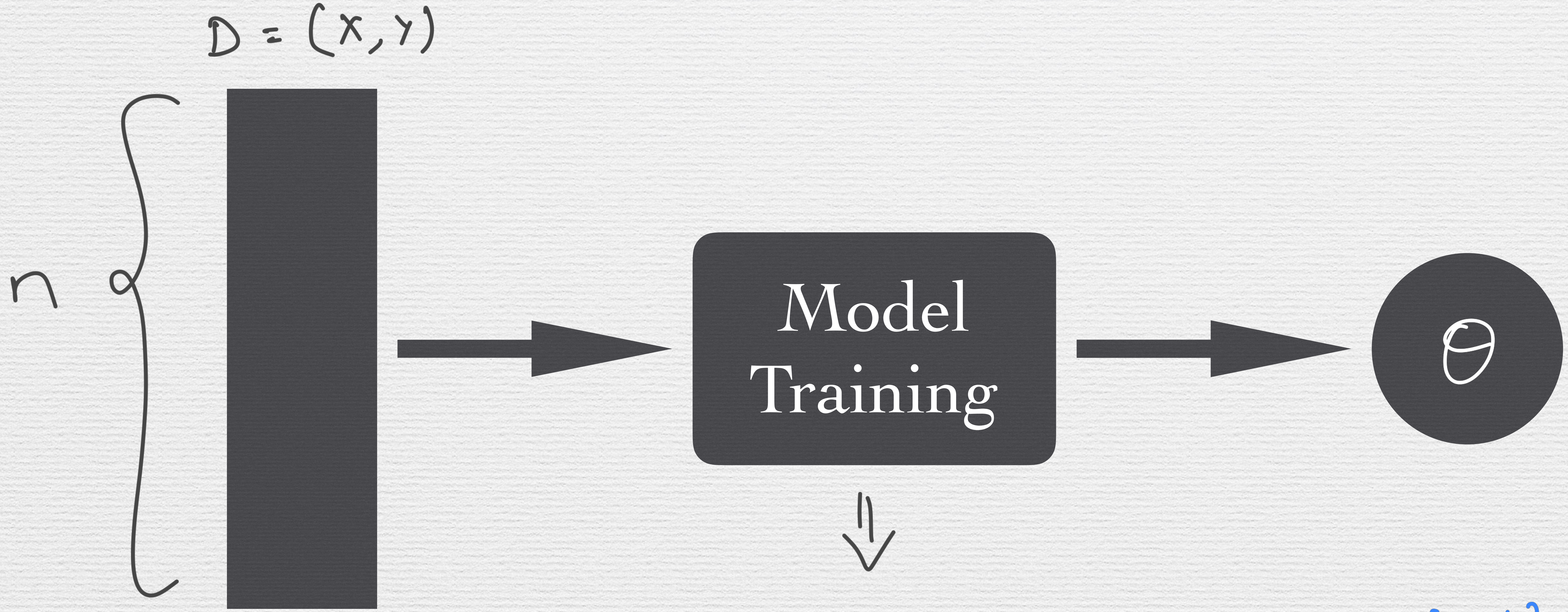
for t in T :

$\theta \leftarrow \theta - \eta \nabla J(\theta)$

return $\theta + \beta\left\{\propto \frac{1}{n}\right\}$

Chaudhuri et al. (2011)
Output Perturbation

Differential Private Solutions for Single Party Setting



$$J(\theta) = \frac{1}{n}\sum_{i=1}^{n} \ell(\theta, x_i, y_i) + \lambda N(\theta) + \beta\{\propto \tfrac{1}{n}\}$$

Chaudhuri et al. (2011)
Objective Perturbation

for t in T:

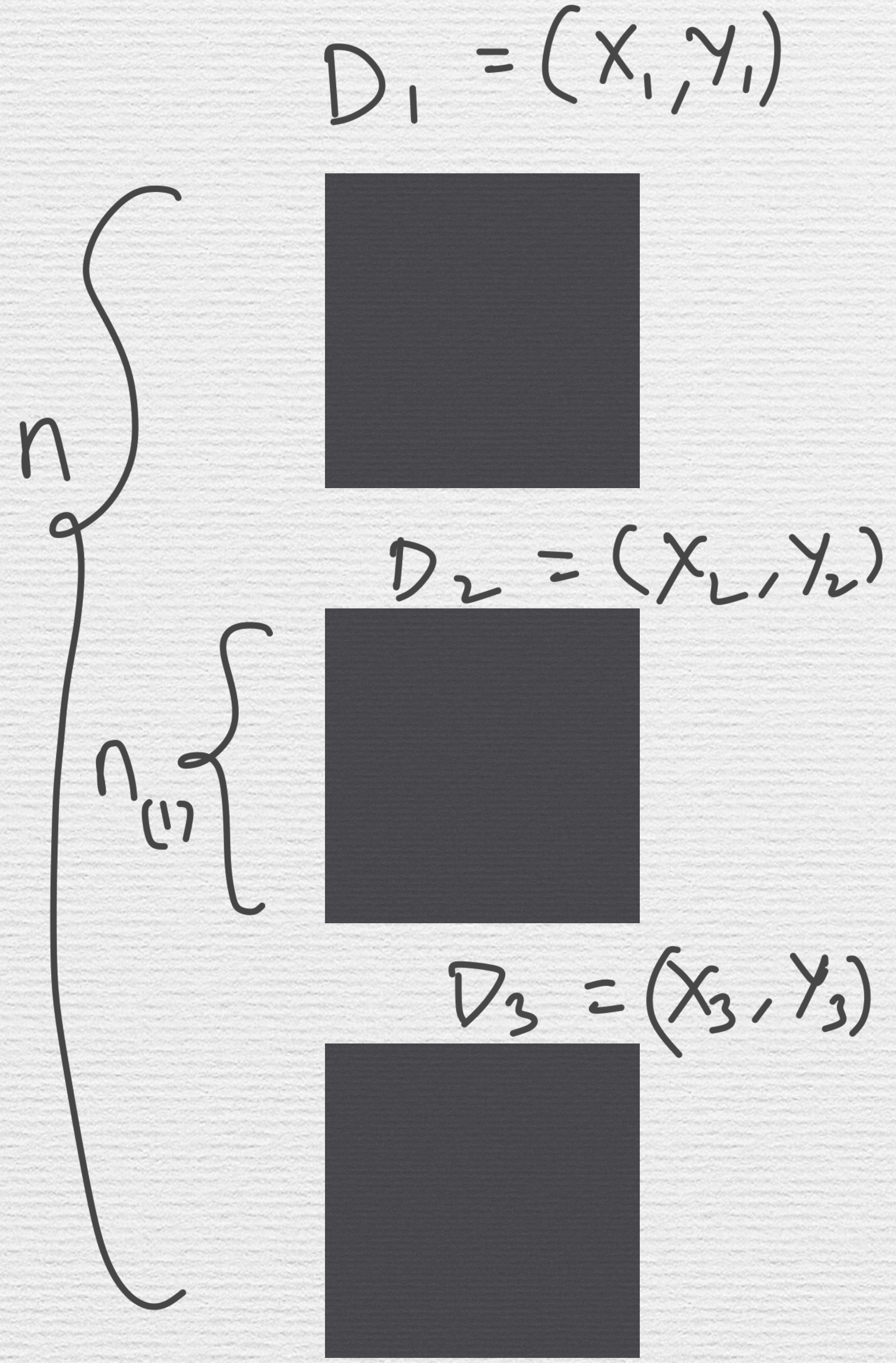
$$\theta \leftarrow \theta - \eta\left(\nabla J(\theta) + \beta\{\propto \tfrac{1}{n}\}\right)$$

Abadi et al. (2016)
Gradient Perturbation

return $\theta + \beta\{\propto \tfrac{1}{n}\}$

Chaudhuri et al. (2011)
Output Perturbation

Multi-Party Setting: Output Perturbation



Multi-Party Setting: Output Perturbation

	Noise Required
Pathak	$\beta \propto \frac{1}{n_{(1)}}$

Pathak et al. (2010)

Multi-Party Setting: Output Perturbation 2

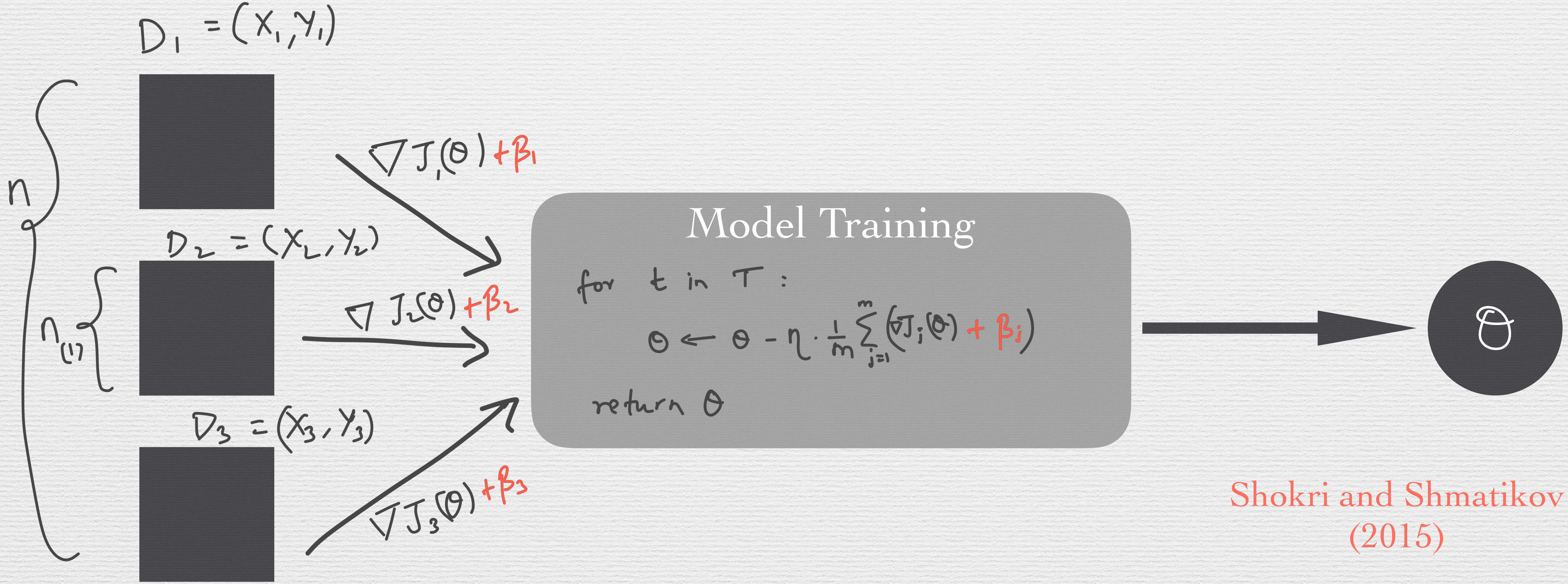
	Noise Required
Pathak	$\beta \propto \frac{1}{n_{(1)}}$
Chaudhuri	$\beta \propto \frac{1}{\sqrt{m} \cdot n_{(1)}}$

Improved Output Perturbation

	Noise Required
Pathak	$\beta \propto \frac{1}{n_{(1)}}$
Chaudhuri	$\beta \propto \frac{1}{\sqrt{m} \cdot n_{(1)}}$
Our Method	$\beta \propto \frac{1}{m \cdot n_{(1)}} \sim \frac{1}{n}$

Multi-Party Setting: Gradient Perturbation

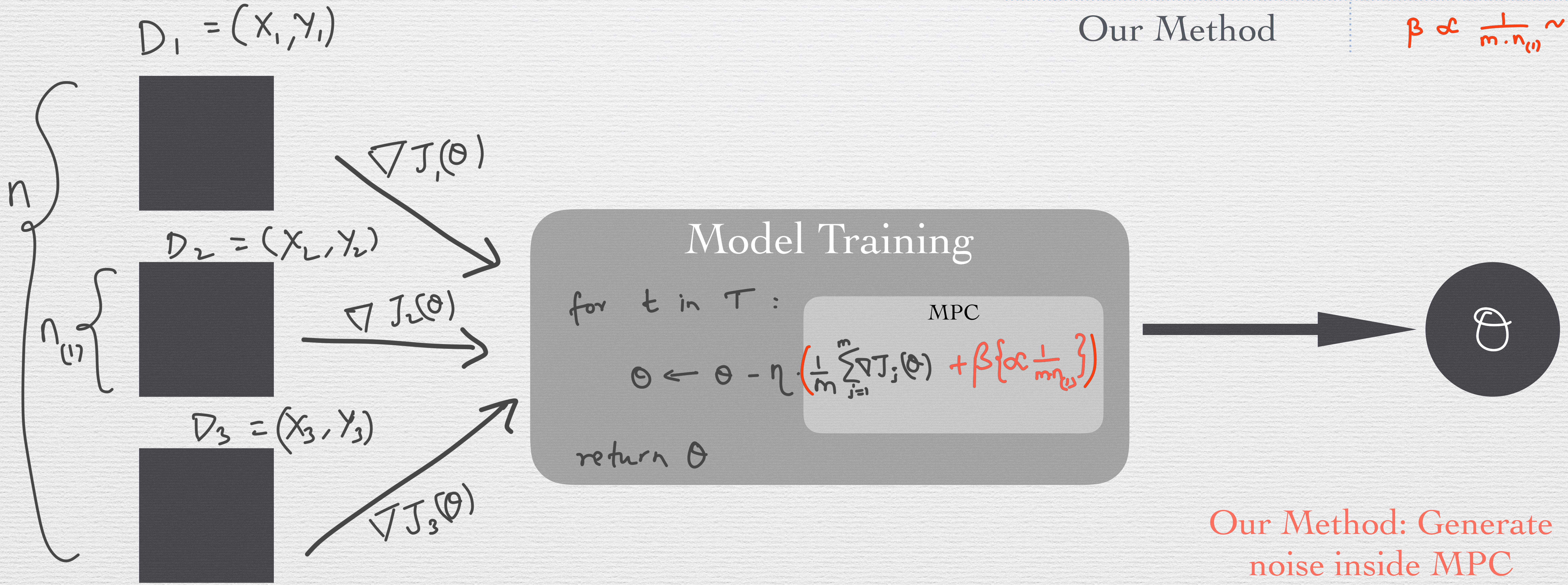
	Noise Required
Shokri & Shmatikov	$\beta \propto \frac{1}{\sqrt{m}\cdot n_{(1)}}$



Shokri and Shmatikov (2015)

Improved Gradient Perturbation

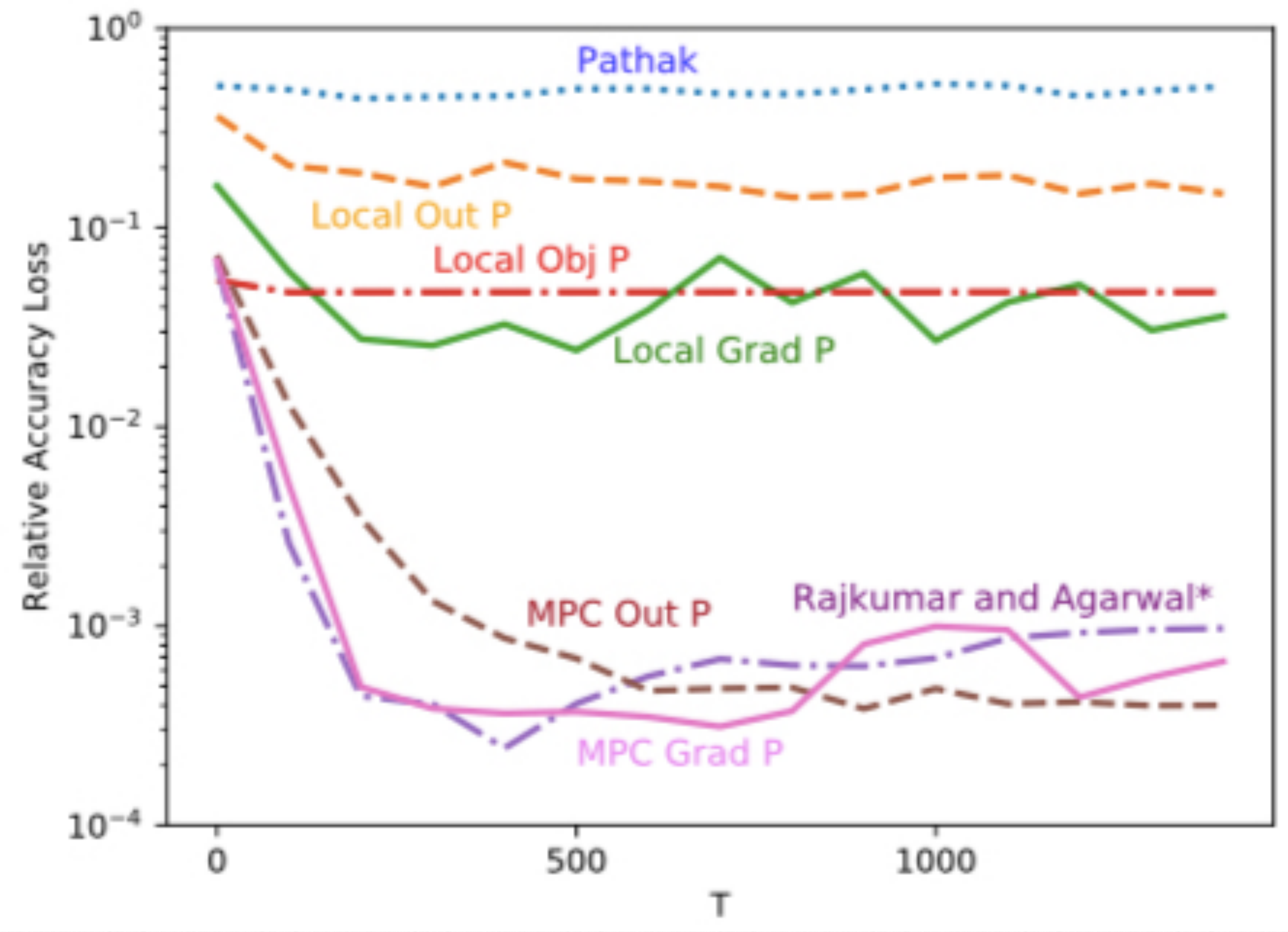
	Noise Required
Shokri & Shmatikov	$\beta \propto \frac{1}{\sqrt{m}\cdot n_{(1)}}$
Our Method	$\beta \propto \frac{1}{m \cdot n_{(1)}} \sim \frac{1}{n}$



Our Method: Generate noise inside MPC

KDDCup99 Dataset - Classification Task

$\epsilon = 0.5$

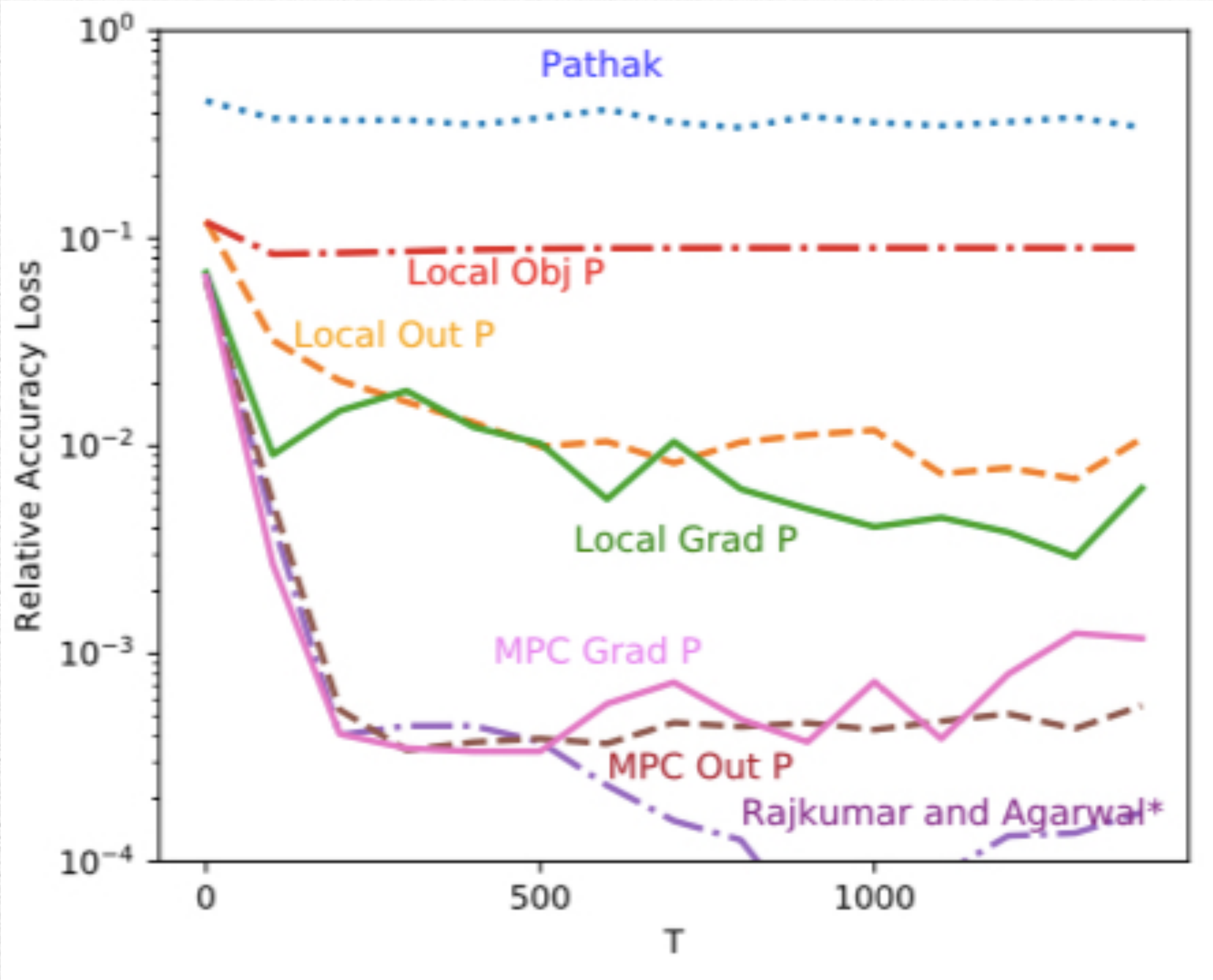


m = 1000

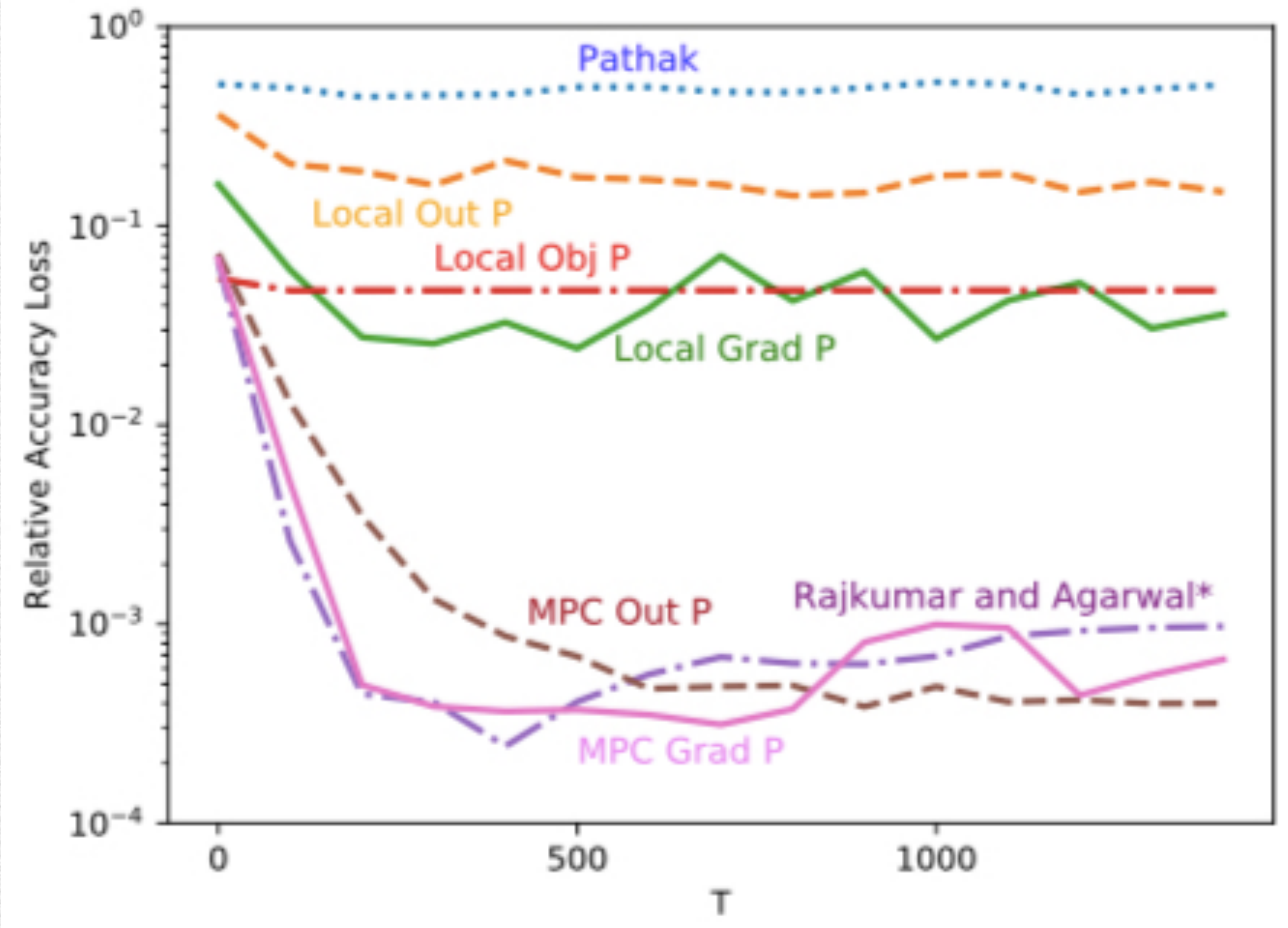
*Violates the privacy budget

KDDCup99 Dataset - Classification Task

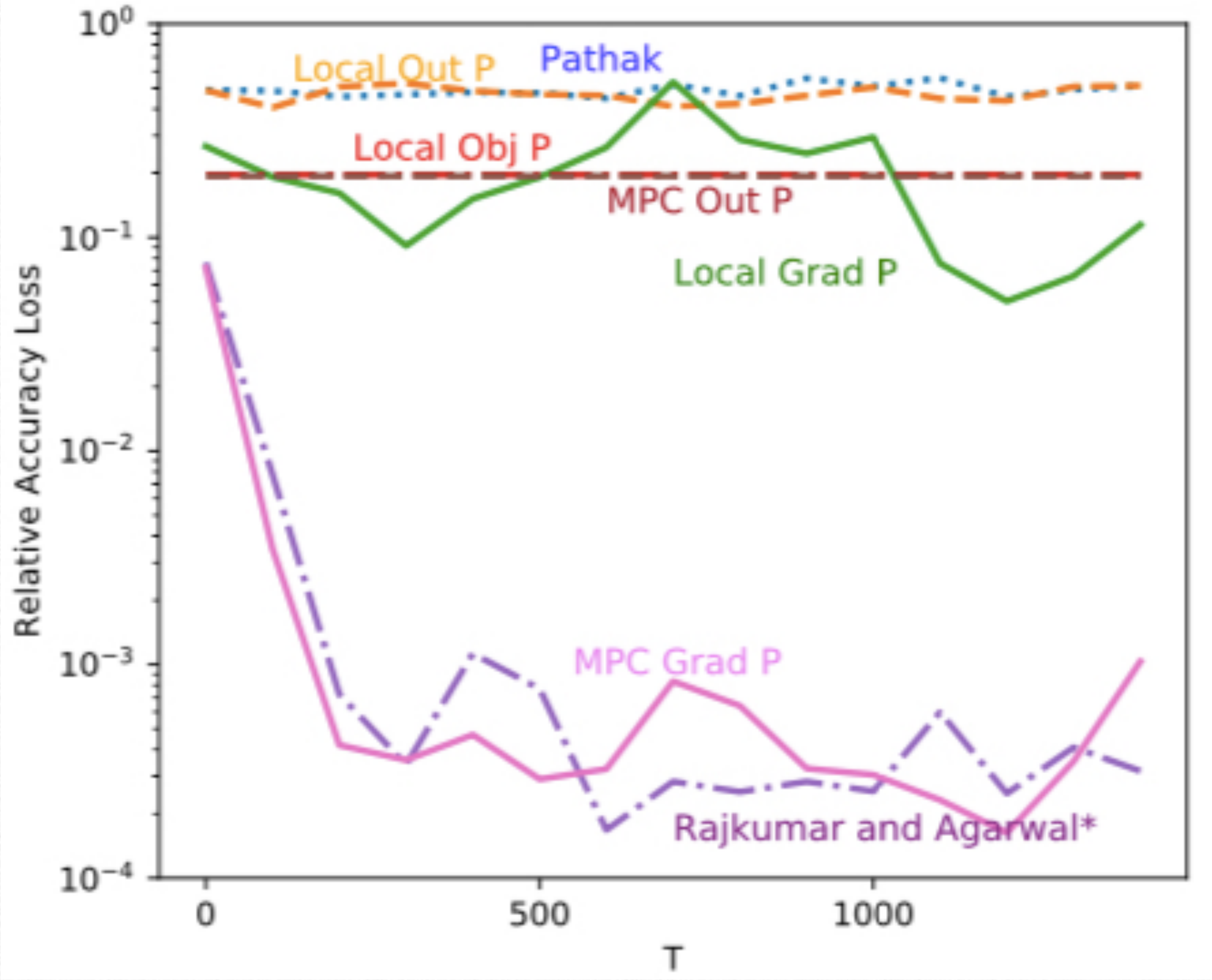
$\epsilon = 0.5$



m = 100



m = 1000

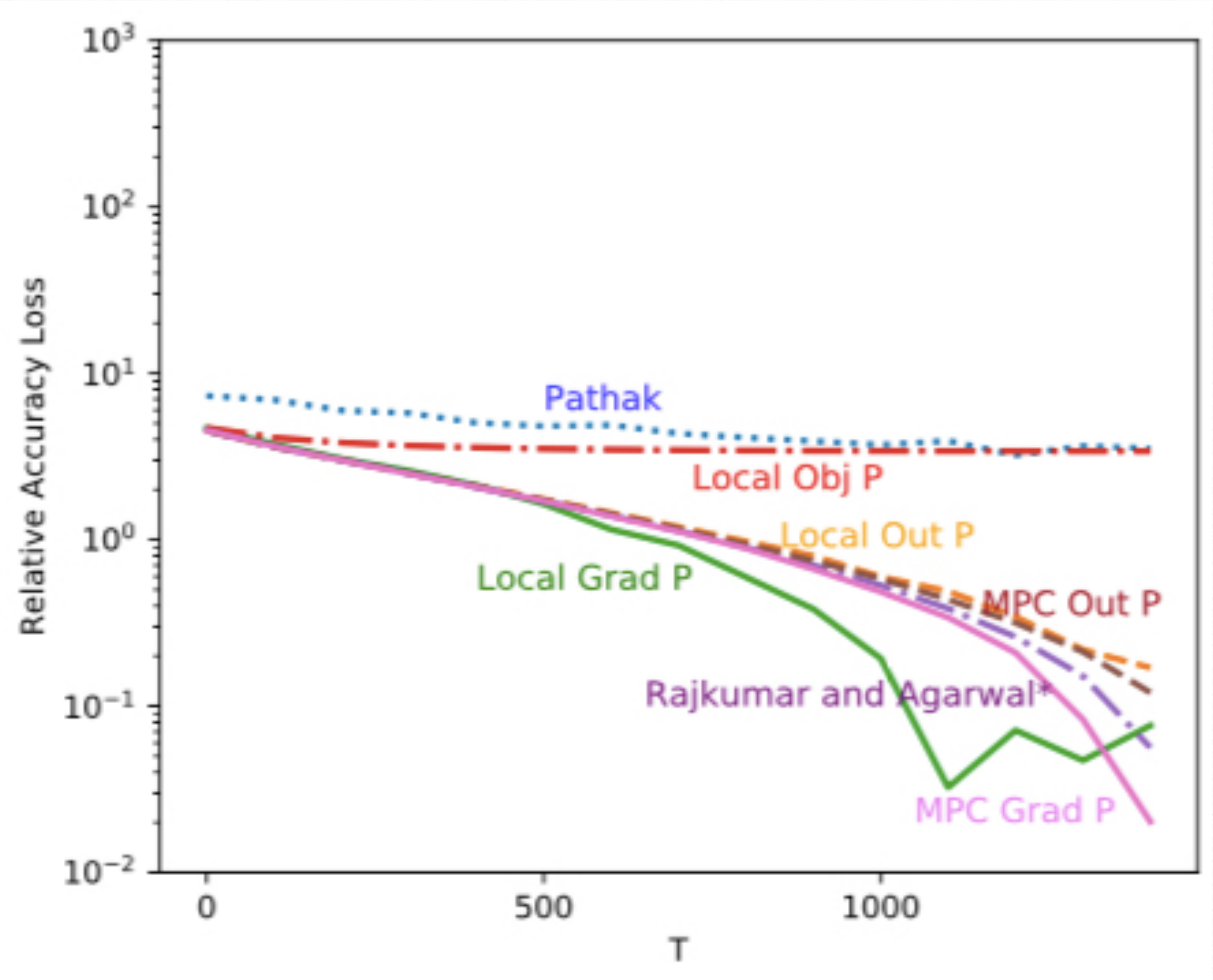


m = 50000

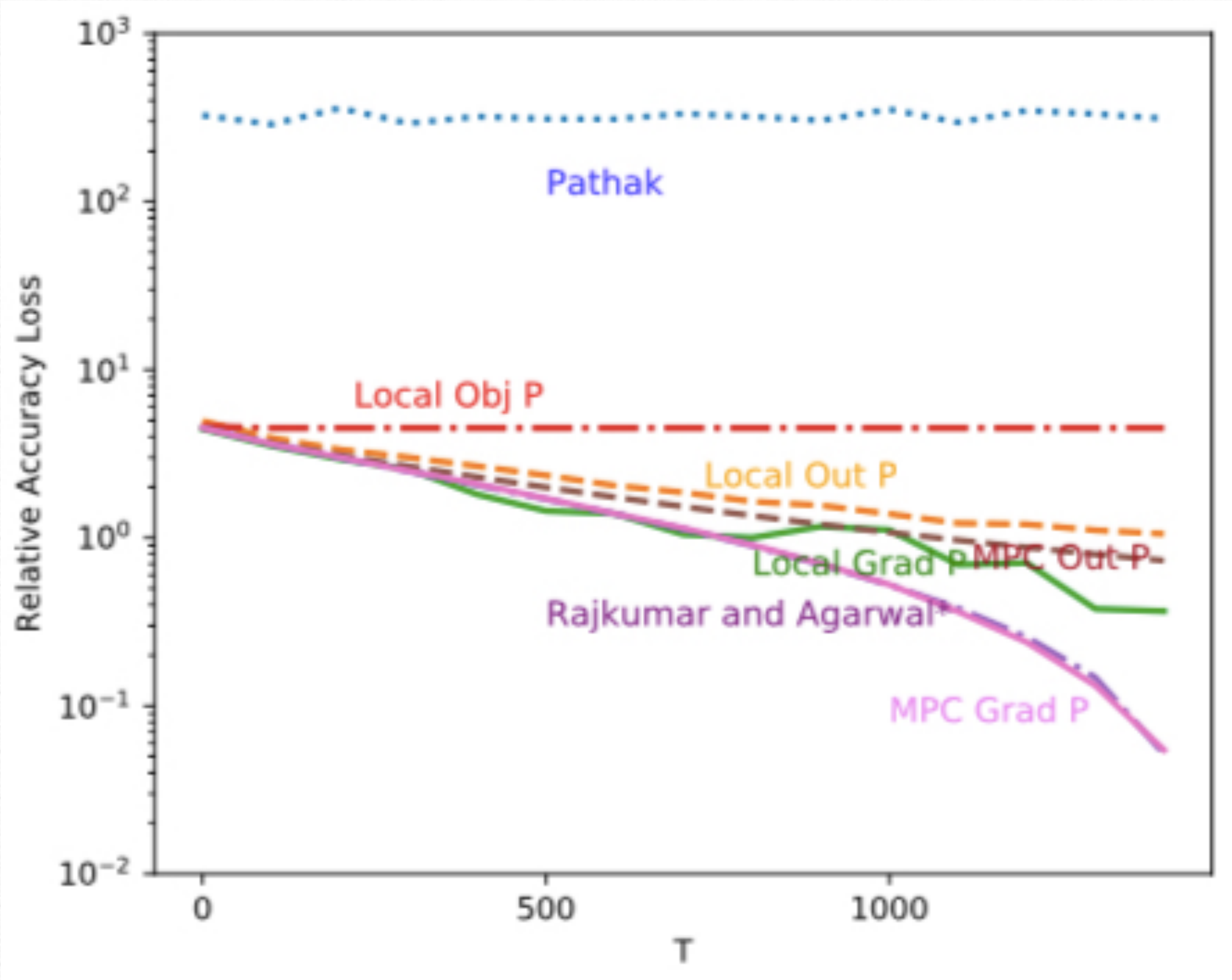
*Violates the privacy budget

KDDCup98 Dataset - Regression Task

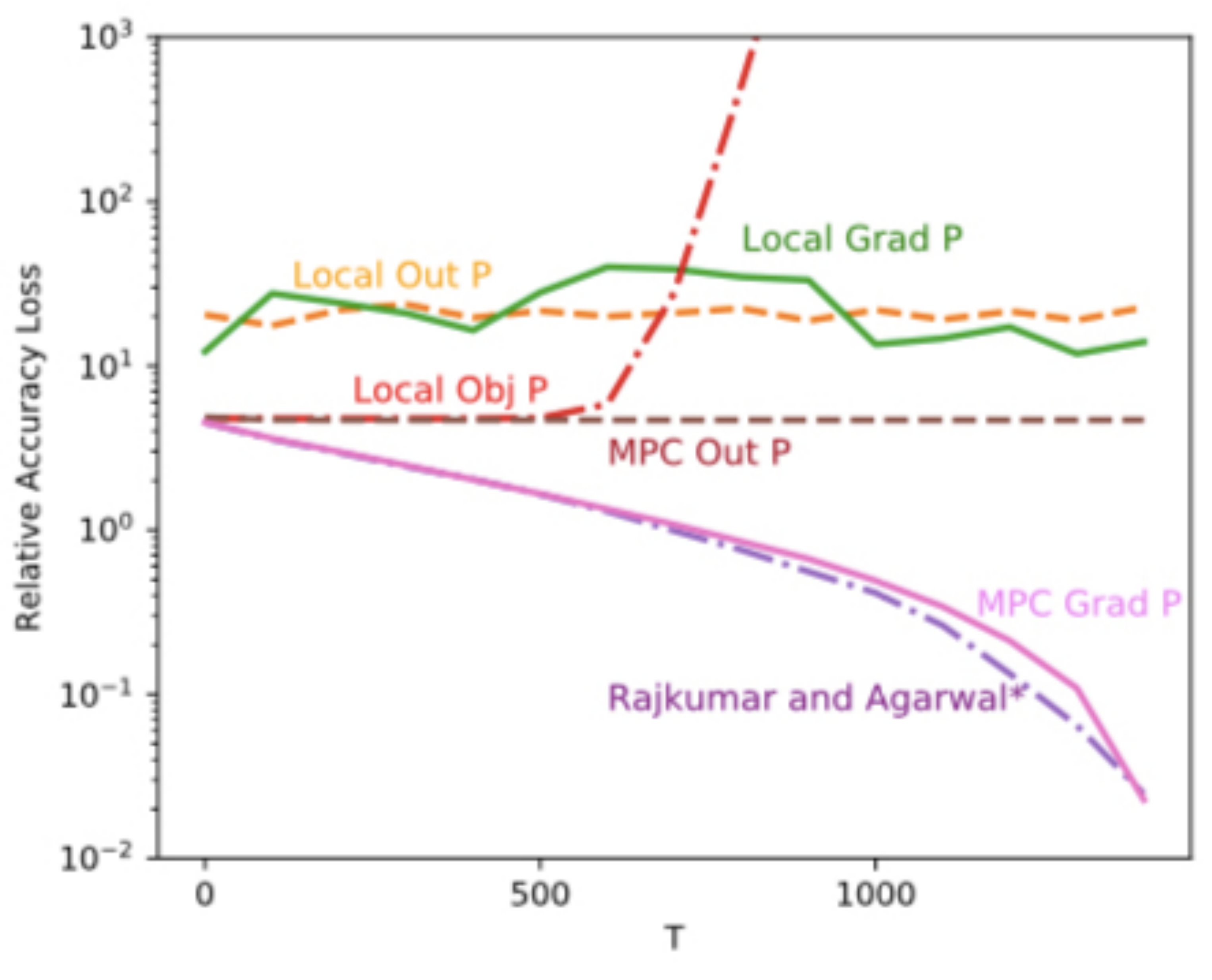
$\epsilon = 0.5$



m = 100



m = 1000



m = 50000

*Violates the privacy budget

Key Conclusion

Generating noise inside MPC and adding it after secure aggregation allows reducing the required noise in multi-party setting.

Shown via two instantiations of Differential Privacy:

1. Output Perturbation
2. Gradient Perturbation

Source Code

https://github.com/bargavj/distributedMachineLearning

References

- Martin Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar and Li Zhang. Deep learning with differential privacy. In *ACM SIGSAC Conference on Computer and Communications Security*, 2016.
- Kamalika Chaudhuri, Claire Monteleoni and Anand D. Sarwate. Differentially private empirical risk minimization. In *Journal of Machine Learning Research*, 2011.
- Manas Pathak, Shantanu Rane and Bhiksha Raj. Multiparty Differential Privacy via Aggregation of Locally Trained Classifiers. In *Advances in Neural Information Processing Systems*, 2010.
- Arun Rajkumar and Shivani Agarwal. A differentially private stochastic gradient descent algorithm for multiparty classification. In *Artificial Intelligence and Statistics*, 2012.
- Reza Shokri and Vitaly Shmatikov. Privacy-preserving deep learning. In *ACM Conference on Computer and Communications Security*, 2015.