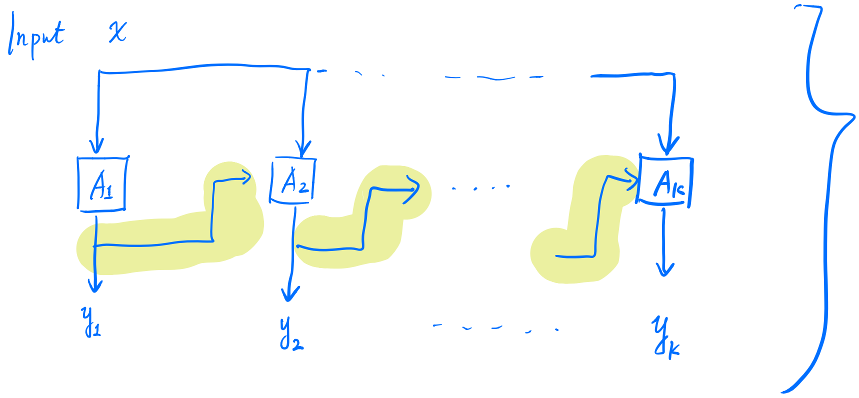
Lecture 12.

- Advanced Composition Recap
- Optimization (ERM)
  - Exponential Mechanism
- Convex sets & functions

Transition to ML

HW2 posted on Canvas.
Due ~~N~~ext Weds

Adaptive Composition.



Composition A is $(\tilde{\varepsilon}, \tilde{\delta})$-DP

If each of $A_1, \dots, A_k$ is (ε, δ)-DP

- Basic Composition : $(\tilde{\varepsilon} = k\varepsilon,\ \tilde{\delta} = k\delta)$ - DP
- Advanced Composition : $\tilde{\varepsilon} = \boxed{\varepsilon \cdot \sqrt{2k \ln\left(\frac{1}{\delta'}\right)}} + k \cdot \varepsilon \cdot \frac{e^{\varepsilon}-1}{e^{\varepsilon}+1}$, $\forall \delta' \in (0,1)$

  $\tilde{\delta} = k\delta + \delta'$

Dominant term $\sqrt{k}\,\varepsilon$

$\approx k\varepsilon^2$ for $\varepsilon \lesssim 1$

$= (\sqrt{k}\,\varepsilon)^2$

If $\sqrt{k}\,\varepsilon \lesssim 1$, $k\varepsilon^2 \leq \sqrt{k}\,\varepsilon$

Privacy Loss as a Random Variable.

Given a randomized algorithm A
inputs x & x', and output $y \in Y$.

$$I^{A}_{x,x'}(y) = \ln\left(\frac{\mathbb{P}[A(x)=y]}{\mathbb{P}[A(x')=y]}\right)$$ ← privacy loss.

ε-DP, $I^{A}_{x,x'} \in [-\varepsilon, \varepsilon]$

KL-divergence
$$\mathbb{E}_{y \sim A(x)}\left[\ln \frac{\mathbb{P}[A(x)=y]}{\mathbb{P}[A(x')=y]}\right]$$

Two random variables U, V are (ε, δ)-indistinguishable

$$U \approx_{\varepsilon,\delta} V$$

$$\forall E \subseteq Y, \quad \left.\begin{aligned} \mathbb{P}[U \in E] &\le e^{\varepsilon}\mathbb{P}[V \in E] + \delta \\ \mathbb{P}[V \in E] &\le e^{\varepsilon}\mathbb{P}[U \in E] + \delta. \end{aligned}\right\}$$

Simulation Lemma.

Original [DRV'08] [DR'14] [KOV]

textbook.

Replace $A_j(x)$, $A_j(x')$

By simpler R.V.

U , V

$U, V \in \{ \underbrace{0, 1}, \text{"U"}, \text{"V"} \}$

"Leaky Randomized Response"

	$\mathbb{P}_U$	$\mathbb{P}_V$
0	$(1-\delta)\frac{e^\varepsilon}{1+e^\varepsilon}$	$(1-\delta)\frac{1}{1+e^\varepsilon}$
1	$(1-\delta)\frac{1}{1+e^\varepsilon}$	$(1-\delta)\frac{e^\varepsilon}{1+e^\varepsilon}$
"U"	δ	0
"V"	0	δ

Simulation Lemma

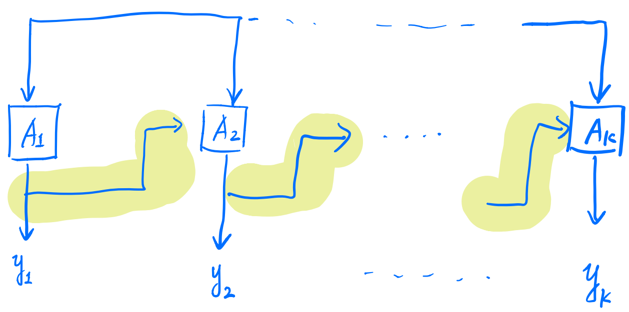
If $Y \approx_{\varepsilon,\delta} Y'$, there exists a (randomized) mapping F such that $F(U) \sim Y$ and $F(V) \sim Y'$.

$$U \longrightarrow \boxed{F} \longrightarrow Y \quad \sim A_j(x)$$

$$V \longrightarrow \boxed{F} \longrightarrow Y' \quad \sim A_j(x')$$

Using the Simulation Lemma.

Neighbors x & x'



Composition $\mathcal{A}$ is $(\tilde{\varepsilon}, \tilde{\delta})$-DP

$\forall j$, $\mathcal{A}_j(x; y_1, \dots, y_{j-1}) \underset{\varepsilon,\delta}{\approx} \mathcal{A}_j(x'; y_1, \dots, y_{j-1})$

By Sim Lemma, $\exists F_j$ s.t.

$F_j(U) \sim \mathcal{A}_j(x; y_1, \dots, y_{j-1})$

$F_j(V) \sim \mathcal{A}_j(x'; y_1, \dots, y_{j-1})$

$\Longrightarrow$ $\exists$ F^*

$F^*(U_1, \dots, U_k) \sim \mathcal{A}(x)$

$F^*(V_1, \dots, V_k) \sim \mathcal{A}(x')$

Proof Idea: suffices to show that

$$(U_1, \dots, U_k) \underset{\tilde{\varepsilon}, \tilde{\delta}}{\approx} (V_1, \dots, V_k)$$

Reduction.

If $(u_1, \dots, u_k) \approx_{\varepsilon,\delta} (v_1, \dots, v_k)$

then by post-processing $A(x) \approx_{\tilde{\varepsilon},\tilde{\delta}} A(x')$.

Intermediate Step

Lemma. $(u_1, \dots, u_k) \approx_{\tilde{\varepsilon},\tilde{\delta}} (v_1, \dots, v_k)$

for
$$\tilde{\varepsilon} = \varepsilon\sqrt{2k\ln(1/\delta')} + k\varepsilon \cdot \underbrace{\frac{e^{\varepsilon}-1}{e^{\varepsilon}+1}}_{\text{per-round expected priv loss.}}$$

$$\tilde{\delta} = k\delta + \delta'$$

Proof Idea: - Union Bound on probability of outputting "U" "V"

- Concentration inequality on privacy losses.

Proof Sketch.

Given $\vec{z} = (z_1, \dots, z_k) \in \{0, 1, \text{"U"}, \text{"V"}\}^k$

$$I(\vec{z}) = \ln \frac{\mathbb{P}[U_1 = z_1, \dots, U_k = z_k]}{\mathbb{P}[V_1 = z_1, \dots, V_k = z_k]} = \sum_{j=1}^{k} \ln \frac{\mathbb{P}[U_j = z_j]}{\mathbb{P}[V_j = z_j]}$$

"Remove" one Bad event:

$$Bad_1 = \{\vec{z} : \text{some } z_j \in \{\text{"U"}, \text{"V"}\}\}$$

$$\mathbb{P}_{z \sim U_1, \dots, U_k}[Bad_1] \le k\delta \quad \leftarrow \text{Union Bound.}$$

Condition on Bad 1 not happening. (which is the case w.p. $\geq 1 - k\delta$)

$\Longrightarrow \forall j, \; z_j \in \{0, 1\}$

$$\ln \frac{\mathbb{P}[U_j = z_j]}{\mathbb{P}[V_j = z_j]} \in \{\varepsilon, -\varepsilon\}$$

$$= \varepsilon \cdot (-1)^{z_j}$$

$$I(\vec{z}) = \sum_j \ln \frac{\mathbb{P}[U_j = z_j]}{\mathbb{P}[V_j = z_j]} = \underbrace{\varepsilon \cdot \sum_{j=1}^{k} (-1)^{z_j}}$$

Sum of i.i.d. random variable $\in [-\varepsilon, \varepsilon]$
and expectation $\varepsilon \cdot \frac{e^{\varepsilon} - 1}{e^{\varepsilon} + 1}$

Apply "Chernoff" Bound, $\forall t > 0$ → Remove Bad event 2.

$$\mathbb{P}\left[I(\vec{z}) \geq \underbrace{k \cdot \varepsilon \cdot \frac{e^{\varepsilon} - 1}{e^{\varepsilon} + 1}}_{k \cdot \text{expectation}} + \underbrace{t \cdot \varepsilon \cdot \sqrt{k}}_{\text{deviation}}\right] \le \boxed{e^{-t^2/2}}$$

→ Set to be δ'

(Private) Optimization for ML. (e.g. fitting a model)

Given a data set $\mathcal{X} = (x_1, \dots, x_n)$

loss function $\ell : C \times \mathcal{X} \to \mathbb{R}$

$C \subseteq \mathbb{R}^d$: feasible set of parameters

Empirical Risk Minimization

$$\underset{w \in C}{\text{minimize}} \quad \underbrace{L(w; x)}_{\text{Empirical risk.}} = \frac{1}{n} \sum_{i=1}^{n} \ell(w; x_i) + \underbrace{\Lambda(w)}_{\text{Optional Regularization}}$$

Find $\hat{w} \in C$ such that

$$L(\hat{w}, x) - \min_{w \in C} L(w, x) \quad \text{to be "small"}$$

Empirical Risk: $L(w; x) = \frac{1}{n} \sum_{i=1}^{n} \ell(w; x_i)$

Population Risk: $L(w; P) = \mathbb{E}_{x' \sim P} \left[\ell(w; x') \right]$

Assumption: $x_1, \dots, x_n \overset{\sim}{\text{i.i.d}} P$

Excess Population risk:

$$L(\hat{w}, P) - \min_{w \in C} L(w, P)$$

Generalization error:

$$\underbrace{L(\hat{w}, x) - L(\hat{w}, P)}$$

Holdout set

Differential Privacy $\longrightarrow$ "Reusable Holdout"

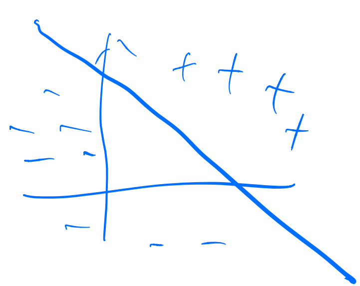
Examples of losses ℓ. or L

Mean estimation: $x_1, \dots, x_n \in \mathbb{R}$, $\ell(w, x_i) = (w - x_i)^2$

$$L(w, x) = \frac{1}{n} \sum_{i=1}^{n} (w - x_i)^2$$

Linear Regression: $(x_1, y_1) \dots, (x_n, y_n) \in \mathbb{R}^d \times \mathbb{R}$,

$$\ell(w; (x_i, y_i)) = \left(\langle w, x_i \rangle - y_i \right)^2$$



Other Examples: Support Vector Machine

Deep Learning

SVM

Input $(x_i, y_i) \in \mathbb{R}^d \times \{\pm 1\}$

$$L(w; x) = \frac{1}{n} \sum_{i=1}^{n} \left(1 - y_i \langle w, x_i \rangle\right)_+$$

$$+ \lambda \|w\|_2^2$$

DP ERM. (as a selection Problem)

First algorithm : exponential Mechanism

$\ell : C \times X \mapsto [0, \Delta]$

$L(w; x) = \frac{1}{n} \sum_{i=1}^{n} \ell(w; x_i)$

Satisfy $(\varepsilon, 0)$-DP

Exp Mech :

Sample $\hat{w} \in C$ with probability

$\mathbb{P}[\hat{w} = w] \propto \exp\left(\frac{-\varepsilon n}{2\Delta} \underbrace{L(w, x)}_{\text{"quality score"}} \right)$

"proportional to"

Efficient Sampler?

Polynomial-time for convex ℓ.

" log-concave distribution"

Theorem. Let $C = \{w \in \mathbb{R}^d : \|w\|_2 \leq R\}$

and ℓ is G-Lipschitz: $\forall w, w', x$
$|\ell(x,w) - \ell(x,w')| \leq G \cdot \|w - w'\|_2$

Run EM with sensitivity $\Delta = G \cdot R$

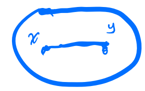
then

$$\mathbb{E}\left[L(\hat{w};x) - \min_{w \in C} L(w;x)\right]$$

$$= O\left(GR \cdot \boxed{\frac{d}{\varepsilon n}} \log\left(\frac{\varepsilon n}{d}\right)\right)$$

"Small" if $n \gg d$

Convexity. (Sets and functions)

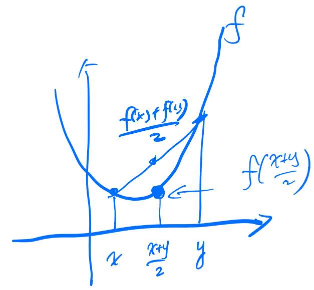


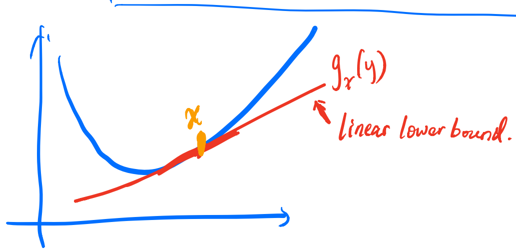
- $C \subseteq \mathbb{R}^d$ is convex if $\forall x, y \in C$, $t \in [0,1]$

$$\underbrace{t \cdot x + (1-t) \cdot y}_{\text{"line Segment"}} \in C$$

- $f : C \longrightarrow \mathbb{R}$ is convex if $\forall x, y \in C$

$$\boxed{f\left(\frac{x+y}{2}\right) \leq \frac{f(x)+f(y)}{2}}$$





$\forall x$,

affine function

$$g_x(y) = f(x) + \langle y - x, \underbrace{\nabla f(x)}_{\text{Gradient}} \rangle$$