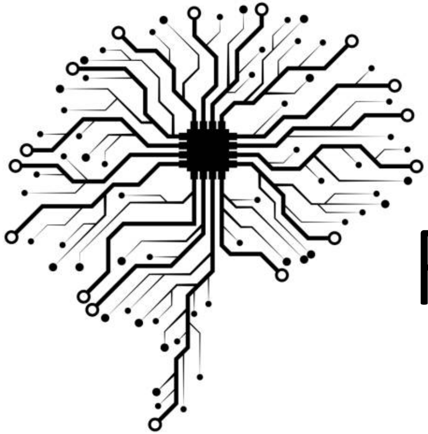


Fairness Through Awareness

Presented By Amir H. Malekijoo

Fairness Through Awareness

What is the definition of Fairness?
How do we protect sensitive information from MISUSE?
Which Domain? Advertising, Insurance, Health Care, etc.
ex. Ethnic, religious, medical, geographic

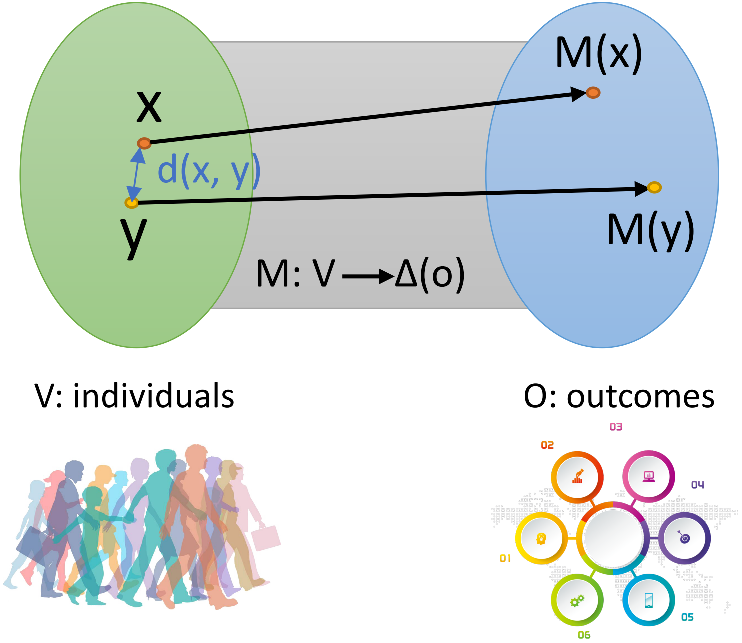
P.S. In this presentation, an overview of the paper entitled "Fairness Through Awareness" is given [1]. (all credit goes to authors of the paper)

Sample Space

Flip a coin: {H, T} 

Space are Discrete and Continuous

Every space V has possible metrics d(x, y)
How do we prevent sensitive information that



1. Introduction

Discriminations

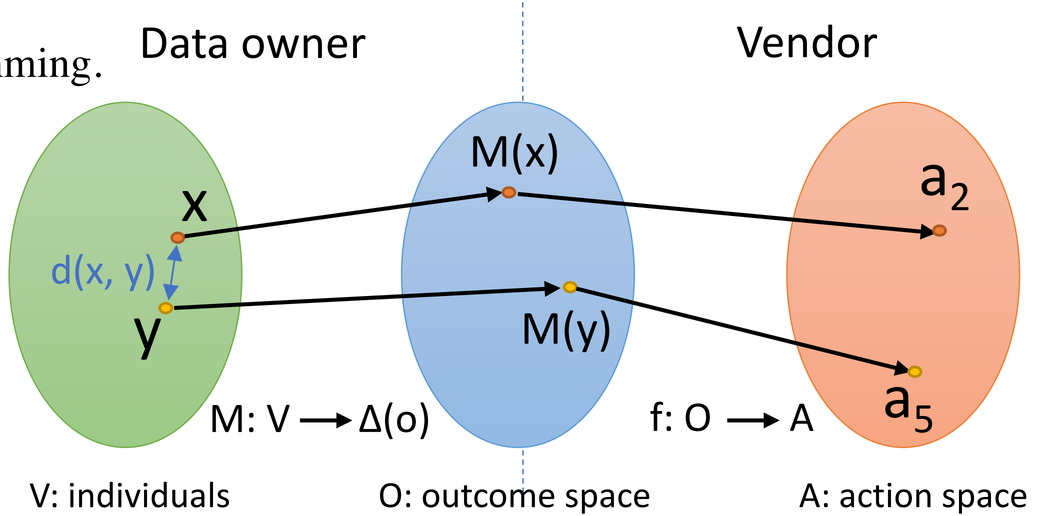
Existing Discrimination examples

- Redlining
- Reverse Tokenism
- Self-fulfilling Prophecy
- Blatant Explicit Discrimination
- Disproportionately High Segmentation (G Redlining)
- Discrimination Based on Redundant Encoding
  - ✓ Fairness Through Blindness

Objective

The following Objective is a linear Programming.

Ultimate Goal: Achieving Fairness
Goal1: Achieving Fairness in Classification
Goal2: Minimize Vendor's Loss Function



$$\mathrm{opt}(\mathcal{I}) \stackrel{\mathrm{def}}{=} \min_{\{\mu_x\}_{x\in V}} \mathop{\mathbb{E}}_{x\sim V} \mathop{\mathbb{E}}_{a\sim\mu_x} L(x,a)$$

$$\text{subject to} \quad \forall x, y \in V, : \quad D(\mu_x, \mu_y) \le d(x,y)$$

$$\forall x \in V : \quad \mu_x \in \Delta(A)$$

2. Key Elements

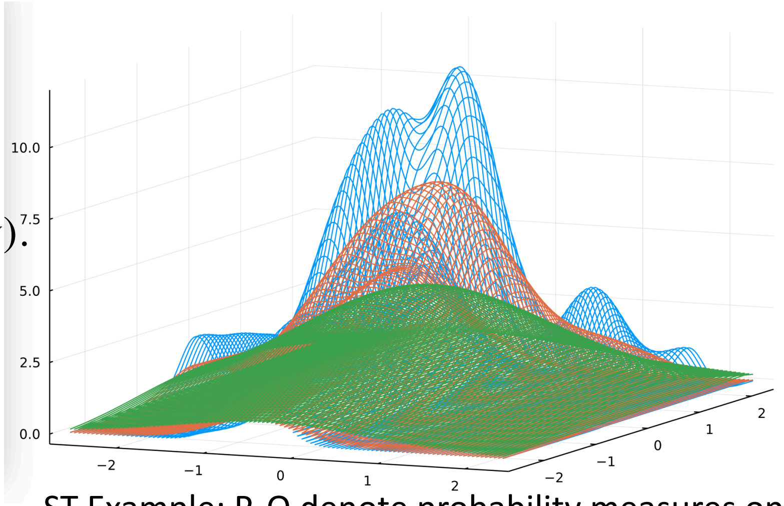
Distance Metrics

A metric on a set X is a function d : X × X → R For all x, y, z in X, this function is required to satisfy the following conditions:

d(x, y) ≥ 0 (non-negativity)
d(x, y) = 0 iff x = y (identity of indiscernible)
d(x, y) = d(y, x) (symmetry)
d(x, z) ≤ d(x, y) + d(y, z) (subadditivity/triangle inequality).

$$D_{\mathrm{tv}}(P, Q) = \frac{1}{2} \sum_{a \in A} |P(a) - Q(a)| .$$

$$D_{\infty}(P, Q) = \sup_{a \in A} \log \left(\max \left\{ \frac{P(a)}{Q(a)}, \frac{Q(a)}{P(a)} \right\} \right) .$$



ST Example: P, Q denote probability measures on a domain A. A= {0,1}
P(0) = P(1) = ½, Q(0) = ¾, Q(1) = ¼
D(P, Q) = ¼

Note: Lipschitz condition is linear when using statistical distance

Group Fairness Approach

Statistical Parity = Group Fairness, Demographic parity
Equalize two groups S, T at the level of outcomes

Pr[outcome o | S] = Pr [outcome o | T]

mapping M : V → Δ(A) satisfies statistical parity between distributions S and T up to bias ε if

$$D_{tv}(\mu_S , \mu_T) \leq \varepsilon$$

SP neutralizes redundant encodings.
Discrimination That SP will not cover.

1. fairness through blindness
2. Self-fulfilling prophecy.
3. Subset Targeting.

Demographic parity, also referred to as **statistical parity**, **acceptance rate parity** and **benchmarking**. A classifier satisfies this definition if the subjects in the protected and unprotected groups have equal probability of being assigned to the positive predicted class.
Statistical parity is a natural way to model equity: members of each group have the same chance of receiving the favorable output.

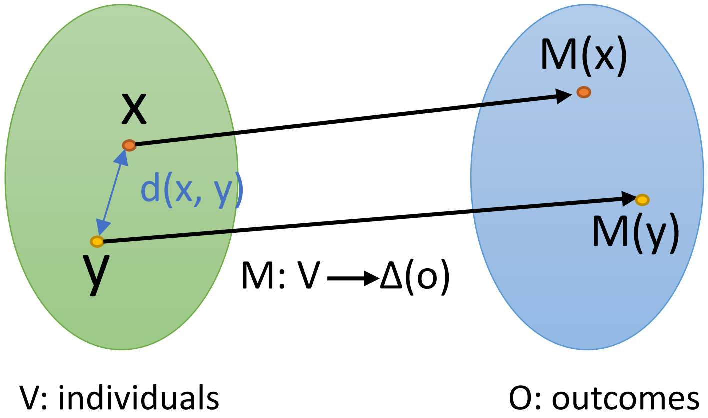
Individual Fairness

Treat similar individuals similarly
We formulize the above sentence with Lipschitz Condition

Metric d: $V \times V \longrightarrow \mathbb{R}^+$

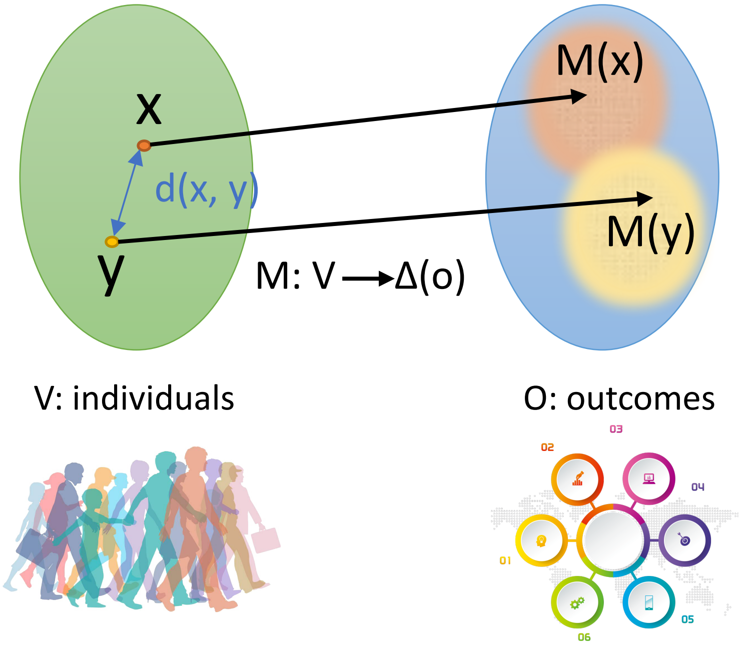
Lipschitz condition:

$$|M(X) - M(y)| \leq d(x, y)$$



Distributional Outcomes

Outcomes can be Distribution not Discrete.



Earthmover Distance

Let σ : V × V → ℝ+ (nonnegative distance function)
The σ-Earthmover distance between two distributions S and T, LP:

$$
\begin{aligned}
\sigma_{\mathrm{EM}}(S,T) \stackrel{\text{def}}{=} \quad & \min \quad \sum_{x,y \in V} h(x,y)\sigma(x,y) \\
& \text{subject to} \quad \sum_{y \in V} h(x,y) = S(x) \\
& \qquad\qquad\quad \sum_{y \in V} h(y,x) = T(x) \\
& \qquad\qquad\quad h(x,y) \geq 0
\end{aligned}
$$

3. Overview

1. LiP vs. SP.

statistical parity is insufficient as a general notion of fairness.

- When Lipschitz condition implies statistical parity between two distributions S and T on V?
- What extent to which any Lipschitz mapping can violate statistical parity?

1. LP vs. SP.

Def 3.2: $\text{bias}_{D,d}(S,T) = \max \mu_S(0) - \mu_T(0)$,

Lemma 3.1. Let $D \in \{D_{tv}, D_\infty\}$ and let $M : V \rightarrow \Delta(A)$ be any (D, d)-Lipschitz mapping. Then, M satisfies statistical parity between S and T up to biasD,d (S , T).

Using Linear Programming and its duality interpretation the following theorem will be proved.

Theorem 3.3:

$$\text{bias}D_{tv,d}(S,T) \leq d_{EM}(S,T).$$

Conclusion: when S and T are "similar" then the Lipschitz condition implies statistical parity

P.S. biasDtv,d(S,T) is a relaxation of biasD∞,d(S,T). biasD∞,d(S,T) ≤ biasDtv,d(S,T).
So D∞ can be proved in a same way as Dtv . Hence, every (D∞, d)-Lipschitz mapping is also (Dtv , d)-Lipschitz.

2. Differential Privacy

fairness is a generalization of the notion of differential privacy

database curator maintains a database $x \subseteq U$
data analyst is allowed to ask a query $F: V \rightarrow A$ on the database.

$M : V \rightarrow \Delta(A)$ satisfies ε-differential privacy iff M satisfies the $(D\infty, d)$-Lipschitz property.

Section 5:
Theorem: Fairness mechanism with "high utility" in metric spaces (V,d) of bounded doubling dimension

$|B(x,R)| \leq O(|B(x,2R))$

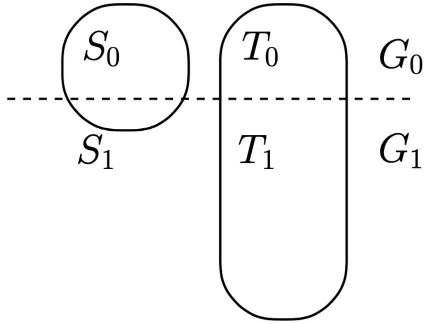
Based on exponential mechanism

3. Fair Affirmative Action

3. Fair Affirmative Action

- What if we want to ensure statistical parity between two groups S and T, but members of S are less likely to be "qualified"?
- What if earthmover distance large?
- Here we consider the complementary case where S and T are very different and imposing statistical parity corresponds to preferential treatment.

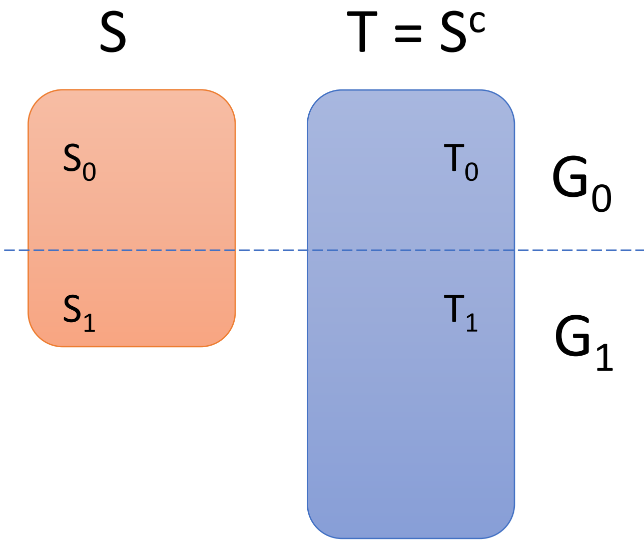
For simplicity, let T = S^c.



3. Fair Affirmative Action

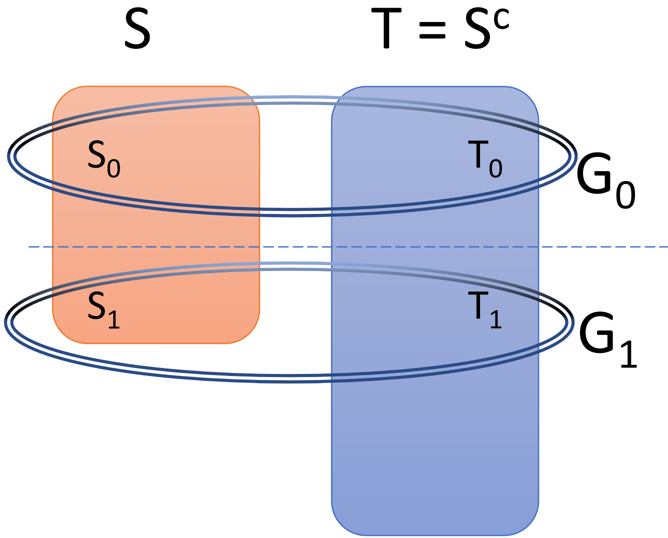
- What if we want to ensure statistical parity between two groups S and T, but members of S are less likely to be "qualified"?
- Or equally, What if earthmover distance large?
- Here we consider the complementary case where S and T are very different and imposing statistical parity corresponds to preferential treatment.

For simplicity, let $T = S^c$.



3. Fair Affirmative Action

1. All in Gi treated the same

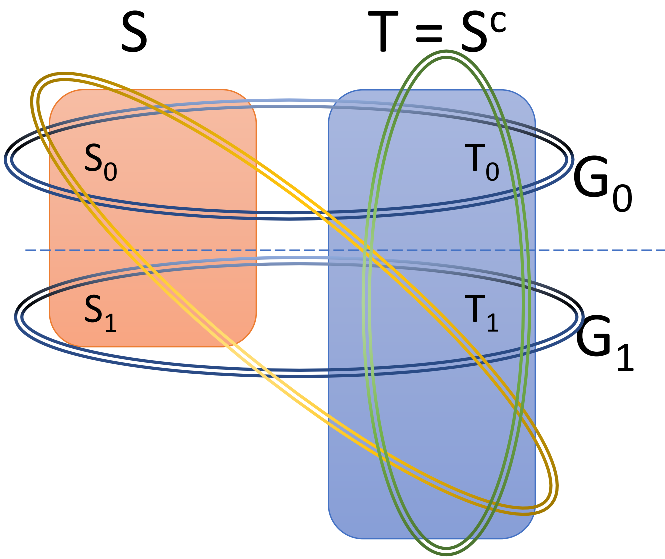


3. Fair Affirmative Action

1. All in G_i treated the same

Statistical Parity, the probability of S_0 must be the same as T_1
Unless the distribution is unbalanced.

Imposing Statistical Parity. (Collapse)
(Lipschitz Condition makes it collapse)(solution: assigning a fixed probability distribution) Yellow oval
Failing to Impose Statistical Parity. (blatant discrimination)
(only T is chosen by vendor) Green oval



3. Fair Affirmative Action

1. (a) First we compute a mapping from elements in S to distributions over T which transports the uniform distribution over S to the uniform distribution over T, while minimizing the total distance traveled. Additionally the mapping preserves the Lipschitz condition between elements within S.

   (b) This mapping gives us the following new loss function for elements of T: For $y \in T$ and $a \in A$ we define a new loss, $L'(y, a)$, as

   $$L'(y,a) = \sum_{x \in S} \mu_x(y) L(x,a) + L(y,a)\,,$$

   where $\{\mu_x\}_{x \in S}$ denotes the mapping computed in step (a). L' can be viewed as a reweighting of the loss function L, taking into account the loss on S (indirectly through its mapping to T).

2. Run the Fairness LP only on T, using the new loss function L'.

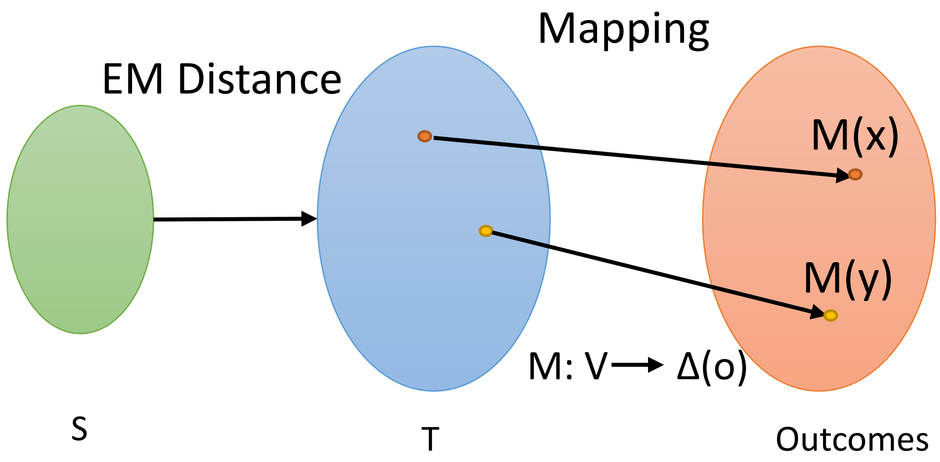
3. Fair Affirmative Action

$$
\begin{aligned}
d_{\mathrm{EM+L}}(S,T) \stackrel{\mathrm{def}}{=} \quad \min \quad & \mathop{\mathbb{E}}_{x\in S}\ \mathop{\mathbb{E}}_{y\sim\mu_x} d(x,y) \\
\text{subject to} \quad & D(\mu_x,\mu_{x'}) \le d(x,x') \quad \text{for all} \quad x,x' \in S \\
& D_{\mathrm{tv}}(\mu_S, U_T) \le \epsilon \\
& \mu_x \in \Delta(T) \quad \text{for all} \quad x \in S
\end{aligned}
$$

$$
M(x) = \begin{cases} \nu_x & x \in T \\ \mathbb{E}_{y\sim\mu_x} \nu_y & x \in S \end{cases} .
$$

3. Fair Affirmative Action

Trade-off between Imposing and failure to impose SP



Summary

Summary

- Group fairness
- Individual fairness
- Relationship between GF & IF
- Group fairness does not imply individual fairness
- Individual fairness implies group fairness if earthmover distance small
- Fair Affirmative Action

References

[1] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, Richard Zemel, ”Fairness Through Awareness”, arXiv:1104.3913v2, (2011). https://doi.org/10.48550/arxiv.1104.3913

Fair Use Notice