Exploiting Problem Structure in Privacy-Preserving Optimization and Machine Learning

PhD Defense – Paul Mangold

Supervisors: Aurélien Bellet, Marc Tommasi

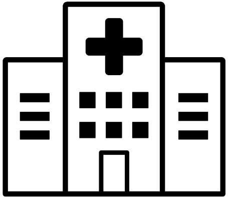
October 11, 2023

Let's Start with a Story

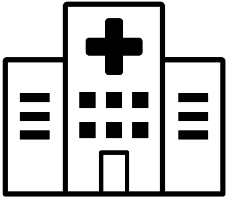
Let's Start with a Story

Let's Start with a Story

Let's Start with a Story

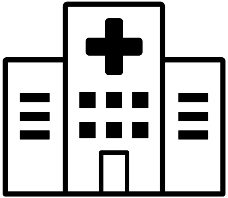


Let's Start with a Story



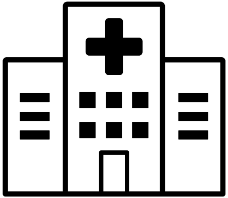
* Examination

Let's Start with a Story



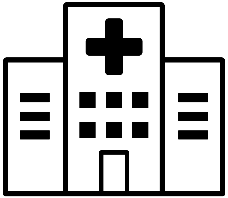
* Examination
* Diagnosis

Let's Start with a Story



* Examination
* Diagnosis
* Cure

Let's Start with a Story



* Examination
* Diagnosis
* Cure

⇒ possible due to years of medical research
(partly using statistical/machine learning)

Record	Age x_1	Pain x_2	... ...	Drug x_p	Sick y
#1	27	1	$\cdots$	1	1
#2	47	0	$\cdots$	1	0
#3	52	0	$\cdots$	0	0
#4	81	1	$\cdots$	0	1
...	...	...	...	...	...
#n	13	1	$\cdots$	0	1

How to study influence of possibly many features x_i's on an outcome y?

Record	Age x_1	Pain x_2	... ...	Drug x_p	Sick y
#1	27	1	$\cdots$	1	1
#2	47	0	$\cdots$	1	0
#3	52	0	$\cdots$	0	0
#4	81	1	$\cdots$	0	1
...	...	...	...	...	...
#n	13	1	$\cdots$	0	1

How to study influence of possibly many features x_i's on an outcome y?

One way: model $\log(\frac{\mathbb{P}(\text{sick})}{\mathbb{P}(\text{not sick})})$ as

$$h_{w^*}(x) = w_0^* + w_1^* \cdot x_1 + \cdots + w_p^* \cdot x_p$$

Record	Age x_1	Pain x_2	... ...	Drug x_p	Sick y
#1	27	1	⋯	1	1
#2	47	0	⋯	1	0
#3	52	0	⋯	0	0
#4	81	1	⋯	0	1
...	...	...	...	...	...
#n	13	1	⋯	0	1

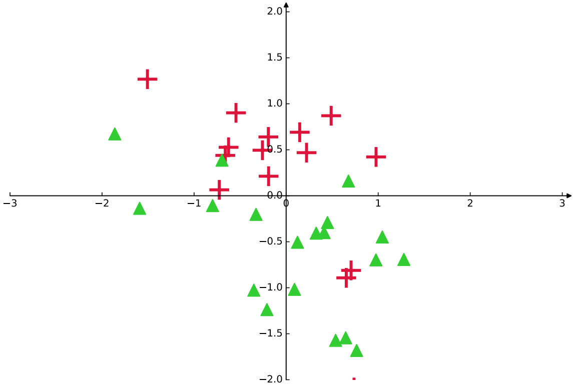
How to study influence of possibly many features x_i's on an outcome y?

One way: model $\log(\frac{\mathbb{P}(\text{sick})}{\mathbb{P}(\text{not sick})})$ as

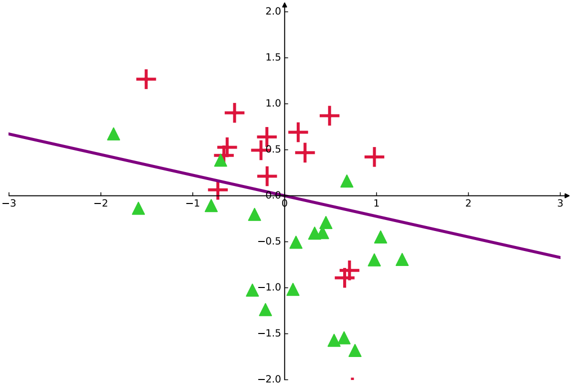
$$h_{w^*}(x) = w_0^* + w_1^* \cdot x_1 + \cdots + w_p^* \cdot x_p$$

Core remark: w^* is **computed from the data**!

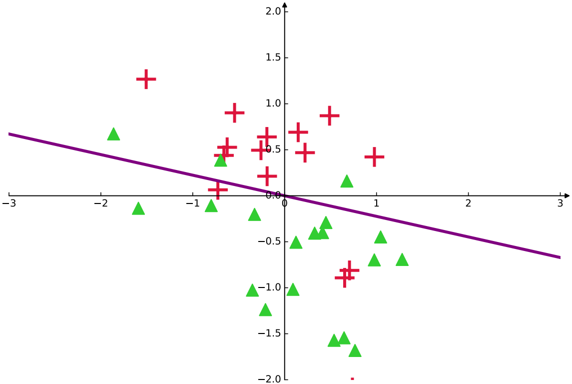
⇒ Trained Classification Model



⇒ Trained Classification Model



⇒ Trained Classification Model



The resulting model:

* is (quite) accurate
* contains info on data

Two Societal Concerns

#1 Privacy of training data

* guarantee that no confidential information is leaked

#2 Fairness of predictions

* guarantee similar predictions on all groups of population

Privacy Issues

Membership inference*:

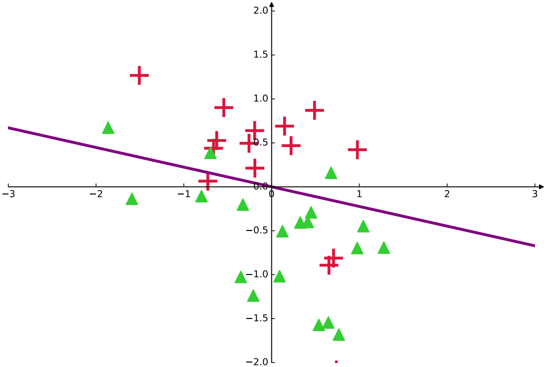
“ determine whether a given record was part of a model’s training dataset “

*R. Shokri et al. “Membership Inference Attacks Against Machine Learning Models”. 2017.

Privacy Issues

Membership inference*:

> *" determine whether a given record was part of a model's training dataset "*

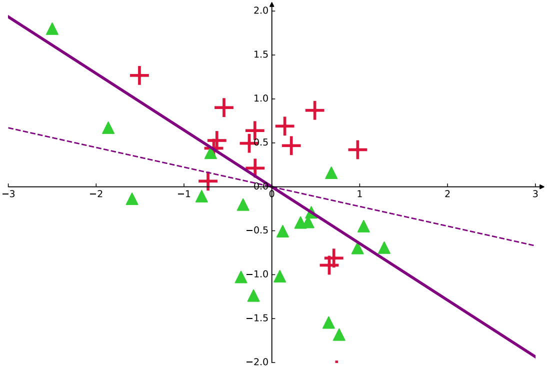


*R. Shokri et al. "Membership Inference Attacks Against Machine Learning Models". 2017.

Privacy Issues

Membership inference*:

“ determine whether a given record was part of a model’s training dataset “



*R. Shokri et al. “Membership Inference Attacks Against Machine Learning Models”. 2017.

Guaranteeing Privacy

Perturb the linear predictor:

$$h_{w^*}(x) = w_0^* + w_1^* \cdot x_1 + \cdots + w_p^* \cdot x_p$$

Guaranteeing Privacy

Perturb the linear predictor:

$$h_{w^*+\eta}(x) = (w_0^* + \eta_0) + (w_1^* + \eta_1) \cdot x_1 + \cdots + (w_p^* + \eta_p) \cdot x_p$$

Guaranteeing Privacy

Perturb the linear predictor:

$$h_{w^*+\eta}(x) = (w_0^* + \eta_0) + (w_1^* + \eta_1) \cdot x_1 + \cdots + (w_p^* + \eta_p) \cdot x_p$$

✓ noise gives *plausible deniability* → better privacy

✗ noisy predictions → lower accuracy

Guaranteeing Privacy

Perturb the linear predictor:

$$h_{w^*+\eta}(x) = (w_0^* + \eta_0) + (w_1^* + \eta_1) \cdot x_1 + \cdots + (w_p^* + \eta_p) \cdot x_p$$

✓ noise gives *plausible deniability* → better privacy

✗ noisy predictions → lower accuracy

⇒ **tension between privacy and utility**

How Strong is the Protection?

$\mathcal{A} : D \mapsto w$ is (ϵ, δ)-*Differentially Private*[*]

*C. Dwork. "Differential Privacy". 2006.

How Strong is the Protection?

$\mathcal{A} : D \mapsto w$ is (ϵ, δ)-*Differentially Private**

$$\mathbb{P}(\mathcal{A}(D) \in \mathcal{S}) \leq \exp(\boldsymbol{\epsilon}) \cdot \mathbb{P}(\mathcal{A}(D') \in \mathcal{S}) + \boldsymbol{\delta}$$

for *all* D, D' that differ on one element

*C. Dwork. "Differential Privacy". 2006.

How Strong is the Protection?

$\mathcal{A} : D \mapsto w$ is (ϵ, δ)-*Differentially Private**

$$\mathbb{P}(\mathcal{A}(D) \in \mathcal{S}) \leq \exp(\boldsymbol{\epsilon}) \cdot \mathbb{P}(\mathcal{A}(D') \in \mathcal{S}) + \boldsymbol{\delta}$$

for *all* D, D' that differ on one element

Rule of thumb: $\boldsymbol{\epsilon} \leq 1$, $\boldsymbol{\delta} = o(1/|D|)$

---

*C. Dwork. "Differential Privacy". 2006.

Two Societal Concerns

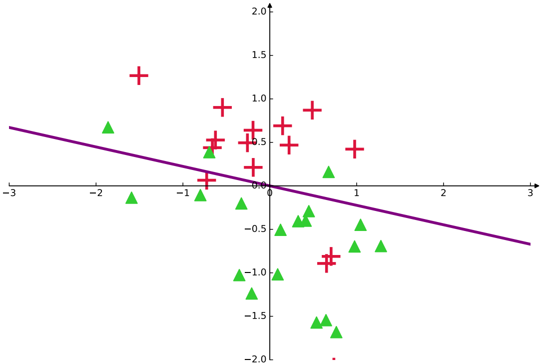
#1 Privacy of training data

* guarantee that no confidential information is leaked

#2 Fairness of predictions

* guarantee similar predictions on all groups of population

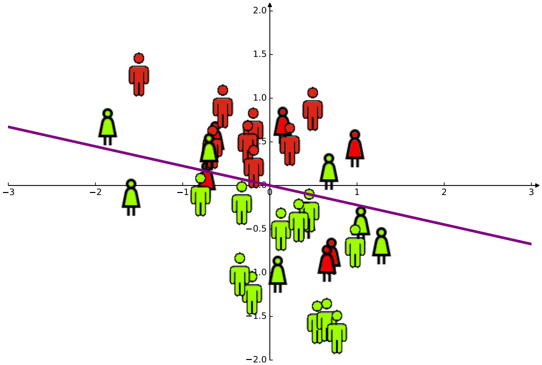
Fairness Issues



Group Fairness:

Different groups can be treated differently

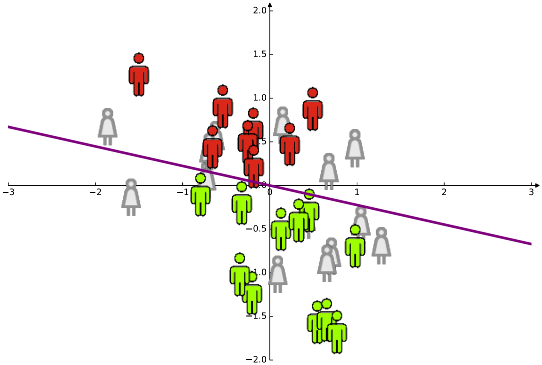
Fairness Issues



Group Fairness:

Different groups can be treated differently

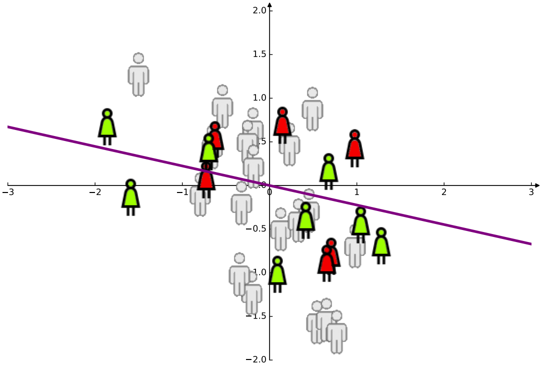
Fairness Issues



Group Fairness:

Different groups can be treated differently

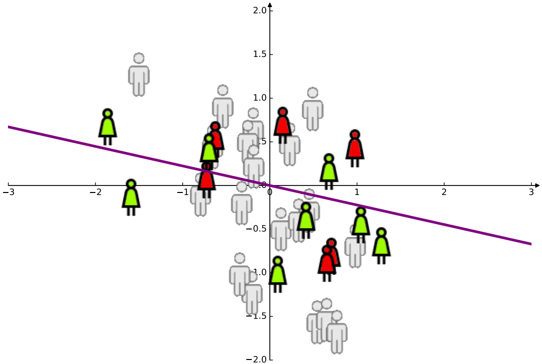
Fairness Issues



Group Fairness:

Different groups can be treated differently

Fairness Issues



GROUP FAIRNESS:

Different groups can be treated differently

Note: Perturbing the model can have a disparate impact[a]

[a] E. Bagdasaryan et al. "DP Has Disparate Impact on Model Accuracy". 2019.

How to exploit problem's structure to:

* obtain better utility?
* study the impact of privacy on fairness?

Contributions

Contributions

* Private learning algorithms exploiting structure
  1. Imbalanced parameter scales and variations
  2. High-dimensional models with imbalanced solutions

Contributions

* Private learning algorithms exploiting structure
  1. Imbalanced parameter scales and variations
  2. High-dimensional models with imbalanced solutions

* Study interplay between privacy and fairness
  3. Bound on the impact of privacy on fairness

Contributions

* Private learning algorithms exploiting structure
  1. Imbalanced parameter scales and variations
  2. High-dimensional models with imbalanced solutions

* Study interplay between privacy and fairness
  3. Bound on the impact of privacy on fairness

Empirical Risk Minimization

Note: Most results also hold for composite ERM with Proximal algorithms

$$w^* \in \operatorname*{arg\,min}_{w \in \mathcal{W}} \left\{ f(w) = \frac{1}{n} \sum_{i=1}^{n} \ell(w; d_i) \right\}$$

Empirical Risk Minimization

Note: Most results also hold for composite ERM with Proximal algorithms

$$w^* \in \arg\min_{w \in \mathcal{W}} \left\{ f(w) = \frac{1}{n} \sum_{i=1}^{n} \ell(w; d_i) \right\}$$

Where $\mathcal{W} \subseteq \mathbb{R}^p$, has diameter $\|\mathcal{W}\|_2$, and ℓ is

- convex: $\ell(w; d) \geq \ell(w'; d) + \langle \nabla \ell(w'; d), w - w' \rangle$
- smooth: $\|\nabla \ell(w; d) - \nabla \ell(w'; d)\| \leq M \|w - w'\|$
- Lipschitz: $|\ell(w; d) - \ell(w'; d)| \leq \Lambda \|w - w'\|$

Empirical Risk Minimization

Note: Most results also hold for composite ERM with Proximal algorithms

$$w^* \in \arg\min_{w \in \mathcal{W}} \left\{ f(w) = \frac{1}{n} \sum_{i=1}^{n} \ell(w; d_i) \right\}$$

Where $\mathcal{W} \subseteq \mathbb{R}^p$, has diameter $\|\mathcal{W}\|_2$, and ℓ is

* convex: $\ell(w; d) \geq \ell(w'; d) + \langle \nabla \ell(w'; d), w - w' \rangle$
* smooth: $\|\nabla \ell(w; d) - \nabla \ell(w'; d)\| \leq M \|w - w'\|$
* Lipschitz: $|\ell(w; d) - \ell(w'; d)| \leq \Lambda \|w - w'\|$

Empirical Risk Minimization

Note: Most results also hold for composite ERM with Proximal algorithms

$$w^* \in \arg\min_{w \in \mathcal{W}} \left\{ f(w) = \frac{1}{n} \sum_{i=1}^{n} \ell(w; d_i) \right\}$$

Where $\mathcal{W} \subseteq \mathbb{R}^p$, has diameter $\|\mathcal{W}\|_2$, and ℓ is

* convex: $\ell(w; d) \geq \ell(w'; d) + \langle \nabla \ell(w'; d), w - w' \rangle$
* smooth: $\|\nabla \ell(w; d) - \nabla \ell(w'; d)\| \leq M\|w - w'\|$
* Lipschitz: $|\ell(w; d) - \ell(w'; d)| \leq \Lambda \|w - w'\|$

Empirical Risk Minimization

Note: Most results also hold for composite ERM with Proximal algorithms

$$w^* \in \arg\min_{w \in \mathcal{W}} \left\{ f(w) = \frac{1}{n} \sum_{i=1}^{n} \ell(w; d_i) \right\}$$

Where $\mathcal{W} \subseteq \mathbb{R}^p$, has diameter $\|\mathcal{W}\|_2$, and ℓ is

* convex: $\ell(w; d) \geq \ell(w'; d) + \langle \nabla \ell(w'; d), w - w' \rangle$
* smooth: $\|\nabla \ell(w; d) - \nabla \ell(w'; d)\| \leq M \|w - w'\|$
* Lipschitz: $|\ell(w; d) - \ell(w'; d)| \leq \Lambda \|w - w'\|$

Empirical Risk Minimization

Note: Most results also hold for composite ERM with Proximal algorithms

$$w^* \in \arg\min_{w \in \mathcal{W}} \left\{ f(w) = \frac{1}{n} \sum_{i=1}^{n} \ell(w; d_i) \right\}$$

Where $\mathcal{W} \subseteq \mathbb{R}^p$, has diameter $\|\mathcal{W}\|_2$, and ℓ is

* convex: $\ell(w; d) \geq \ell(w'; d) + \langle \nabla \ell(w'; d), w - w' \rangle$
* smooth: $\|\nabla \ell(w; d) - \nabla \ell(w'; d)\| \leq M \|w - w'\|$
* Lipschitz: $\|\nabla \ell(w; d)\| \leq \Lambda$

Empirical Risk Minimization

Note: Most results also hold for composite ERM with Proximal algorithms

How to solve ERM privately?

* smooth: $\|\nabla \ell(w; d) - \nabla \ell(w'; d)\| \le M\|w - w'\|$
* Lipschitz: $\|\nabla \ell(w; d)\| \le \Lambda$

DP-SGD[*,†]

Differentially Private Stochastic Gradient Descent

For $t = 0$ to $T - 1$:

- Choose a data record d_i
- Draw noise $\boldsymbol{\eta^t \sim \mathcal{N}(0; \sigma^2 \mathbb{I}_p)}$
- Update $w^{t+1} = w^t - \gamma^t \left(\nabla \ell(w^t; d_i) + \eta^t\right)$

Return w^T

---

[*] S. Song et al. "Stochastic Gradient Descent with Differentially Private Updates". 2013.

[†] R. Bassily et al. "Private ERM: Efficient Algorithms and Tight Error Bounds". 2014.

Privacy of DP-SGD*,†

For (ϵ, δ)-differential privacy we need

$$\sigma^2 = O\left(\frac{\Lambda T}{n^2\epsilon^2}\right) \quad , \quad \text{where } \|\nabla \ell\| \leq \Lambda$$

- Noise increases with number of iterations
- Sampling amplifies privacy

---

*S. Song et al. "Stochastic Gradient Descent with Differentially Private Updates". 2013.

†R. Bassily et al. "Private ERM: Efficient Algorithms and Tight Error Bounds". 2014.

Utility of DP-SGD*

$$\mathbb{E}(f(w^{SGD}) - f(w^*)) = O\left(\frac{\Lambda\|\mathcal{W}\|_2}{\epsilon\sqrt{T}} + \frac{\sqrt{T}p\Lambda\|\mathcal{W}\|_2\log(1/\delta)}{n^2\epsilon}\right)$$

optimization error — privacy error

*R. Bassily et al. "Private ERM: Efficient Algorithms and Tight Error Bounds". 2014.

Utility of DP-SGD*

$$\mathbb{E}(f(w^{SGD}) - f(w^*)) = O\left(\frac{\Lambda \|\mathcal{W}\|_2 \sqrt{p \log(1/\delta)}}{n\epsilon}\right)$$

after balancing the two terms

*R. Bassily et al. "Private ERM: Efficient Algorithms and Tight Error Bounds". 2014.

Utility of DP-SGD*

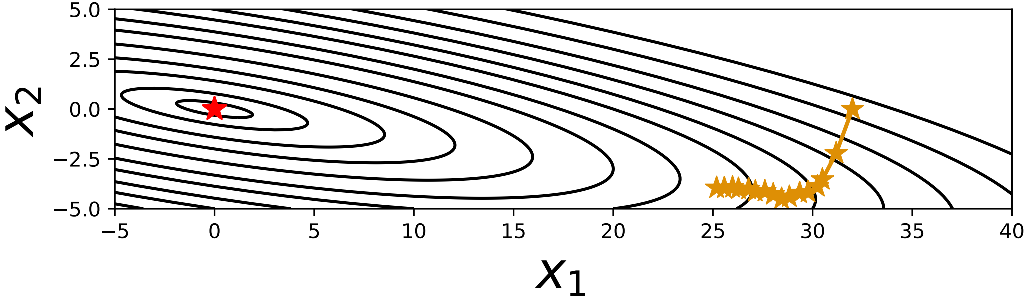
$$\mathbb{E}(f(w^{SGD}) - f(w^*)) = \boldsymbol{\Theta}\left(\frac{\Lambda\|\mathcal{W}\|_2\sqrt{p\log(1/\delta)}}{n\epsilon}\right)$$

$\Rightarrow$ and the result is *tight* (under these assumptions)

*R. Bassily et al. "Private ERM: Efficient Algorithms and Tight Error Bounds". 2014.

The Problem of DP-SGD

It fails on imbalanced problems...



We need to refine measure of regularity of f:

* smoothness:

$$\|\nabla f(w+t) - \nabla f(w)\| \leq M\|t\|$$

* Lipschitzness:

$$\|\nabla f(w)\| \leq \Lambda$$

We need to refine measure of regularity of f:

* coordinate-wise smoothness:

$$|\nabla_j f(w + te_j) - \nabla_j f(w)| \leq M_j |t|$$

* coordinate-wise Lipschitzness:

$$|\nabla_j f(w)| \leq L_j$$

We need to refine measure of regularity of f:

* coordinate-wise smoothness:

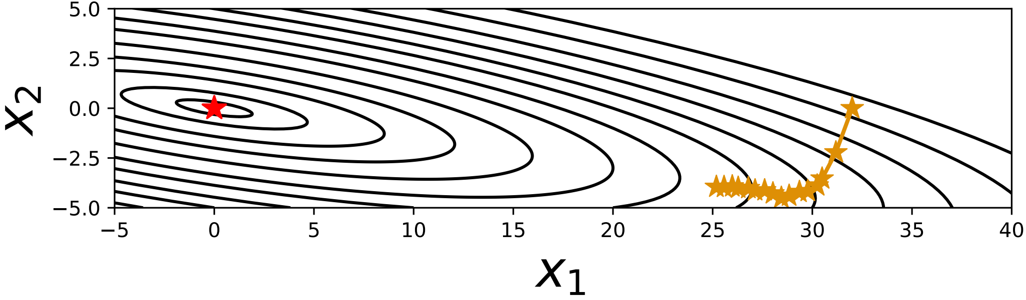
$$|\nabla_j f(w + te_j) - \nabla_j f(w)| \leq M_j |t|$$

* coordinate-wise Lipschitzness:

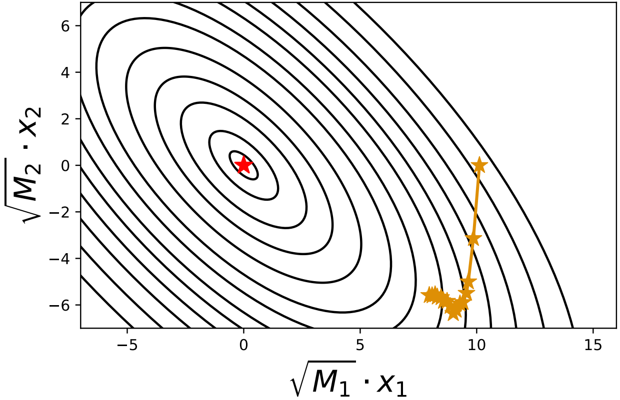
$$|\nabla_j f(w)| \leq L_j$$

Important: $M_j \leq M$, and $L_j \leq \Lambda$

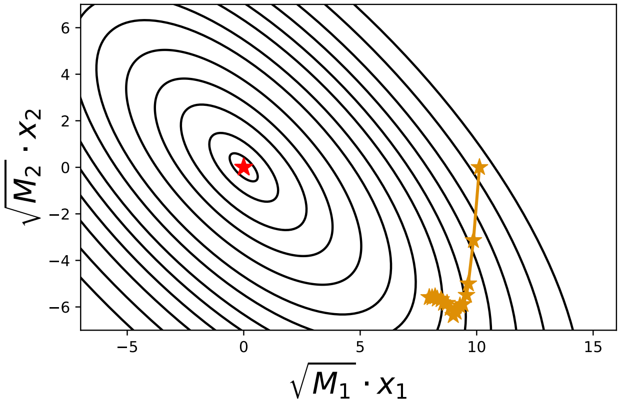
We can now use a more appropriate measure of our space!



We can now use a more appropriate measure of our space!



We can now use a more appropriate measure of our space!



Scaled norm: $\|w\|_{M,q} = \left(\sum_{j=1}^{p} M_j^{\frac{q}{2}} |w_j|^q \right)^{\frac{1}{q}}$ for $q \in \{1, 2\}$

Contribution 1: DP-CD*

Differentially Private Coordinate Descent

For $t = 0$ to $T - 1$:

* Choose a *coordinate* $j \in [p]$
* Draw noise $\eta_j^t \sim \mathcal{N}\left(0; \sigma_j^2\right)$
* Update $w_j^{t+1} = w_j^t - \gamma_j (\nabla_j f(w^t) + \eta_j^t)$

Return $w^{CD} = \frac{1}{T} \sum_{t=1}^{T} w^t$

---

*P. Mangold et al. "Differentially Private Coordinate Descent for Composite ERM". 2022.

Contribution 1: DP-CD*

Differentially Private Coordinate Descent

For $t = 0$ to $T - 1$:

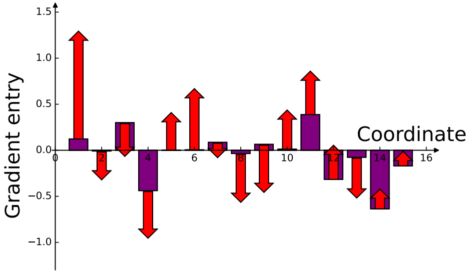
- Choose a *coordinate* $\boldsymbol{j} \in [p]$
- Draw noise $\eta_{\boldsymbol{j}}^t \sim \mathcal{N}\left(0; \boldsymbol{O}\left(\frac{\boldsymbol{L_j T}}{\boldsymbol{n^2 \epsilon^2}}\right)\right)$
- Update $w_{\boldsymbol{j}}^{t+1} = w_{\boldsymbol{j}}^t - \gamma_{\boldsymbol{j}}(\nabla_{\boldsymbol{j}} f(w^t) + \eta_{\boldsymbol{j}}^t)$

Return $w^{CD} = \frac{1}{T} \sum_{t=1}^{T} w^t$

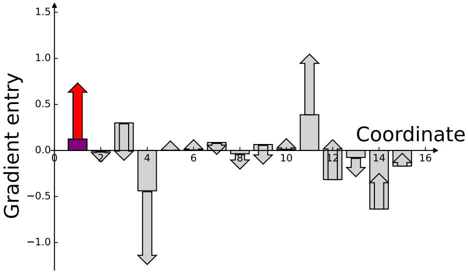
---

*P. Mangold et al. "Differentially Private Coordinate Descent for Composite ERM". 2022.

DP-SGD noise:



DP-CD noise:



Utility of DP-CD

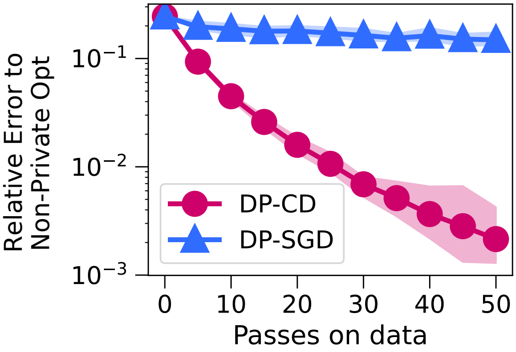
$$\mathbb{E}(f(w^{CD}) - f(w^*)) \leq O\left(\frac{\sqrt{p \log(1/\delta)}}{n\epsilon} \|L\|_{M^{-1}} \|\mathcal{W}\|_M \right)$$

Recall that for DP-SGD:

$$\mathbb{E}(f(w^{SGD}) - f(w^*)) \leq O\left(\frac{\sqrt{p \log(1/\delta)}}{n\epsilon} \Lambda \|\mathcal{W}\|_2 \right)$$

Numerical Illustration

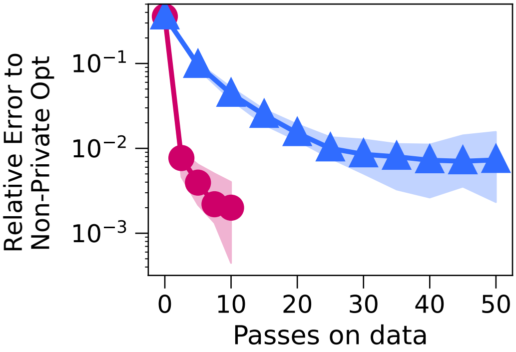
DP-CD uses more appropriate step sizes



* Regularized logistic regression
* Raw (imbalanced) data
* $n = 45,312$ records
* $p = 8$ features
* $\epsilon = 1,\ \delta = 1/n^2$

Numerical Illustration

DP-CD does not require amplification by sampling



- Regularized logistic regression
- Standardized data
- $n = 45,312$ records
- $p = 8$ features
- $\epsilon = 1,\ \delta = 1/n^2$

Contribution 2: DP-GCD*

Differentially Private **Greedy** Coordinate Descent

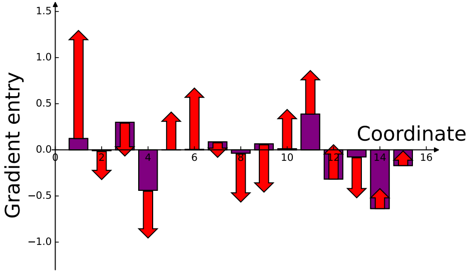
For $t = 0$ to $T - 1$:

- Draw noise $\eta_j^t, \zeta_j^t \sim \text{Lap}\left(0; O\left(\frac{L_j T}{n^2 \epsilon^2}\right)\right)$
- Choose $j = \arg\max_{j' \in [p]} |\nabla_{j'} f(w^t) + \zeta_{j'}|$
- Update $w^{t+1} = w^t - \gamma_j (\nabla_j f(w^t) + \eta_j^t)$

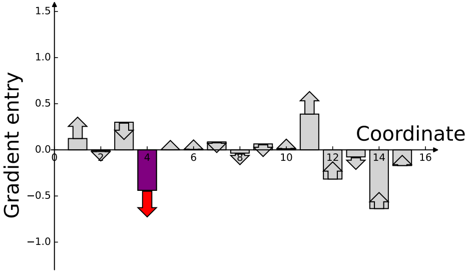
Return $w^{GCD} = w^T$

*P. Mangold et al. "High-Dimensional Private ERM by Greedy Coordinate Descent". 2023.

DP-SGD noise:



DP-GCD noise:



Utility of DP-GCD

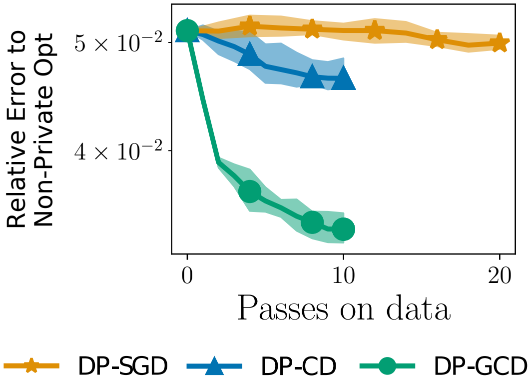
$$\mathbb{E}(f(w^{GCD}) - f(w^*)) \leq O\left(\frac{\log(1/\delta)\boldsymbol{\log(p)}}{n^{2/3}\epsilon^{2/3}} L_{\max}^{2/3} \|\mathcal{W}\|_{M,1}^{4/3}\right)$$

Recall that:

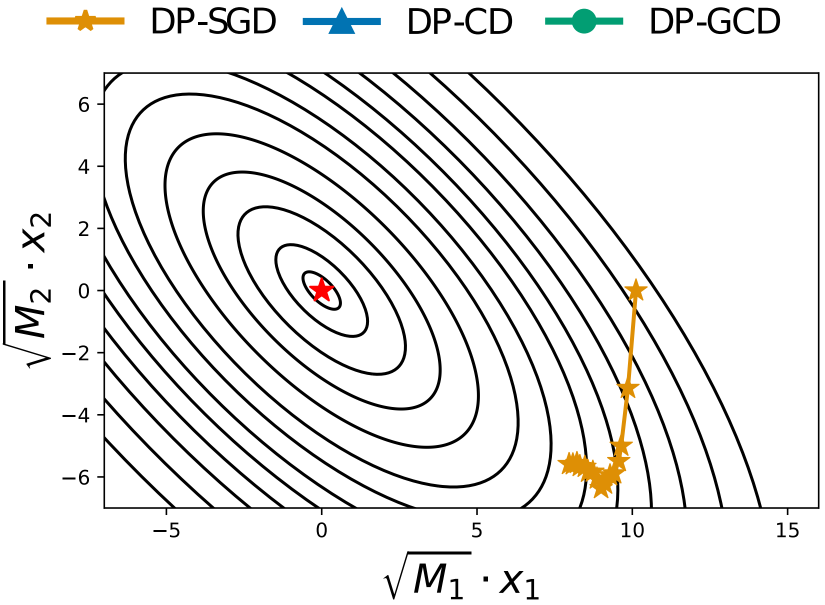
$$\mathbb{E}(f(w^{SGD}) - f(w^*)) \leq O\left(\frac{\sqrt{\boldsymbol{p}\log(1/\delta)}}{n\epsilon} \Lambda \|\mathcal{W}\|_2\right)$$

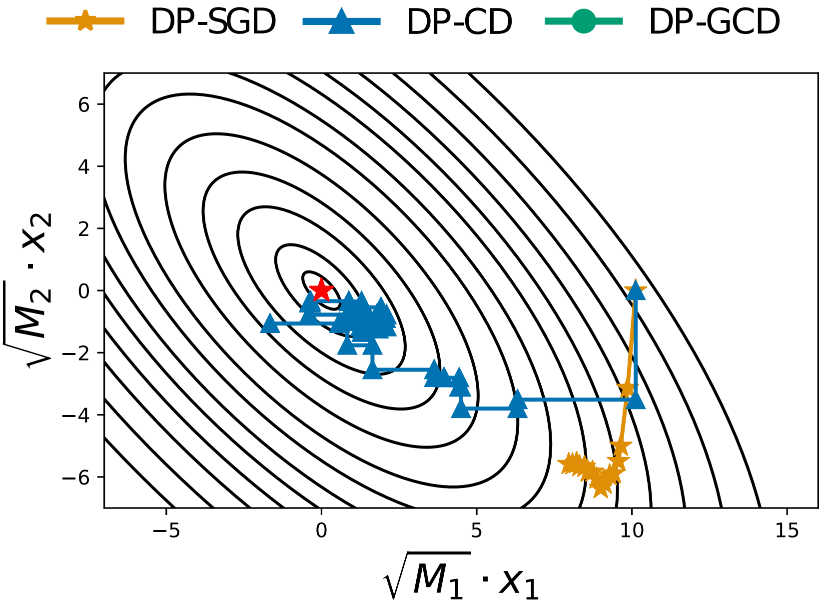
Numerical Illustration

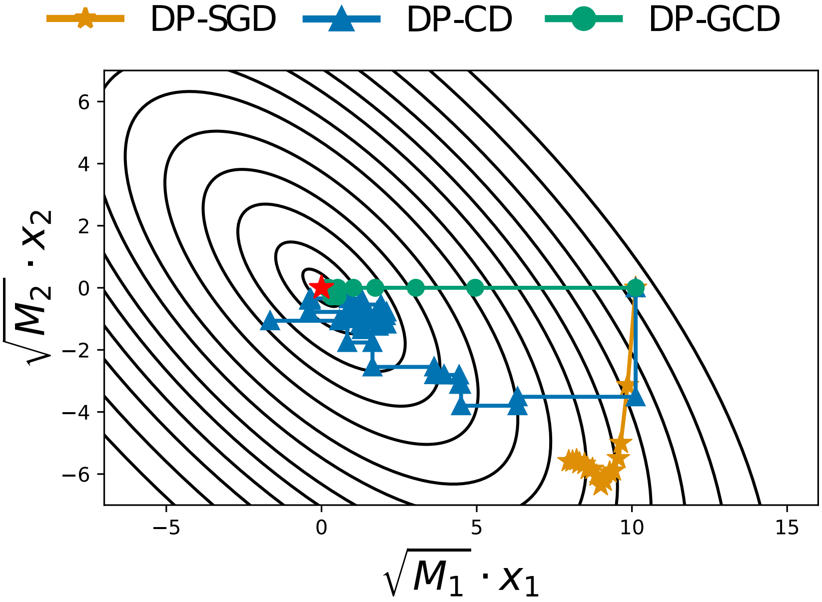
DP-GCD can focus on relevant coordinates



* Regularized logistic regression
* Standardized data
* $n = 2,600$ records
* $p = 501$ features
* $\epsilon = 1$, $\delta = 1/n^2$


DP-SGD
DP-CD
DP-GCD
6
4
2
0
−2
−4
−6
−5
0
5
10
15
$\sqrt{M_2} \cdot x_2$
$\sqrt{M_1} \cdot x_1$


DP-SGD
DP-CD
DP-GCD
6
4
2
0
−2
−4
−6
−5
0
5
10
15
$\sqrt{M_2} \cdot x_2$
$\sqrt{M_1} \cdot x_1$


DP-SGD
DP-CD
DP-GCD
6
4
2
0
−2
−4
−6
−5
0
5
10
15
$\sqrt{M_2} \cdot x_2$
$\sqrt{M_1} \cdot x_1$

Additional Results

* Utility for strongly-convex functions
* Refined lower bounds
* Proximal DP-CD and DP-GCD
* Quasi-sparse problems
* Private estimation of constants
* Clipping

Summary of this Part

Private coordinate descent methods can exploit:

* imbalance in parameter scales and variations
* imbalance/sparsity of the solution
* adapt to underlying structure

Summary of this Part

Private coordinate descent methods can exploit:

* imbalance in parameter scales and variations
* imbalance/sparsity of the solution
* adapt to underlying structure

Open questions: adaptive step sizes and clipping, better sampling of coordinates, analyze proximal greedy CD...

Contributions

* Private learning algorithms exploiting structure
  1. Imbalanced parameter scales and variations
  2. High-dimensional models with imbalanced solutions

* Study interplay between privacy and fairness
  3. Bound on the impact of privacy on fairness

Classification Problem

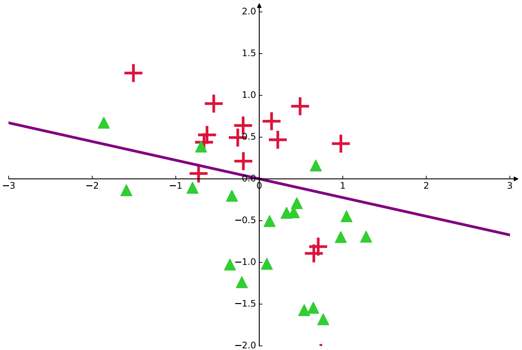
Classical Setting

Take: $\mathcal{X} \to \{-1, 1\}$

Goal: learn $h : \mathcal{X} \to \mathbb{R}$

$\to$ classify $x \in \mathcal{X}$ as

$$\hat{y} = \text{sign}(h(x))$$



Classification Problem

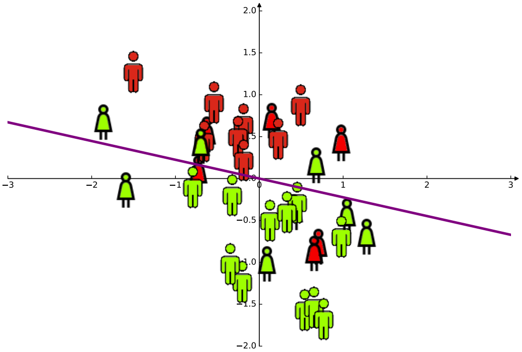
Sensitive Group $\mathcal{S}$ Setting

Take: $\mathcal{X} \times \boldsymbol{\mathcal{S}} \to \{-1, 1\}$

Goal: learn $h : \mathcal{X} \to \mathbb{R}$

$\to$ classify $x \in \mathcal{X}$ as

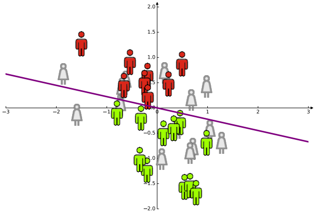
$$\hat{y} = \text{sign}(h(x))$$

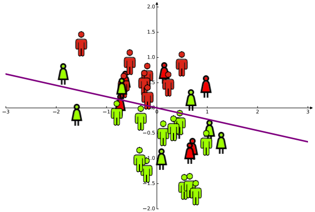


Measuring Group Fairness

Example: demographic parity*

$$F_k(h) = \mathbb{P}(h(X) > 0 | S = k) - \mathbb{P}(h(X) > 0)$$

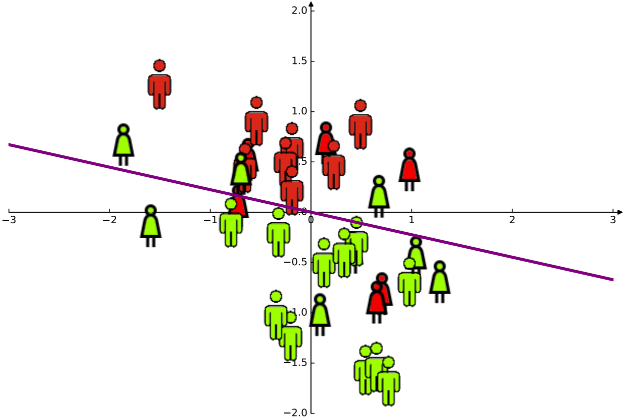




*T. Calders et al. "Building Classifiers with Independency Constraints". 2009.

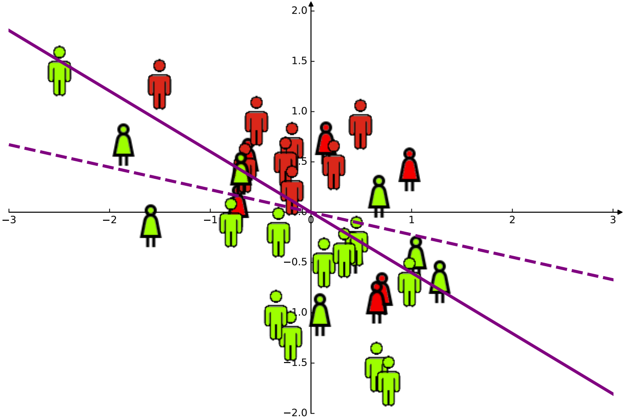
Fairness and Privacy

How much can fairness be impacted by privacy?



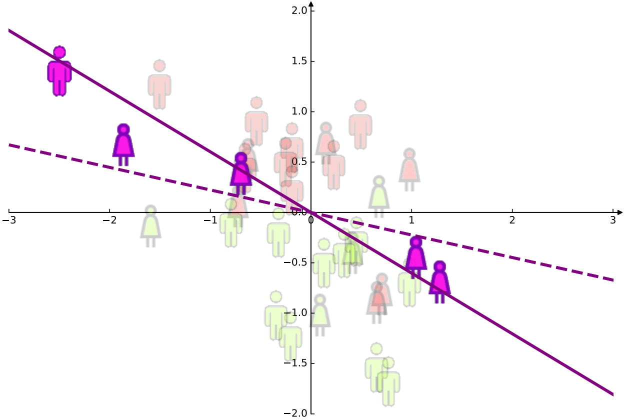
Fairness and Privacy

How much can fairness be impacted by privacy?



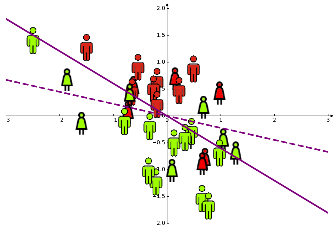
Fairness and Privacy

How much can fairness be impacted by privacy?



Fairness and Privacy

How much can fairness be impacted by privacy?



Key assumption:
confidence margin is Lipschitz

$$|h(x) - h'(x)| \leq L_{x,y} \|h - h'\|$$

for $x, y \in \mathcal{X} \times \mathcal{Y}$

Contribution 3: Privacy, Fairness*

Bound on Difference of Fairness

Difference of fairness:

$$|F_k(h) - F_k(h')| \leq \chi_k(h)\|h - h'\|$$

*P. Mangold et al. "DP Has Bounded Impact on Fairness in Classification". 2023.

Contribution 3: Privacy, Fairness*

Bound on Difference of Fairness

Difference of fairness:

$$|F_k(h) - F_k(h')| \leq \chi_k(h) \| h - h' \|$$

Where $\chi_k(h) = \mathbb{E}\left(\frac{L_{X,Y}}{|h(X)|} \middle| S = k \right) + \mathbb{E}\left(\frac{L_{X,Y}}{|h(X)|} \right)$

*P. Mangold et al. "DP Has Bounded Impact on Fairness in Classification". 2023.

Contribution 3: Privacy, Fairness*

Loss of Fairness due to Privacy is Bounded

Take $h = h^{\text{priv}}$ and $h' = h^*$:

$$|F_k(h^{\text{priv}}) - F_k(h^*)| = O\left(\chi_k(h^{\text{priv}}) \frac{\sqrt{p}}{n\epsilon} \right)$$

Where $\chi_k(h^{\text{priv}}) = \mathbb{E}\left(\frac{L_{X,Y}}{|h^{\text{priv}}(X)|} \middle| S = k \right) + \mathbb{E}\left(\frac{L_{X,Y}}{|h^{\text{priv}}(X)|} \right)$

*P. Mangold et al. "DP Has Bounded Impact on Fairness in Classification". 2023.

Contribution 3: Privacy, Fairness*

Loss of Fairness due to Privacy is Bounded

Take $h = h^{\text{priv}}$ and $h' = h^*$:

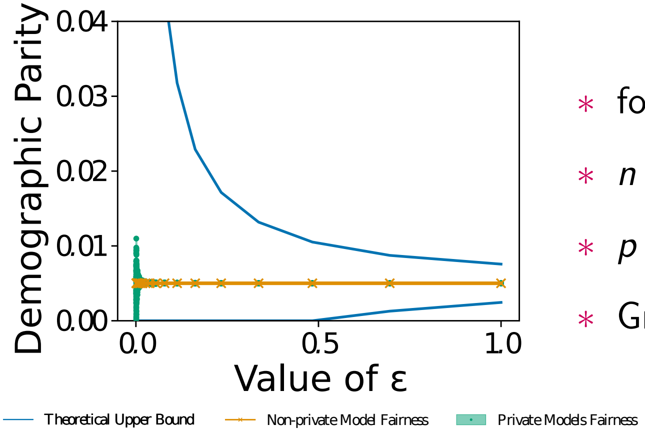
$$|F_k(h^{\text{priv}}) - F_k(h^*)| = O\left(\chi_k(h^{\text{priv}})\frac{\sqrt{p}}{n\epsilon}\right)$$

$\Rightarrow$ No need to know optimal model h^*!!

*P. Mangold et al. "DP Has Bounded Impact on Fairness in Classification". 2023.

Numerical Illustration

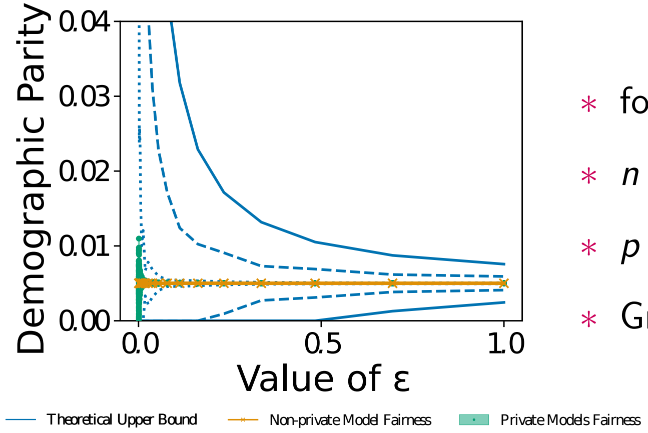
Not super tight, but meaningful!



* folktables dataset
* $n = 182,339$ records
* $p = 40$ features
* Green = private models

Numerical Illustration

Not super tight, but meaningful!



- folktables dataset
- $n = 182,339$ records
- $p = 40$ features
- Green = private models

Additional Results

* General result on conditional accuracy
* Results for other fairness measures and accuracy
* Multi-class setting

Summary of this Part

Fairness of private models:

* is “close” to the one of non-private model
* is influenced by confidence margin of the model

Open questions: use fairness-promoting methods, broader study of large-margin classifiers...

Conclusion

Structure is central to private machine learning:

* allows to improve over generic lower bounds
* can be exploited with *ad hoc* algorithms
* influences impact of privacy on fairness

More General Open Questions

* Fully adaptive private optimization algorithms
* Greedy vs. non-greedy in privacy
* Evaluating robustness of a convergence analysis
* DP mechanisms that preserve properties like fairness
* Vertical private/fair federated learning

Thank you! :)

Please ask questions!!

Publications presented in the thesis

- P. Mangold et al. "Differentially Private Coordinate Descent for Composite ERM". 2022 (ICML)
- P. Mangold et al. "High-Dimensional Private ERM by Greedy Coordinate Descent". 2023 (AISTATS)
- P. Mangold et al. "DP Has Bounded Impact on Fairness in Classification". 2023 (ICML)

Other publications

- H. Hendrikx et al. "The Relative Gaussian Mechanism and its Application to DP-GD". 2023 (working paper)
- J. O. du Terrail et al. "FLamby: Datasets and Benchmarks for Cross-Silo FL in Healthcare". 2022 (NeurIPS)
- A. Lamer et al. "Specifications for the Routine Implementation of FL in Hospitals Networks". 2021 (MIE)
- P. Mangold et al. "A Decentralized Framework for Biostatistics and Privacy Concerns". 2020 (EFMI STC)