Private Learning / Optim

work in progress with

Etienne Boursier

ENS Paris-Saclay

and Criteo AI Lab
Privacy?

* Differential privacy:
  "cannot retrieve specific info in databases"

* Private learning:
  "use/exploit some private info about you without revealing too much"
Privacy for users

* I want to visit websites I like... without Google knowing exactly how depraved I am.

* I want to watch movies on Netflix without being classified as white/male/30's.
Film fans see red over Netflix ‘targeted’ posters for black viewers

The streaming service’s customers say they are being duped by marketing that shows minor cast members as leading characters.

▲ Set It Up is made to look like a two-hander between Taye Diggs and Lucy Liu, rather than the white couple. Photograph: Twitter Kelly Quantrill @codetrill
Privacy for a company

* Start-up wants to use Deep Learning/High Perf Comp.
  → Send data to Amazon/Microsoft/IBM...
  → They do the computations.
  → Start-up gets the result.
Another Concrete Example

* Criteo is buying “ad slots” via auctions

* These auctions are run by Google

* The major competitor of Criteo is Google
A simple model

- \( \max_{x \in \mathbb{R}^d} \ x^T A_k y \)
- \( y \in \mathbb{R}^n \)
- \( A_k \in \mathbb{R}^{d \times n} \)

- \( k \in \{1, \ldots, K\} \) is the private type

prior \( \pi_k \in \Delta_k \), \( k \neq 0 \).

- Privacy value is the amount of info leaked on \( k \).

\[ \xi_k \leq KL (\pi_0, \pi_k) \]
What is the posterior $\pi_1$?

$$\max_{x \in X} z^T A_k y \quad , \quad k \rightarrow k_0$$

- Given \( k \in \{1, \ldots, K\} \), choose \( x \sim \mu_k \in \mathfrak{P}(x) \)

- \( \pi_1 \in \Delta_k \) posterior knowing \( x \)

$$\pi_1^k \mid x = \frac{\pi_0^k \mu_k(x)}{\sum_{j=1}^{\infty} \pi_0^j \mu_j(x)} \quad , \quad \text{Bayes}$$
Private learning obj.

$$\max_x \sum_{k=1}^{K} \pi_k^n \mathbb{E}_{x \sim \mu_k} \left[ x^T A_k y - \lambda \text{KL}(\pi_{x|y}, \pi_0) \right]$$

more generally

$$\inf_{\gamma \in \mathcal{P}(X \times [k])} \int c(x, k) + \lambda \mathcal{D}(\pi_{x|y}, \pi_0) \, d\gamma(x, k)$$

$$\pi, \# \gamma = \pi_0$$
The f-divergence case

\[ f \text{ convex} \quad f(1) = 0 \]

\[ D (p, q) = \mathbb{E}_{X \sim q} \left[ f \left( \frac{p(x)}{q(x)} \right) \right] \]

\[ KL (q, p) = \mathbb{E}_{X \sim q} \left[ \log \left( \frac{p(x)}{q(x)} \right) \right] \]

\[ KL (p, q) = \mathbb{E}_{X \sim q} \left[ \frac{p(x)}{q(x)} \log \left( \frac{p(x)}{q(x)} \right) \right] \]

\[ TV (p, q) = \mathbb{E}_{X \sim q} \left[ \frac{1}{2} \left| \frac{p(x)}{q(x)} - 1 \right| \right] \]
Convexity result

\[ f \text{ convex, } f(1) = 0 \]

\[ \inf_{x \in \Delta(X \times k)} \sum_{x \times k} c \, d\xi + \lambda \sum_{x \times k} E \left( \frac{d \pi_{12} \, d\xi}{d \pi_0} \right) \, d\xi \]

is convex in \( \gamma \).
Finiteness result

If \( k \) is finite (\( \text{card} \ k \)).

\[ \forall \varepsilon > 0, \exists \varepsilon \text{-optimal } y \text{ with finite support } k(k+2) \]

\( X \) compact + \( c(\cdot, k) \) lsc true for \( \varepsilon = 0 \)
Finite reformulation?

\[
\inf \sum_{i, k} y_{i, k} c(x_i, k) + \sum_{i, k, j} y_{i, k} \pi_{j, i} \prod_{i_{j}} \frac{\sum_{\rho} y_{i_{\rho}, \rho}}{\pi_{j, i_{\rho}}}
\]

\(\forall \in \mathbb{R}^{(k+2)k} \times \mathbb{R}^{k+2} \times \mathbb{R}\)

\(\sum_{i} y_{i, j} = \pi_{j, i}^{0}\)

No longer convex!
Ex: the linear case

\[ c(x, k) = \omega_k^T x + \beta_k, \quad x \in [-1, 1]^d \]

\[
\inf_{\tilde{\chi}} \sum \| \sum_{k, i, j} \omega_{k, i, j} \| + \sum \pi_{i, k} \beta_k + 2 \sum \chi_{i, k, j} \prod \left( \frac{\chi_{i, j}}{\pi_{i, k} \sum \chi_{i, j}} \right)
\]

\[ = -G(\chi) + F(\chi) \]

\[ \rightarrow \text{Difference of convex} \]
KL-divergence and OT

\[ \inf_{\gamma \in \mathcal{P}(X \times \{k\})} \left\{ c(x, k) + \lambda \int \mathbb{E} \log \left( \frac{d\pi_x}{d\gamma} \right) d\gamma(x, k) \right\} \]

\( \| \) (\( \nu \) is the prior)

\[ \inf_{\nu \in \mathcal{P}(X)} \inf_{\pi \in \mathcal{T}(\nu, \nu)} \left\{ c d\pi + \lambda \int \log \left( \frac{d\pi}{d\nu d\nu} \right) d\pi \right\} \]

\( \text{OT}_{c, \lambda} (\nu, \nu) \)
Regularized OT

$$\mathcal{OT}(\mu, \nu) = \inf_{\pi \in \mathcal{T}(\mu, \nu)} \int c \, d\pi + t \, \int \log \left( \frac{d\pi}{d\mu d\nu} \right) d\pi$$

\[\text{Let solve with Sinkhorn (in } \pi)\]

\[\text{Let optimize (in } \mu)\]
Parametric Family

\[
\min_{\nu \in \mathcal{P}(x)} \quad \text{OT}_{\epsilon, \lambda}(\mu, \nu)
\]

So look for \( \mu_\theta = \sum_{j=1}^{n} \xi_j(\theta) \delta_{\chi_j(\theta)} \)

Compute \( \frac{\partial}{\partial \xi} \text{OT}_{\epsilon, \lambda}(\mu, \nu) \) and \( \frac{\partial}{\partial \lambda} \text{OT}_{\epsilon, \lambda}(\mu, \nu) \)

(autodual diff - dual)
Applications

Private LP: \( c(x,k) = x^T w_k, \)
\( v = \sum_{i=0}^{k} \delta w_k \)

Private Regression: \( v = \sum_{i=0}^{k} \delta \beta_k \)

\[ c(A_k, A) = \| (x^T A^T A x)^T - (x^T A_k A_k x)^T \|_2 \]

\( y = A_k x + \epsilon \); \((y, x)\) public; \( A_k\) private
So many questions...

* Statistical guarantees
* Computational issues
* Private optimization algo
  (query AWS repeatedly)
* General f-divergence