Lecture 15 Propose-Test-Release

Yu-Xiang Wang

COMPUTER SCIENCE
UC SANTA BARBARA
Computing. ReInvented.

Logistics

- Please submit your HW2.
- The coding part should be pretty easy given my template.
  - Let me know if you run into troubles.
- HW3 will be light-weighted so you have time to work on your project.

Recap: Beyond worst-case noise in DP query release

- Global sensitivity

$$GS_q(x) = \max_{x \sim x'} \| q(x) - q(x') \|_1$$

- Local sensitivity

$$\mathrm{LS}_q(x) = \max\left\{ |q(x) - q(x')| : x' \sim x \right\}.$$

- Smooth sensitivity

$$SS_q(x,\beta) = \max\left\{ LS_q(x),\ \max_{x''} LS_q(x'') \cdot e^{-\beta d(x,x'')} \right\}$$

Recap: Admissible noise

Notation. For a subset $\mathcal{S}$ of $\mathbb{R}^d$, we write $\mathcal{S} + \Delta$ for the set $\{z + \Delta \mid z \in \mathcal{S}\}$; and $e^{\lambda} \cdot \mathcal{S}$ for the set $\{e^{\lambda} \cdot z \mid z \in \mathcal{S}\}$. We also write $a \pm b$ for the interval $[a - b, a + b]$.

Definition 2.5 (Admissible Noise Distribution). *A probability distribution on $\mathbb{R}^d$, given by a density function h, is (α, β)-*admissible *(with respect to ℓ_1) if, for $\alpha = \alpha(\epsilon, \delta), \beta = \beta(\epsilon, \delta)$, the following two conditions hold for all $\Delta \in \mathbb{R}^d$ and $\lambda \in \mathbb{R}$ satisfying $\|\Delta\|_1 \leq \alpha$ and $|\lambda| \leq \beta$, and for all measurable subsets $\mathcal{S} \subseteq \mathbb{R}^d$:*

Sliding Property: $$\Pr_{Z \sim h}\left[Z \in \mathcal{S}\right] \leq e^{\frac{\epsilon}{2}} \cdot \Pr_{Z \sim h}\left[Z \in \mathcal{S} + \Delta\right] + \frac{\delta}{2}.$$

Dilation Property: $$\Pr_{Z \sim h}\left[Z \in \mathcal{S}\right] \leq e^{\frac{\epsilon}{2}} \cdot \Pr_{Z \sim h}\left[Z \in e^{\lambda} \cdot \mathcal{S}\right] + \frac{\delta}{2}.$$

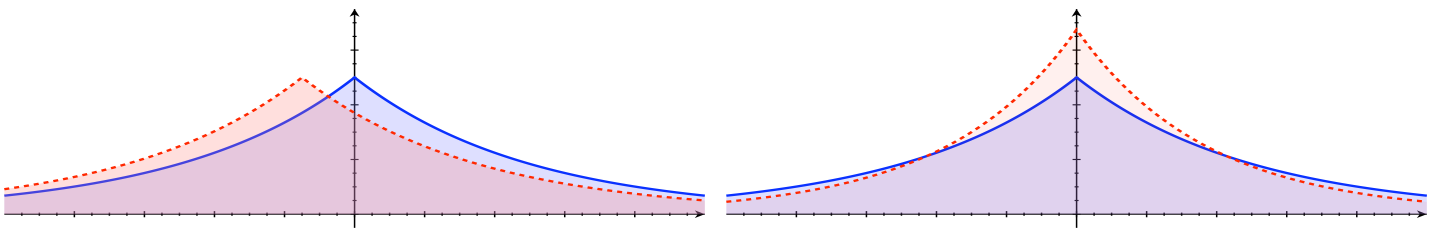


Figure 1: Sliding and dilation for the Laplace distribution with p.d.f. $h(z) = \frac{1}{2}e^{-|z|}$, plotted as a solid line. The dotted lines plot the densities $h(z + 0.3)$ (left) and $e^{0.3}h(e^{0.3}z)$ (right).

- Then $\mathcal{A}(x) = f(x) + \frac{S(x)}{\alpha} \cdot Z$ satisfies (ε,δ)-DP.

$\frac{(e^{\varepsilon} f_1)}{2}$

Recap: Summary of the noises that are known to work

- Cauchy distribution
- Student t-distribution

$\Big\}$ ε-DP

- Laplace-log-normal
- Uniform-log-normal
- Arcsinh-normal

$\Big\}$ ρ-zCDP

- Gaussian
- Laplace

$\Big\}$ (ε,δ)-DP $\quad \delta > 0$

Recap: Laplace-log-normal noise and CDP

- Adding log-normal noise

$$Z = X \cdot \underline{e^{\sigma Y}}$$

- X drawn from Laplace and Y from a standard Normal.

Proposition 3. *Let $f : \mathcal{X}^n \to \mathbb{R}$ and let $Z \leftarrow \mathsf{LLN}(\sigma)$ for some $\sigma > 0$. Then, for all $s, t > 0$, the algorithm $M(x) = f(x) + \frac{1}{s} \cdot \mathsf{S}_f^t(x) \cdot Z$ guarantees $\frac{1}{2}\varepsilon^2$-CDP for $\varepsilon = t/\sigma + e^{3\sigma^2/2}s$.*

Bun and Steinke (2019): "Average case averages": https://arxiv.org/pdf/1906.02830.pdf

This lecture

- Finish smooth sensitivity
  - Sketching the idea of the zCDP proof for Laplace log-normal.
  - Empirical results on truncated mean.

- Propose-Test-Release

- Easy-to-use recipes for PTR and examples

Reading materials

- Vadhan book Section 3.2 – 3.4
- Dwork and Lei “Differential Privacy and Robust Statistics”
  - Original paper for PTR.
- W. (2018) “Revisiting Differentially Private Linear Regression” https://arxiv.org/abs/1803.02596
  - A good example for deriving data—dependent DP algorithm

Concentrated DP analysis of Smoothed Sensitivity

- Adding log-normal noise

$$Z = X \cdot e^{\sigma Y}$$

- X drawn from Laplace and Y from a standard Normal.

> **Proposition 3.** *Let* $f : \mathcal{X}^n \to \mathbb{R}$ *and let* $Z \leftarrow \mathsf{LLN}(\sigma)$ *for some* $\sigma > 0$*. Then, for all* $s, t > 0$*, the algorithm* $M(x) = f(x) + \frac{1}{s} \cdot \mathsf{S}_f^t(x) \cdot Z$ *guarantees* $\frac{1}{2}\varepsilon^2$*-CDP for* $\varepsilon = t/\sigma + e^{3\sigma^2/2} s$.

$$D_\alpha\left(f(x)+g(x)\cdot Z \,\|\, f(x')+g(x')Z\right)$$
$$= D_\alpha\left(g(x)\cdot Z \,\|\, f(x')-f(x)+g(x')Z\right)$$
$$= D_\alpha\left(Z \,\|\, \frac{f(x')-f(x)}{g(x)} + \frac{g(x')}{g(x)} Z\right)$$

translation

Dilation

Bun and Steinke (2019): "Average case averages": https://arxiv.org/pdf/1906.02830.pdf

Summary of the noises that are known to work

- Cauchy distribution
- Student t-distribution

- Laplace-log-normal
- Uniform-log-normal
- Arcsinh-normal

- Gaussian
- Laplace

Sketch of the proof for the Laplace-Log-Normal

- Let's say for all neighboring datasets

$$\underline{|f(x) - f(x')| \leq g(x)} \qquad \text{and} \qquad \underline{e^{-t} g(x) \leq g(x') \leq e^{t} g(x)}.$$

- **Algorithm:** $M(x) = f(x) + \frac{g(x)}{s} \cdot Z \quad \text{for} \quad Z \leftarrow \mathsf{LLN}(\sigma).$

- We have that $\mathrm{D}_\alpha\left(M(x) \| M(x')\right) = \mathrm{D}_\alpha\left(Z \middle\| \frac{f(x') - f(x)}{g(x)} \cdot s + \frac{g(x')}{g(x)} \cdot Z\right).$

$$-1 \leq \cdot \leq 1 \qquad e^{-t} \leq \cdot \leq e^{t}$$

$$\leq \max \left\{ \begin{array}{c} D_\alpha(Z \| s + e^{t} Z) \\ \hline D_\alpha(s + e^{t} Z \| Z) \end{array} \right\}$$

Technical tools

$(P\|Q)\ (\sqrt{a}+\sqrt{b})^2\text{-CDP}$

$(P\|R)\ a\text{-CDP}$ $\qquad$ $(R\|Q)\ b\text{-CDP}$

- Group privacy for CDP:

Lemma 11. *Let P, Q, R be probability distributions. Suppose $\mathrm{D}_\alpha(P\|R) \le a\cdot\alpha$ and $\mathrm{D}_\alpha(R\|Q) \le b\cdot\alpha$ for all $\alpha \in (1,\infty)$. Then, for all $\alpha \in (1,\infty)$,*

$$\mathrm{D}_\alpha(P\|Q) \le \alpha\cdot(\sqrt{a}+\sqrt{b})^2 \le 2\alpha\cdot(a+b).$$

- Decompose what we want to bound

$R = Z$

$$\mathrm{D}_\alpha(\underset{P}{Z}\|\underset{Q}{e^tZ+s}) = \mathrm{D}_\alpha(\underset{P}{Z-s}\|\underset{\alpha}{e^tZ}) \le (\sqrt{a}+\sqrt{b})^2\cdot\alpha$$

$\mathrm{D}(Z-s\|Z) \le a\cdot\alpha$

$\mathrm{D}_\alpha(Z\|e^tZ) \le b\cdot\alpha$

$$\mathrm{D}_\alpha(e^tZ+s\|Z)$$

Bounding the two parts separately

$\frac{1}{2}e^{3\sigma^2}\cdot s^2 =: b$

Lemma 19. *Let* $Z \leftarrow \mathsf{LLN}(\sigma)$ *for* $\sigma > 0$. *Let* $t \in \mathbb{R}$ *and* $\alpha \in (1, \infty)$. *Then*

$$\mathrm{D}_\alpha\left(Z \middle\| e^t Z\right) \le \frac{\alpha t^2}{2\sigma^2}.$$

$\frac{t^2}{2\sigma^2} =: a$

- **Proof:**

quasi-convexity of $D_\alpha(\cdot\|\cdot)$ post-processing

$$\mathrm{D}_\alpha\left(Z \middle\| e^t Z\right) = \mathrm{D}_\alpha\left(X e^{\sigma Y} \middle\| X e^{\sigma Y + t}\right) \le \sup_x \mathrm{D}_\alpha\left(x e^{\sigma Y} \middle\| x e^{\sigma Y + t}\right) \le \mathrm{D}_\alpha\left(\sigma Y \| \sigma Y + t\right). \le \frac{\alpha t^2}{2\sigma^2}$$

x x'

$|f(x) - f(x')| \le t$

Lemma 20. *Let* $Z \leftarrow \mathsf{LLN}(\sigma)$ *for* $\sigma > 0$. *Let* $s \in \mathbb{R}$ *and* $\alpha \in (1, \infty)$. *Then*

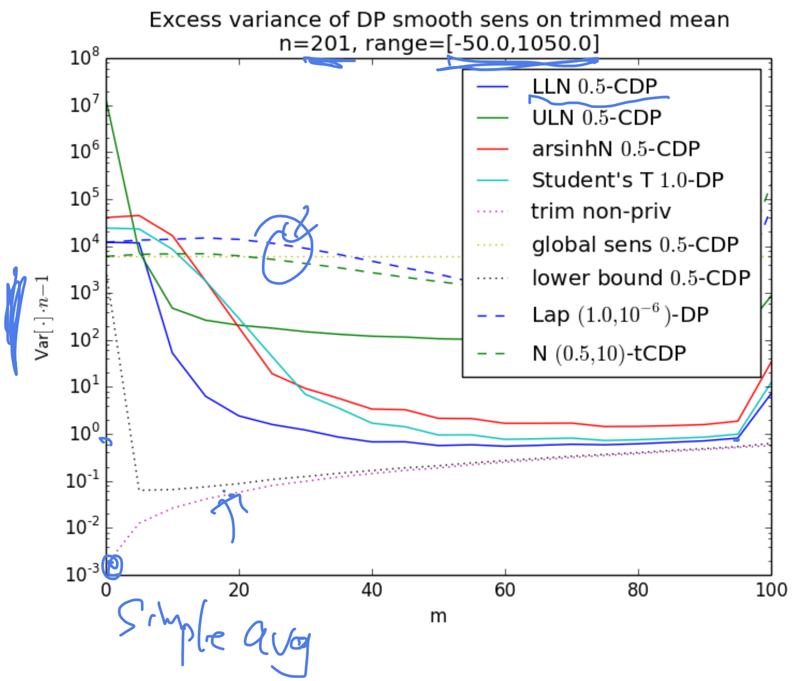
$$\mathrm{D}_\alpha\left(Z \| Z + s\right) \le \min\left\{\frac{1}{2}e^{3\sigma^2}s^2\alpha,\ e^{\frac{3}{2}\sigma^2}s\right\}.$$

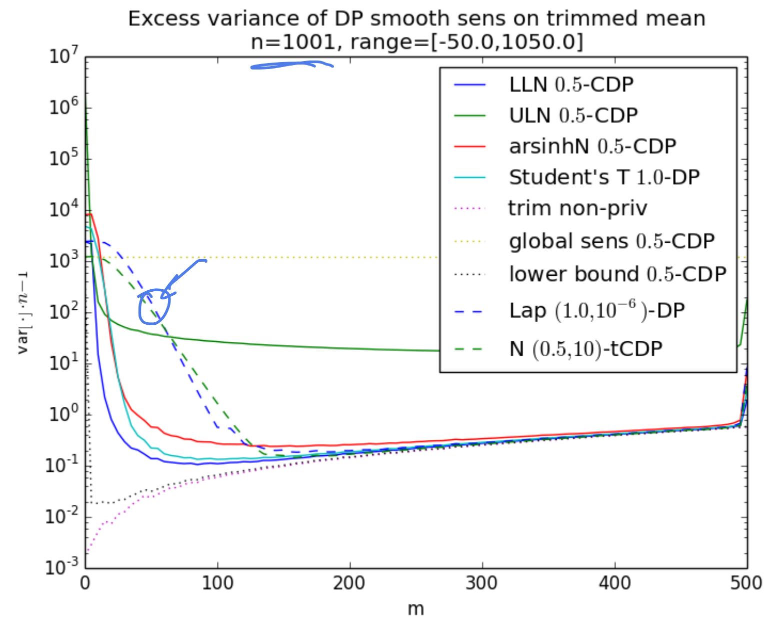
- **Proof:** $|\text{log density ratio}| \le e^{\frac{3}{2}\sigma^2}\cdot s$

ε-DP $\Rightarrow$ $\frac{\varepsilon^2}{2}$-CDP

Improvement from running smoothed sensitivity is substantial!

$$\text{trim}_m(x) = \frac{x_{(m+1)} + x_{(m+2)} + \cdots + x_{(n-m)}}{n - 2m},$$





Bun and Steinke (2019): "Average case averages": https://arxiv.org/pdf/1906.02830.pdf

Drawbacks of Smooth Sensitivity

- Restricted to numerical valued outputs.
- Requires elaborate design of the noise, generally with a much heavier tail
- Does not generalize well to high-dimension
- Are there more flexible recipes for deriving data-dependent DP algorithms?

Example: Releasing reciprocal

\# of students grade ≥ A-

- Let f(D) be a counting query, define g(D) = 1/f(D)
  - What is the global and local sensitivity of g(D)?

$\to \infty$

$$\left| \frac{1}{f(D)} - \frac{1}{f(D)-1} \right| \le \phi \qquad O\left(\frac{1}{|f(D)|}\right)$$

  - What is the smooth sensitivity of g(D)?

$LS(\tilde{D})$, $f(\tilde{D}) = 0$

$$e^{-n\beta} \cdot \infty = \infty$$

- Example: the prediction variance of linear regression on a new dataset
  - Useful for statistical inference / uncertainty quantification

$\hat{\theta}$ for dog $= (X^T X)^{-1} X^T y$

Want $\tilde{x}^T \hat{\theta}$ to $\tilde{x}^T \theta^*$

$$|\tilde{x}^T \hat{\theta} - \tilde{x}^T \theta^*| \lesssim \sqrt{\tilde{x}^T (X^T X)^{-1} \tilde{x}}$$

Examples: Private Argmax

- Voting: Who won the election?

- Model selection: Which is the best performing model when evaluating on a private dataset?

- Netflix: What is the Top-k most-popular movie last week?

$$\arg\max_{S : |S| = k} \sum_{s \in S} \text{popularity}(s)$$

$$\in \left\{ \binom{n}{k} \right\}$$

Release Stable Values without adding noise.

f: query

Define "Dist2Instability" function:

$d(x) = \min_{x'': f(x'') \neq f(x)} d(x, x'') - 1$ How many "step" (datapoints) before $x \rightsquigarrow x''$ s.t. $f(x'') \neq f(x)$

$d(x)$ is 1-insensitivity

$GS_d = 1$

"Dist2Instability":

1. $\hat{d} = d(x) + \mathrm{Lap}(\frac{1}{\varepsilon})$

2. if $\hat{d} > \frac{\log \frac{1}{\delta}}{\varepsilon}$, then output $f(x)$

   else ($\hat{d} \leq \frac{\log \frac{1}{\delta}}{\varepsilon}$), then output "$\perp$"

Case A: $f(x') \neq f(x) \Rightarrow d(x) = d(x') = 0$

All together (ε, δ)-DP

The privacy analysis of "Dist2Instability"

$(0, \delta)$-DP

- Case A:

$d(x) = 0 \Rightarrow \forall x'$ neighbor x, s.t. $f(x') \neq f(x)$ then $d(x') = 0$

$\Downarrow$

$\exists x''$ s.t. $d(x, x'') \leq 1$ and also $f(x) \neq f(x'')$

if $d(x', x'') \leq 1$, then $d(x') = \min_{f(x') \neq f(x'')} (d(x', x'') - 1) = 0$

$d(x') = \max d(x', x'')$, $f(x'') \neq f(x')$, choose $x'' = x$

x': if $f(x') = x +$

if $f(x') = f(x)$ then no conflict

if $f(x'') \neq f(x)$

x: $0 + \mathrm{Lap}(\frac{1}{\varepsilon})$

x': $0 + \mathrm{Lap}(\frac{1}{\varepsilon})$

identical except

$P(\mathrm{Lap}(\frac{1}{\varepsilon}) > \frac{\log \frac{1}{\delta}}{\varepsilon}) \leq \delta$

- Case B:

$f(x') = f(x)$

$d(x) \neq 0$

then $f(x) = f(x')$ $\forall x'$ s.t. $d(x', x) \leq 1$

output is $\perp \Leftarrow \hat{d} > \frac{\log \frac{1}{\delta}}{\varepsilon}$

post process of Lap Mech $\hat{d} = d(x) + \mathrm{Lap}(\frac{1}{\varepsilon})$

output is $f(x) \Leftarrow \hat{d} < \frac{\log \frac{1}{\delta}}{\varepsilon}$

ε-DP

Utility of “Dist2Instability”

- Perfect utility with high probability when “margin is large”

- No utility at all when the margin is small.

- Comparing to exponential mechanism
  - Homework 3 question.

Propose-Test-Release

1. Propose a bound on local-sensitivity

β

2. Test the validity of this bound

$$\hat{d} = d(x, \{x' : \mathrm{LS}_q(x') > \beta\}) + \mathrm{Lap}(1/\varepsilon),$$

3. Release: return $q(x) + \mathrm{lap}\left(\frac{\beta}{\varepsilon}\right)$ if $\hat{d} > \frac{\log 1/\delta}{\varepsilon}$

else return "$\perp$"

Proposition 3.2 (propose-test-release [33]). *For every query $q : \mathcal{X}^n \to \mathbb{R}$ and $\varepsilon, \delta, \beta \geq 0$, the above algorithm is $(2\varepsilon, \delta)$-differentially private.*

$(2\varepsilon, \delta)$-DP

$q: \mathcal{X} \to \mathbb{R}$ $\quad \mathbb{R} \cup \{\perp\}$ output space

The privacy analysis of PTR

- Case 1: $LS_q(x) > \beta \Rightarrow d(x, \{x'' : LS_q(x'') > \beta\}) = 0$

(ε, δ)-DP

$\hat{d}(x) = 0 + \mathrm{Lap}(\frac{1}{\varepsilon})$ $\quad \mathbb{P}(\underbrace{\hat{d}(x) \geq \frac{\log \frac{1}{\delta}}{\varepsilon}}_{\text{event } E}) \leq \delta$

$\mathbb{P}(M(x) \in S) = \mathbb{P}(M(x) \in S \cap E^c) + \mathbb{P}(M(x) \in S \cap E)$

$\leq \mathbb{P}(M(x) = \text{``}\perp\text{''}) + \delta$

$\leq e^{\varepsilon} \mathbb{P}(M(x') \in S \cap E^c) + \delta$

$\leq e^{\varepsilon} \mathbb{P}(M(x') \in S) + \delta$

- Case 2:

$LS(x) \leq \beta$

$|q(x) - q(x')| \leq \beta$

Laplace mech that output $M(x) \in \mathbb{R} \cup \{\perp\}$

consisting of $\hat{d}$ and $q(x) + \mathrm{Lap}(\frac{\beta}{\varepsilon})$, $q(x') + \mathrm{Lap}(\frac{\beta}{\varepsilon})$ are ε-indistinguishable

$(2\varepsilon, 0)$-DP $\quad$ post-processing of ε-DP

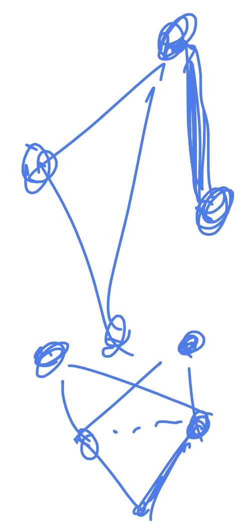
Two remaining issues with PTR

1. How do I know what bound to propose?

2. Isn't it still relying on local sensitivity and noise-adding? How does it help to go beyond releasing numerical queries?

How do I know what bound to propose? Privately releasing "a high probability bound" of local sensitivity.

- Example: Estimating the number of triangles in a graph under Edge Differential Privacy.



- Global sensitivity: n-2
- Local sensitivity: the max degree of G
- Private releasing local sensitivity?

$\Delta_{LS_q} = 1$

Privacy analysis of the approach to release local sensitivity privately.

$\Delta_f = LS_f$

Lemma: Let $\tilde{\Delta}_f(D)$ satisfies ε-DP and

$$\mathbb{P}\left[\Delta_f(D) \geq \tilde{\Delta}_f(D)\right] \leq \delta \iff \text{w.p. } 1-\delta,\ \Delta_f(D) \le \tilde{\Delta}_f(D)$$

Then $f(D) + \mathrm{Lap}(\tilde{\Delta}_f(D)/\epsilon)$ satisfies $(2\varepsilon,\delta)$-DP.

- Proof: $(y, \check{\Delta})$ $\quad \check{\Delta}$ is ε-DP, $y = f(x) + \mathrm{Lap}(\frac{\check{\Delta}}{\varepsilon})$, $E = \{\check{\Delta}_f > \Delta_f\}$

$$\mathbb{P}_x[(y,\check{\Delta}) \in S_1 \times S_2] = \mathbb{P}_x[(y,\check{\Delta}) \in S_1 \times S_2 \cap E] + \underbrace{\mathbb{P}[(y,\check{\Delta}) \in S_1 \times S_2 \cap E^c]}_{\le \delta}$$

$$\le \int_{\check{\Delta} \in S_2 \cap E} \mathbb{P}_x[y \in S_1 \mid \check{\Delta}] \cdot \mathbb{P}_x[\check{\Delta} \in S_2 \cap E]\, d\check{\Delta} + \delta$$

$$\le \int_{\check{\Delta} \in S_2 \cap E} e^{\varepsilon} \mathbb{P}_{x'}(y \in S_1 \mid \check{\Delta})\, d\check{\Delta} \cdot e^{\varepsilon} \mathbb{P}_{x'}[\check{\Delta} \in S_2 \cap E] + \delta$$

$$\le e^{2\varepsilon} \mathbb{P}_{x'}((y,\check{\Delta}) \in S_1 \times (S_2 \cap E)) + \delta \le e^{2\varepsilon} \mathbb{P}_{x'}((y,\check{\Delta}) \in S_1 \times S_2) + \delta$$

See a more general statement and proof in
Appendix G.6 of this paper: https://sites.cs.ucsb.edu/~yuxiangw/docs/spectral_privatelda.pdf

Beyond local sensitivity / noise-adding approaches

- What happens when the output space is not numerical?

- How to design data-adaptive versions of posterior-sampling, or objective-perturbation, or NoisySGD rather than just noise adding?

Topic of the next (and final) lecture

- Beyond local sensitivity
  - Per-instance differential privacy
  - pDP to DP conversion

- Data-dependent algorithms in differentially private machine learning