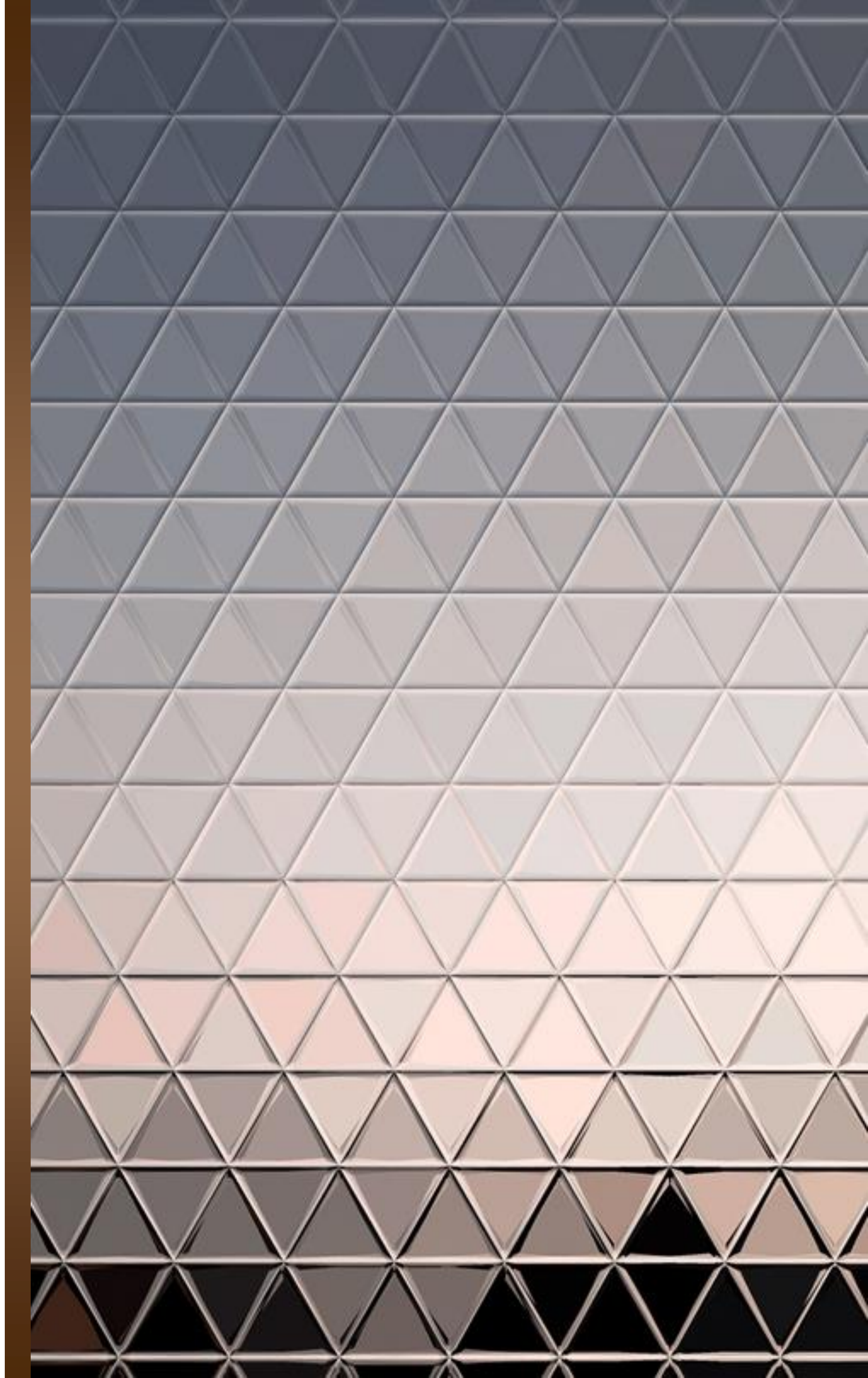
J.P.Morgan

Comparing Apples to Oranges: Learning Similarity Functions for Data Produced by Different Distributions

Leonidas Tsepenekas, Ivan Brugere, Freddy Lecue, Daniele Magazzeni

AI Research, JPMorganChase



Individual Fairness

- Introduced by Dwork et al. (Fairness through Awareness, ITCS 2012)

Similar individuals should be treated similarly

- How can you define similarity between individuals?
  i. For every two elements x, y you are given $\sigma(x, y) \in [0,1]$
  ii. The smaller $\sigma(x, y)$ is the more similar the elements

- The similarity function is always assumed given

Main Obstacle and Prior Work

- Similarity scores are not trivial to obtain (even raised in Dwork et al.)
  - Deferred to third parties
  - Ideally should be learned

- C. Ilvento. (Metric learning for individual fairness, FAccT 2019) learns similarity scores through the use of oracle queries
  - **Assumption:** The $\sigma(x, y)$ form a metric space

- Mukherjee et al. (Two simple ways to learn individual fairness metrics from data, ICML 2020)
  - Learns a specific metric function

- Wang et al. (An empirical study on learning fairness metrics for compas data with human supervision, 2019)
  - Purely empirical and focuses on specific metrics

Our Setting

- **Starting point:** Learning similarities is sometimes easy and other times hard
  - **Easy:** Comparing homogeneous data; same “demographic” group (equivalently data produced by the same distribution)
  - **Hard:** Comparing heterogeneous data; different demographics
  - **E.g.:** PhD admissions. Comparisons for students from different universities are hard

- Feature space $\mathcal{I}$. γ “demographic” groups, where each $\ell \in [\gamma]$ is governed by a distribution $\mathcal{D}_\ell$. $x \sim \mathcal{D}_\ell$ denotes an element $x \in \mathcal{I}$ that is randomly drawn from $\mathcal{D}_\ell$. The support of each distribution corresponds to the members of the group.
- For each $\ell \in [\gamma]$ there exists a given **metric** similarity function $d_\ell\colon \mathcal{I}^2 \mapsto [0,1]$.
- For every distinct ℓ, ℓ' there exists an unknown similarity function $\sigma_{\ell,\ell'}\colon \mathcal{I}^2 \mapsto [0,1]$.

Computational Goal

Goal of Our Problem: We want for any two groups ℓ, ℓ' to compute a function $f_{\ell,\ell'} : \mathcal{I}^2 \mapsto \mathbb{R}_{\geq 0}$, such that $f_{\ell,\ell'}(x, y)$ is our estimate of similarity for any $x \in \mathcal{D}_\ell$ and $y \in \mathcal{D}_{\ell'}$. Specifically, we seek a PAC (Probably Approximately Correct) guarantee, where for any given accuracy and confidence parameters $\epsilon, \delta \in (0, 1)$ we have:

$$\Pr_{x \sim \mathcal{D}_\ell, y \sim \mathcal{D}_{\ell'}} \left[\left| f_{\ell,\ell'}(x, y) - \sigma_{\ell,\ell'}(x, y) \right| > \epsilon \right] \leq \delta$$

Tools for learning:

i. For each ℓ a set S_ℓ of i.i.d. samples from $\mathcal{D}_\ell$
ii. Access to an expert oracle. You provide the oracle with $x \in \mathcal{D}_\ell$ and $\mathrm{y} \in \mathcal{D}_{\ell'}$, and it returns $\sigma_{\ell,\ell'}(\mathrm{x}, \mathrm{y})$

Objectives:

i. Polynomial number of samples
ii. Minimum queries

Results

- Algorithm with provable PAC guarantees:
    i. Almost optimal error probability (no free lunch theorem)
    ii. Almost optimal number of queries (lower bound on queries required)
    iii. Experimental validation