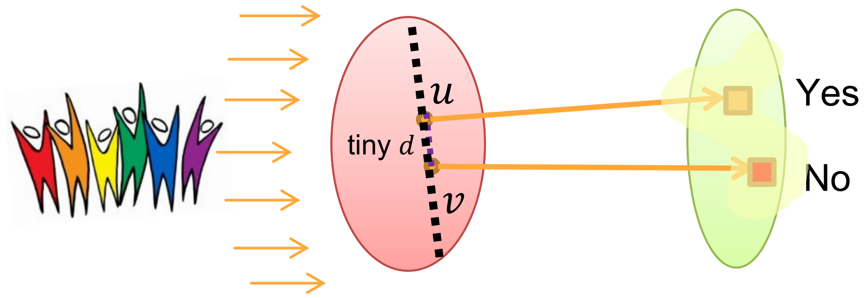
Composition of Metric-Fair Algorithms

Recall: Individual (aka Metric) Fairness

"Similar people" have similar *probabilities* of "Yes" and "No" outcomes



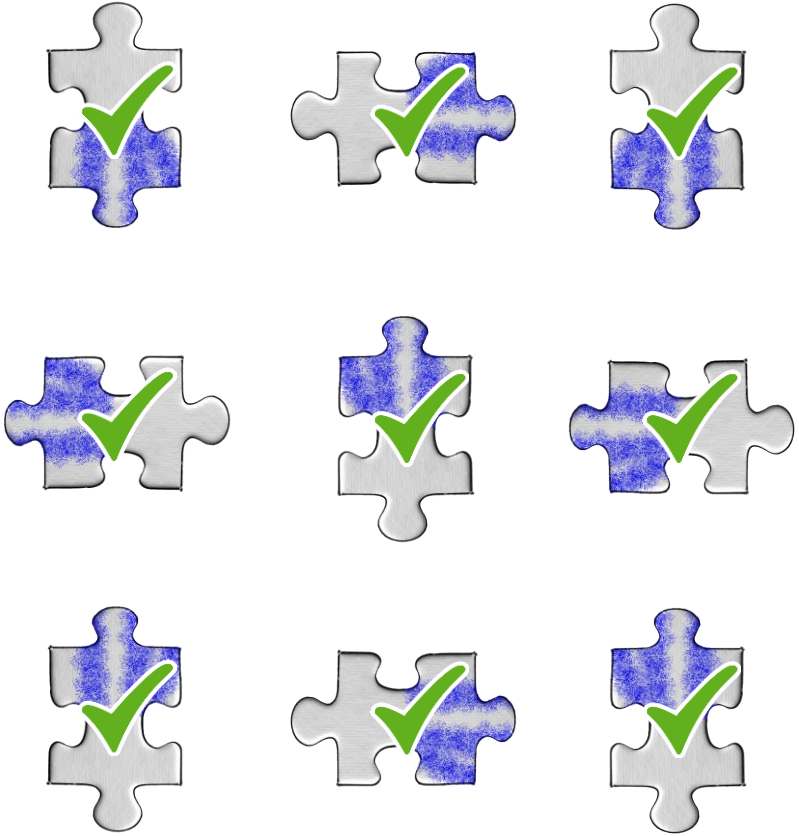
$$C: U \to \Delta(O)$$
$$\left|\left|C(x) - C(y)\right|\right| \leq d(x, y)$$

Pop Quiz

- Does Individual Fairness address the equal FPR, FNR, PPV problem?

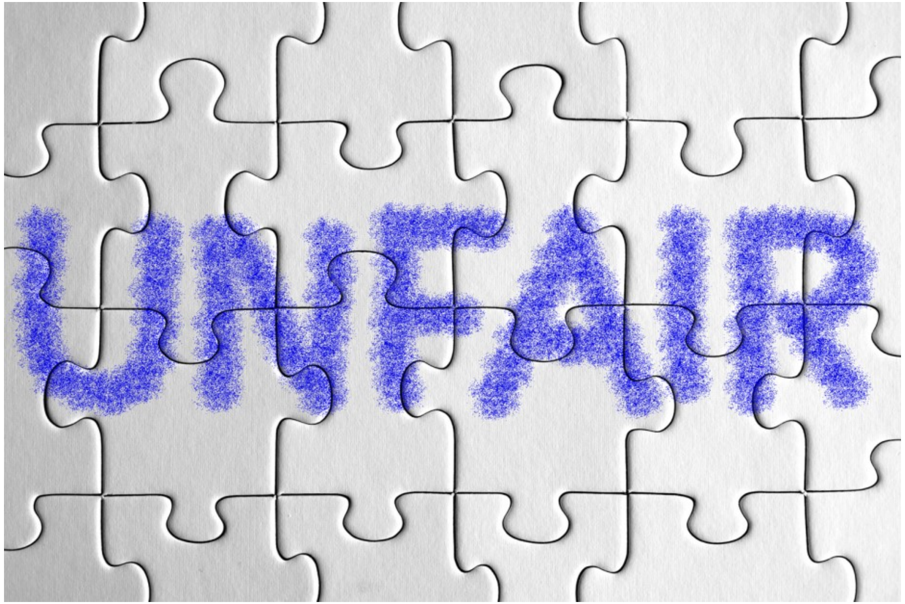
Intuition

If all of the parts are fair, then the whole should be fair.



Reality

It's complicated.



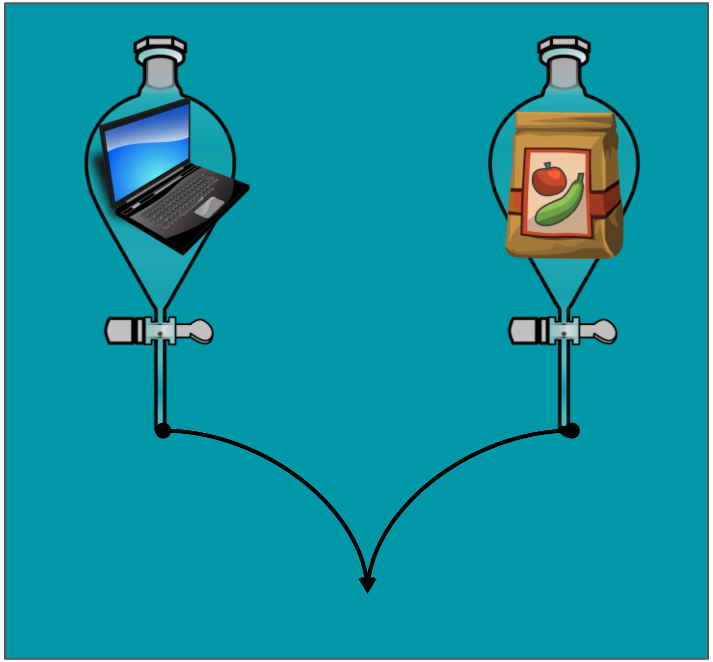
Task-Competitive Composition

Tasks ‘compete’ for individuals.

- Example: Advertisers compete for a single ad slot
- Goal: Individual Fairness for tech jobs advertising and groceries advertising *simultaneously*

tech firm vs grocery delivery service

Naïve Task-Competitive Composition



1 Grocery service decides whether to bid ($1)

2 Tech company bids among those not claimed by groceries ($0.50)

Not Guaranteed to be Fair

Theorem. For any two tasks T and T′ with nontrivial metrics D, D′, and for any tie-breaking function, not necessarily the same for each individual, there exist classifiers C and C′ that are individually fair in isolation, but when naïvely combined violate multiple task fairness.

A metric is trivial if all distances are in {0,1}

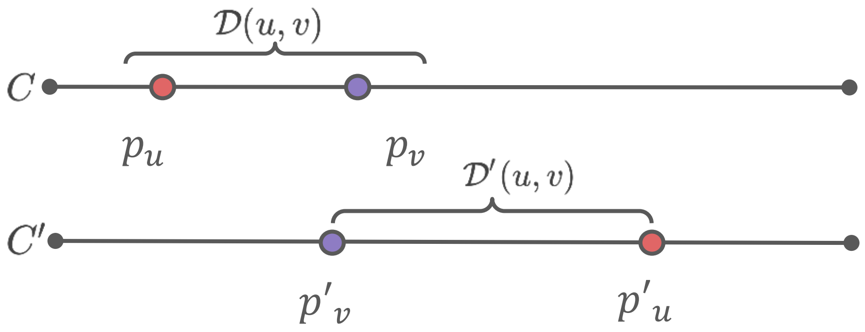
Not Guaranteed to be Fair

Theorem. For any two tasks T and T′ with nontrivial metrics D, D′, and for any tie-breaking function, not necessarily the same for each individual, there exist classifiers C and C′ that are individually fair in isolation, but when naïvely combined violate multiple task fairness.

A metric is trivial if all distances are in {0,1}

Proof Sketch (for case ties go to T)

- $0 < p_u < p_v$
- $p_u' \geq p_v' > 0$ and the distance is maximized subject to D′.



Proof Sketch (continued)

The difference in probability of positive classification for T′:

$(1-p_u)p_u' - (1-p_v)p_v' = D'(u,v) + p_v p'_v - p_u p'_u$

If $D'(u,v) = 0$ then done: ($p_u < p_v$; $p_u' = p_v' > 0$)

Write $\alpha = p_v/p_u$ so $p_v p'_v - p_u p'_u = \alpha p_u p'_v - p_u p'_u$

$\alpha p_u p'_v - p_u p'_u > 0 \Leftrightarrow$

$\alpha p'_v - p'_u > 0 \Leftrightarrow$

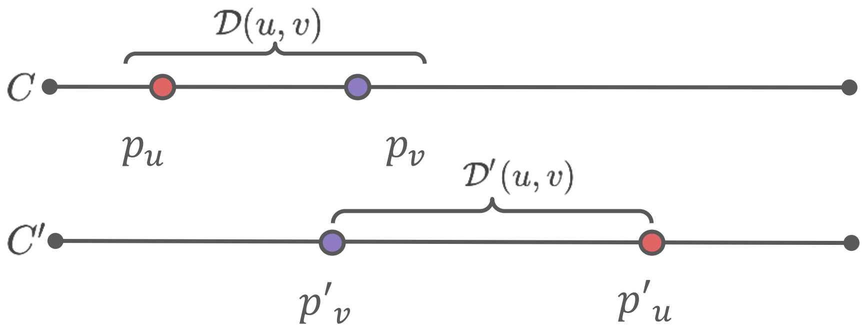
$\alpha = p_v / p_u > p'_u / p'_v$

Easy to ensure w/o violating fairness for T.

Omitted: a bit of cleanup for other elements.

Proof Sketch (for case ties go to T)

- $0 < p_u < p_v$
- $p_u' \geq p_v' > 0$ and the distance is maximized subject to D′.



An Algorithm: Randomize Then Classify

Procedure:

- Fix a probability distribution X over the tasks.
- Choose a task T~ *X*
- Classify using a fair classifier for T.

Homework: Prove RTC is individually fair.

Functional Composition

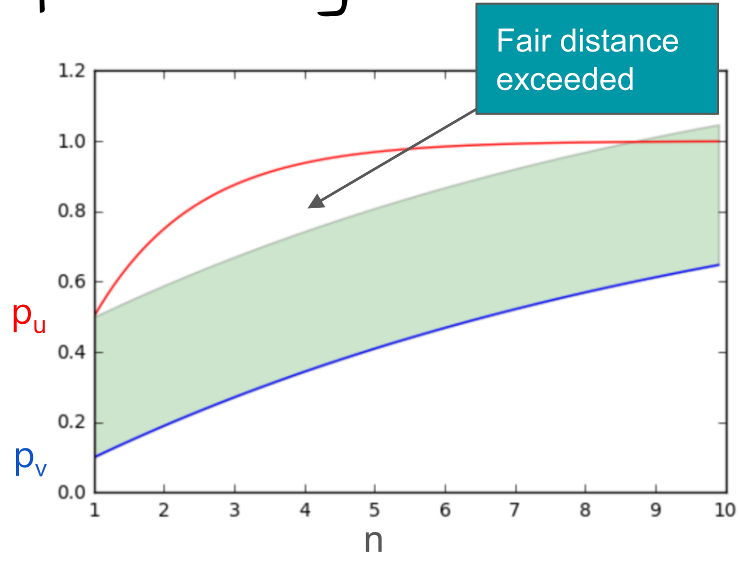
OR Fairness: Applying to multiple colleges

Relevant outcome: get in to at least one college.

Definition. A set of classifiers *C* for task T with metric *D* satisfy *OR-Fairness* if for all u, v in U $|p_u - p_v| \leq D(u,v)$, where p_w is the probability that w is accepted by at least one classifier in *C*.

Theorem. For any nontrivial task, there exists a set of classifiers that are fair in isolation, but violate OR-Fairness.

Proof Sketch: Characterize when $1-(1-p_u)^n$ grows faster than $1-(1-p_v)^n$.



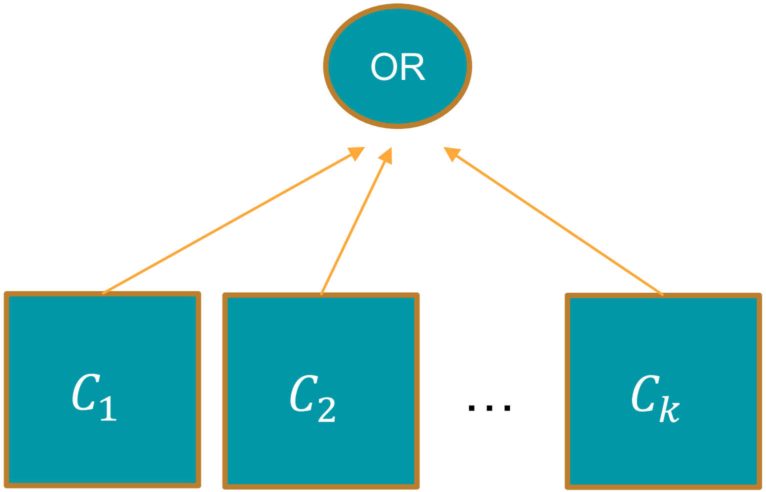
OR Fairness: Applying to multiple colleges

Relevant outcome: get in to at least one college.

Definition. A set of classifiers *C* for task T with metric *D* satisfy *OR-Fairness* if for all u, v in U $|p_u - p_v| \leq D(u,v)$, where p_w is the probability that w is accepted by at least one classifier in *C*.

Theorem. For any nontrivial task, there exists a set of classifiers that are fair in isolation, but violate OR-Fairness.

Observation. If for all C_i and all u in U the probability of positive classification for u under C_i is above ½, then fairness is preserved under OR-composition.



OR of "heavy" ORs

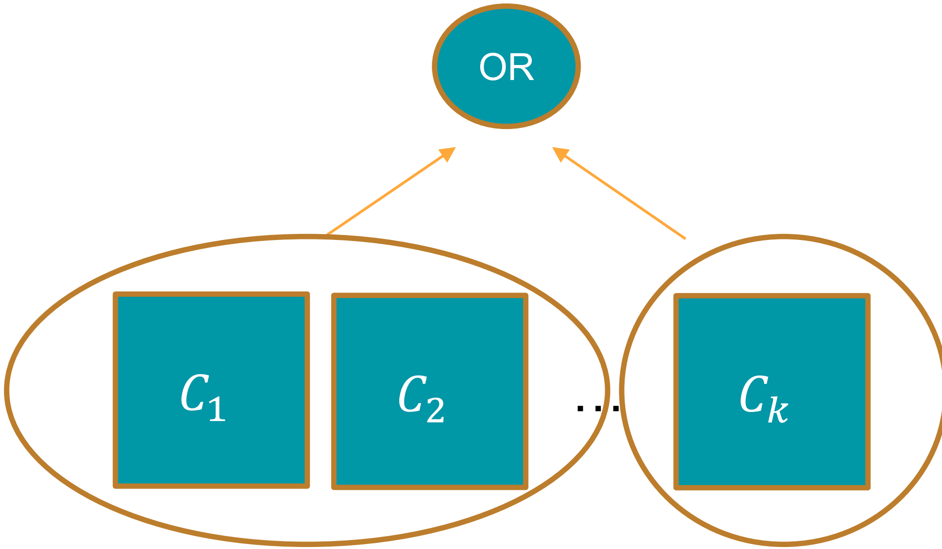
OR Fairness: Applying to multiple colleges

Relevant outcome: get in to at least one college.

Definition. A set of classifiers *C* for task T with metric *D* satisfy *OR-Fairness* if for all u, v in U $|p_u - p_v| \leq D(u,v)$, where p_w is the probability that w is accepted by at least one classifier in *C*.

Theorem. For any nontrivial task, there exists a set of classifiers that are fair in isolation, but violate OR-Fairness.

Theorem. Any set of individually fair classifiers for a task which have an aggregate probability of positive classification > ½ for all u ∈ U also satisfy OR-Fairness.



OR of “heavy” ORs

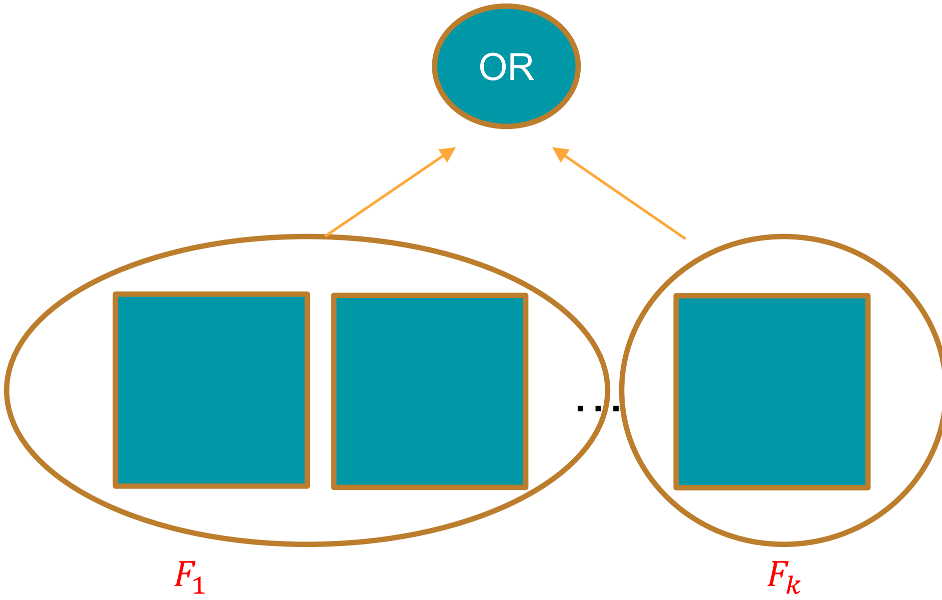
OR Fairness: Applying to multiple colleges

Relevant outcome: get in to at least one college.

Definition. A set of classifiers *C* for task T with metric *D* satisfy *OR-Fairness* if for all u, v in U $|p_u - p_v| \leq D(u,v)$, where p_w is the probability that w is accepted by at least one classifier in *C*.

Theorem. For any nontrivial task, there exists a set of classifiers that are fair in isolation, but violate OR-Fairness.

Theorem. Any set of individually fair classifiers for a task which have an aggregate probability of positive classification > ½ for all u ∈ U also satisfy OR-Fairness



OR of "heavy" circuits

Dependent Classifications

Dwork and Ilvento 2018

Dependent Compositions

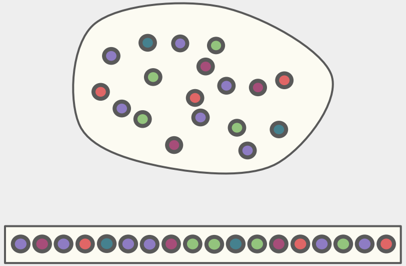
In many cases outcomes are not independent:

- Can only accept *n* students
- Must accept at least *n/2* students who can pay tuition
- Can't grant too many loans or hire too many people on a particular day

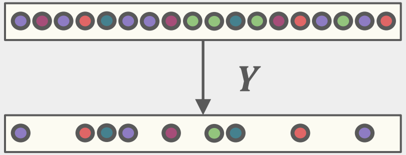
Two main settings:

- Cohort Selection (fixed size n)
- Universe Subset Problems: operate on subset, but still want fairness wrt all pairs

Dependent Composition: Many Possible Axes...

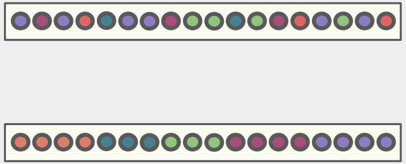


Offline
vs
Online

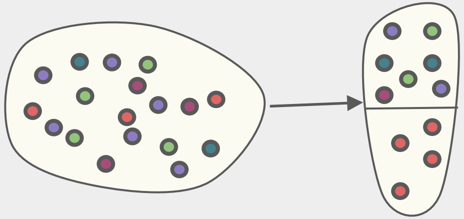


Universe
vs
Subset

Uniform Random
vs
Adversarial Ordering



Constrained
Outputs

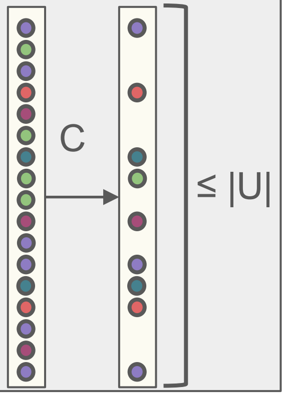


The Cohort Selection Problem

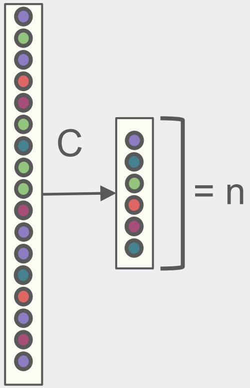
Cohort Selection: Given a universe of individuals U and an integer n < |U|, select n individuals from U such that for every u and v in U the difference in probability of selection p_u and p_v respectively satisfies $D(u,v) \geq |p_u - p_v|$.

Cannot independently classify each element, as the number of previously selected elements must be taken into account.

Independent Classification: By applying a fair classifier C independently to each element, can select up to |U| elements.



Cohort Selection: Must select n elements without increasing distances. Probability of selection dependent on other elements.

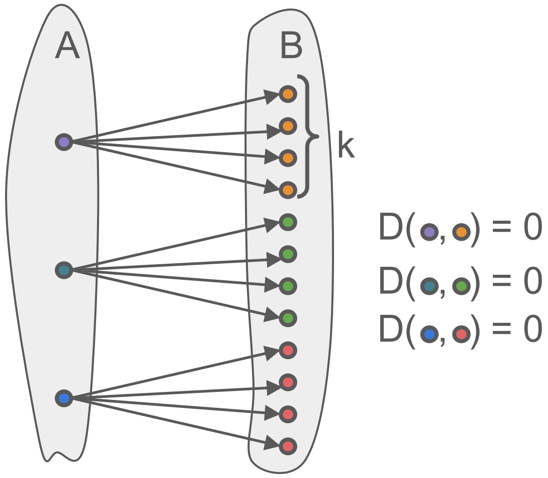


Constrained Cohort Selection

Definition. For $A \subseteq U$, $p \in [0,1]$, fairly select a set of n elements of U such that at least a p fraction of those selected are in A.

- Must have at least n/2 students who pay full tuition to cover operating costs
- Must satisfy statistical parity for legal reasons
- Must accept at least p*n students of each gender who can play a particular sport to field a team

An Impossibility Result (simple special case)



B is a blown-up version of A, and |A| = n/2=pn

$\forall u \in A$: $p_u = 1$

$\exists v \in B$: $p_v \leq 1/k$. Let $u \in A$ satisfy $D(u, v) = 0$.

QED.

The Universe Subset (Cohort Selection) Problem

Let $\boldsymbol{Y}$ be a distribution over subsets of U. Let $\boldsymbol{X}=\{\boldsymbol{X}(\mathrm{V})\}_{V\subseteq U}$ be family a distributions, where $\boldsymbol{X}$(V) is a distribution on permutations of the elements of V. For a system S_n: $\Pi(2^U)$ x r → U^*, `Experiment`(S_n,$\boldsymbol{X}$,$\boldsymbol{Y}$,u):

1. Choose r ~ {0,1}*
2. Choose V ~ $\boldsymbol{Y}$
3. Choose $\boldsymbol{\pi}$ ~ $\boldsymbol{X}$(V)
4. Run S_n, and output 1 if u is selected

The system is individually fair (and a solution to the Cohort Selection Problem) if ∀u, v ∈ U,

$$|\mathbb{E}[\texttt{Experiment}(S_n, \boldsymbol{X}, \boldsymbol{Y}, u)]-\mathbb{E}[\texttt{Experiment}(S_n, \boldsymbol{X}, \boldsymbol{Y}, v)]| \leq D(u,v)$$

(and S_n outputs a set of n distinct elements of U).

Solutions for Basic, Offline Cohort Selection

Easiest Setting: decisions made offline, with access to the entire universe U and the metric D with no constraints on the output set other than size.

Permute then Classify

1. Choose $\pi \sim S_{|U|}$, where $S_{|U|}$ is the set of all permutations of |U| elements
2. Apply π to U, and classify each element as usual until either:
   - n elements are selected: stop
   - there are exactly enough elements left in the permutation to select n total: take all remaining elements

Weighted Sampling

1. Enumerate all sets of size n, and for each set T assign weight $w(T) \propto \sum_{u \in T} E[C(u)]$
2. Sample from all of the sets with probability proportional to the weights

Individual Fairness in Pipelines

- Hire a cohort; one year later, promote a cohort member to team leader
    - Whether or not you are promoted depends on the cohort
    - Not so crazy: hiring decisions not necessarily made by team’s organizational head; hiring manager often different than manager one year later
- Gives rise to yet another catalog of evils

Two-Stage Cohort Pipeline

- Universe U, Permissible set of cohorts $C \subseteq \mathrm{Pow}(U) \setminus \emptyset$
- Cohort selection mechanism $A: U \to C$, aka the "hiring manager"
- A set of scoring functions $F: C \times U \to [0,1]$ and $f \in F$
  - Scoring is contextual, i.e., may have $f(c,u) \neq f(c',u)$
  - Undefined if $u \notin C$
- The pipeline is $A \circ f$

- C_u: set of cohorts in C containing individual $u \in U$
- Probability that A selects u is $p(u) = \sum_{c \in C_u} \Pr[A(U) = c]$
  - *Assume* intra-cohort fairness $\forall c \in C \; \forall u, v \in c \; |f(c,u) - f(c,v)| \leq d(u,v)$

Informal, Deceptively Simple, Fairness Definition

For $f \in F$, the pipeline instantiated with f is individually fair with respect to similarity metric d and distribution metric $D\colon \Delta(\mathrm{O}_{\text{pipeline}}) \times \Delta(\mathrm{O}_{\text{pipeline}}) \to [0,1]$ if $\forall\, u, v \in U\colon D\big([f \circ A](u), [f \circ A](v)\big) \le d(u,v)$.

If the pipeline is individually fair wrt d, D for all $f \in F$, then it is robust to F.

The hitch: outcome space $\mathrm{O}_{\text{pipeline}} = [0,1] \cup \{\perp\}$: *individuals drop out*, voluntarily or otherwise, from the pipeline.

Unconditional and Conditional Distributions

Probabiltiy that A outputs $c \in C$

Starting point:

$\xi_u \in \Delta(\mathrm{O}_{\mathrm{pipeline}})$ places probability $1 - p(u)$ on $\bot$

$\xi_u \in \Delta(\mathrm{O}_{\mathrm{pipeline}})$ places probability $\sum_{c \in C_u} \Pr[f(c,u) = s] P_A(c)$ on $s \in [0,1]$

Unconditional distribution: as above, but treat $\bot$ as having score of 0

For $s \in (0,1]$: place probability $\sum_{c \in C_u} \Pr[f(c,u) = s]\, P_A(c)$ on s

For $s = 0$: place probability $1 - p(u) + \sum_{c \in C_u} \Pr[f(c,u) = 0]\, P_A(c)$

Conditional distribution: condition on positive $p(u)$:

For $s \in [0,1]$ place probability $\frac{\sum_{c \in C_u} \Pr[f(c,u)=s] P_A(c)}{p(u)}$ on s

Unconditional distribution: treat $\perp$ as having score of 0

Is This Fair?

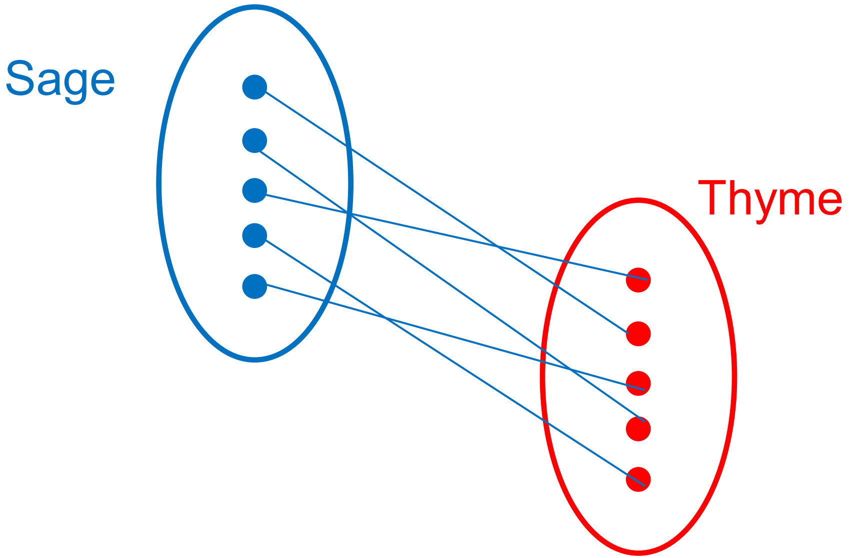
- $d(u,v) = 0.1$
- Under A, $p(u) = p(v) \stackrel{\text{def}}{=} p^*$ but A never outputs a cohort containing both
- Constrain f for the unconditional distribution $|p(u)f(u) - p(v)f(v)| \leq d(u,v)$, Simplifies to $p^*|f(u) - f(v)| \leq d(u,v)$
- Weak fairness constraint when p^* is small!

Congratulations, you are offered a job! After a year you, may expect a promotion with probability $f(u)$ (or, for v, $f(v)$).

Ulfar and Virginia receive offers with the same probability, but correctly perceive the offers very differently.

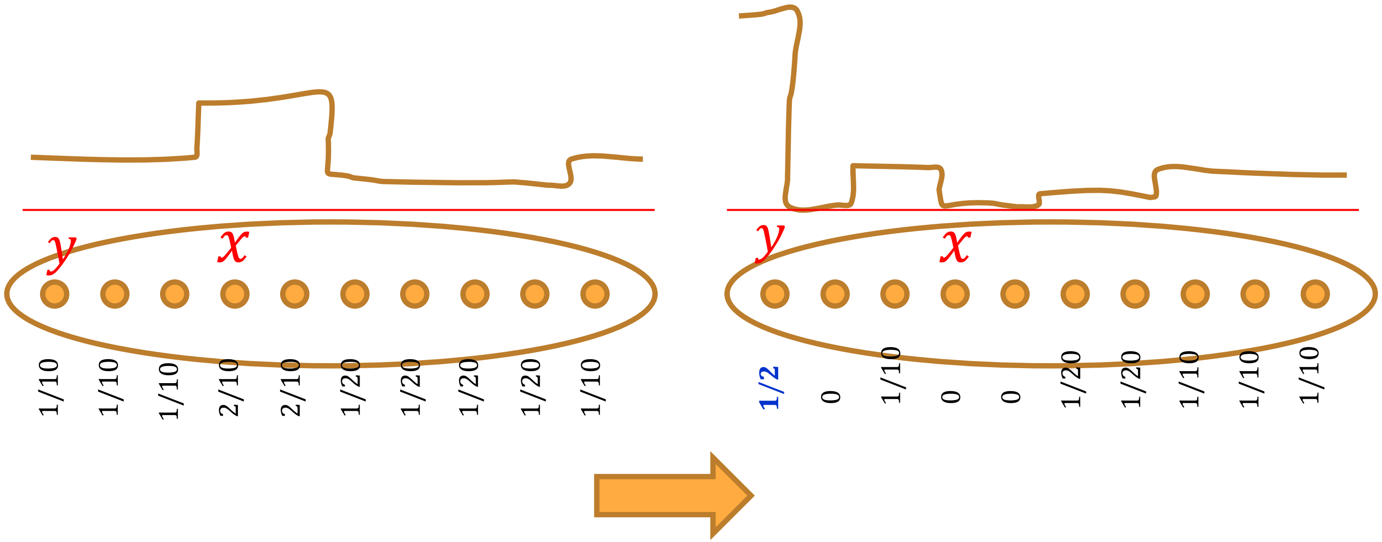
Metric-Fair Affirmative Action

Fair AA via Metrics (highly simlplified)



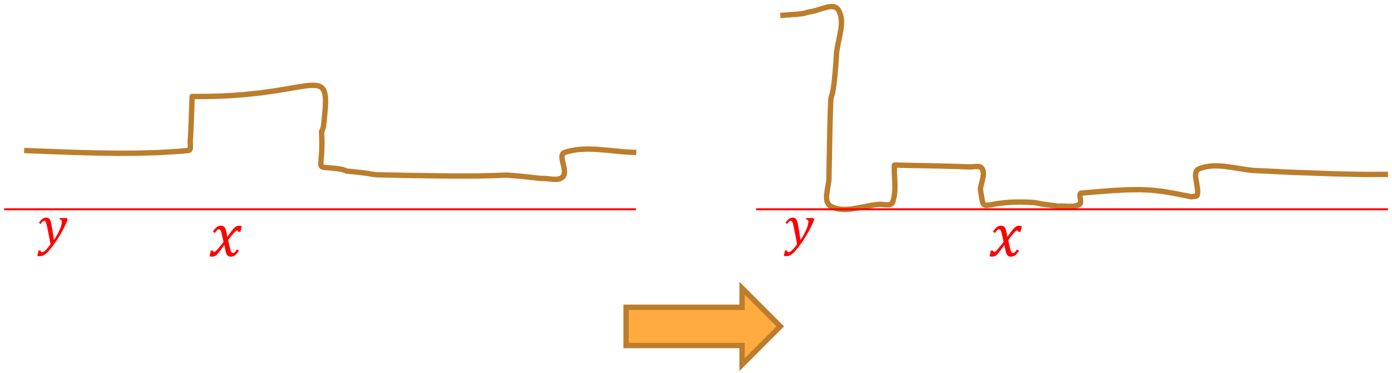
- Pair up S's and T's to minimize $\sum_i d(s_i, t_{\text{pair}(s_i)})$
- Classify s_i by classifying $t_{\text{pair}(s_i)}$

Transforming One Distribution to Another



Transforming One Distribution to Another

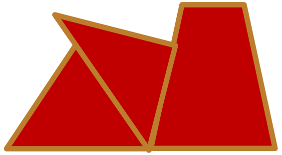
"Cost" (in clay-moving) captures difference between distributions

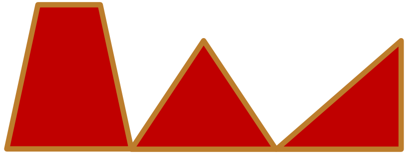


$d(y, y) = 0$

$d(x, y)$ depends on the metric

The Earthmover Linear Program





$$EM(S,T) = \min \Sigma_{x,y \in V}\, h(x,y)\, d(x,y)$$

Amount to "haul" from x to y

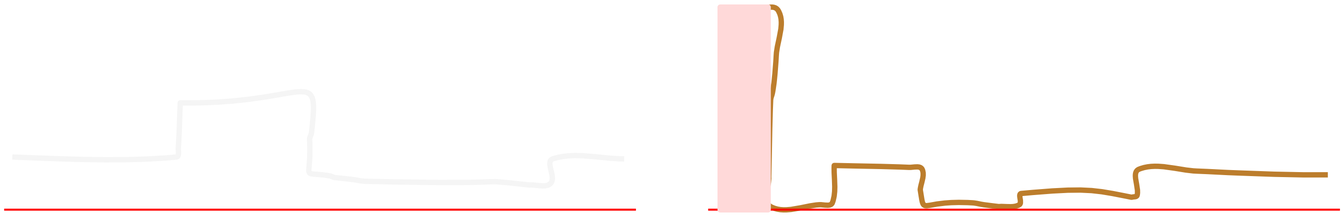
s.t.

$$\Sigma_{x \in V} h(x,y) = T(y)$$

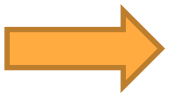
$$\Sigma_{y \in V} h(x,y) = S(x)$$

$$h(x,y) \geq 0$$

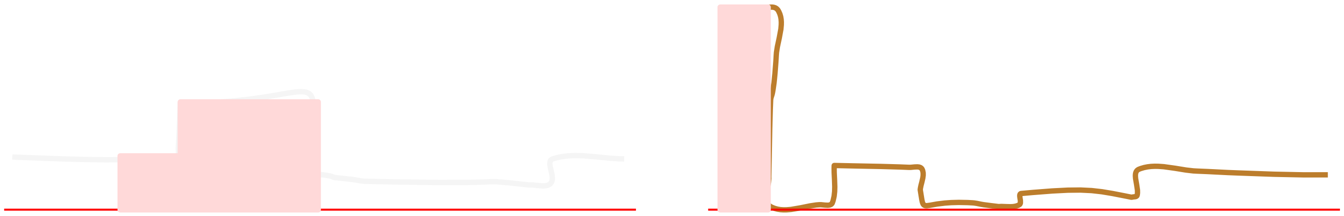
Transforming One Distribution to Another



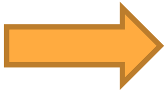
Suppose 1/2 the clay on the right is pink…



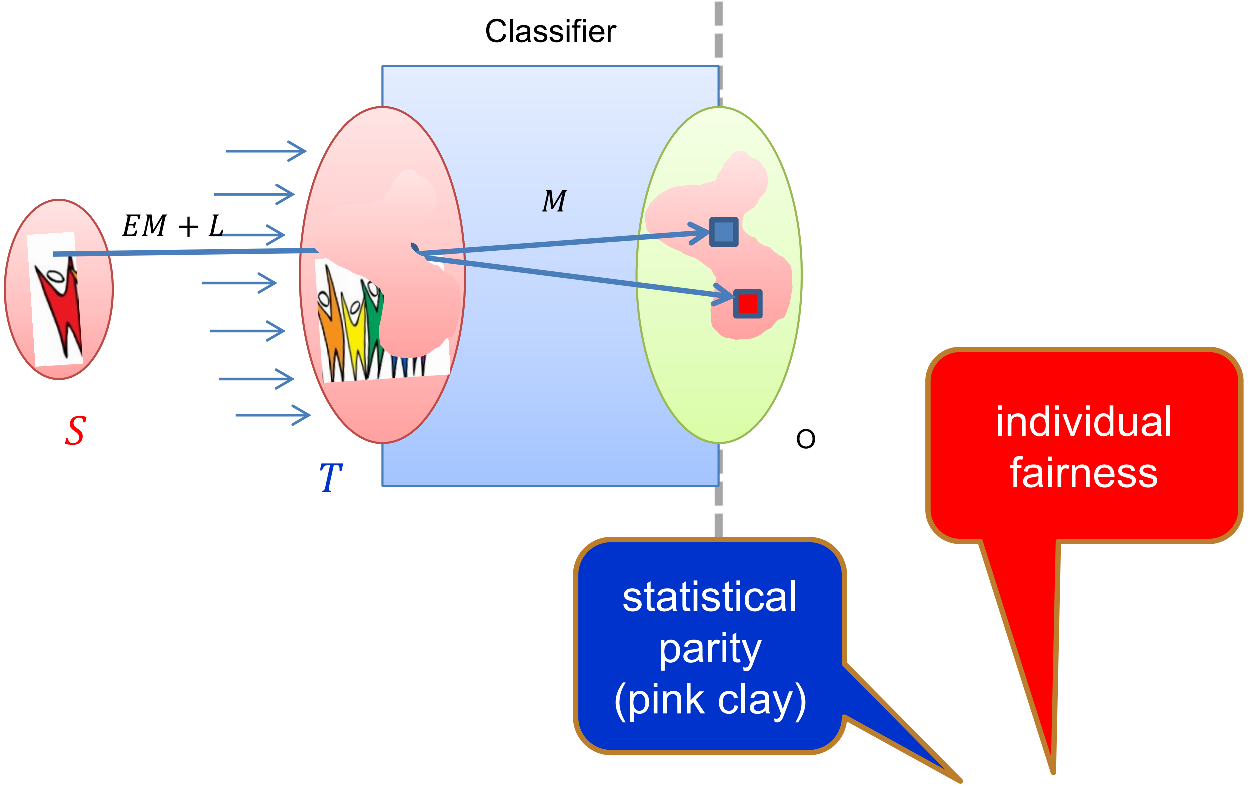
Transforming One Distribution to Another



Then 1/2 the clay on the left is pink!

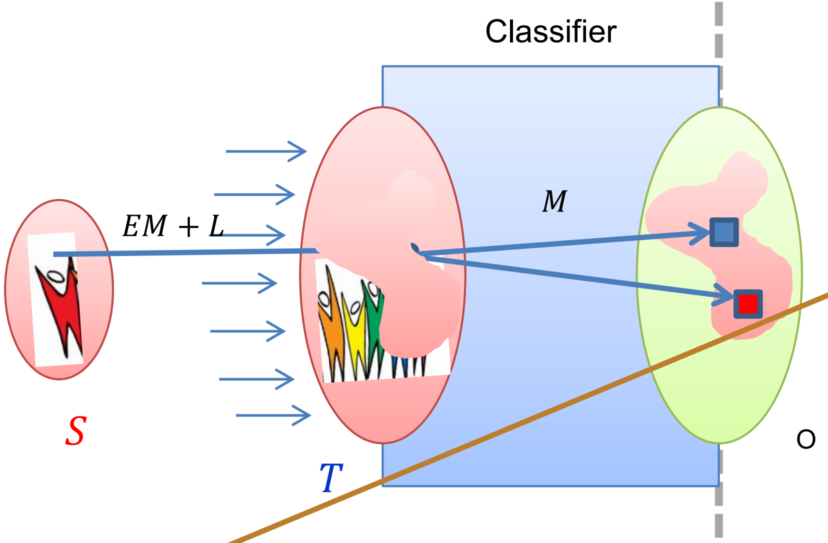


Fair Affirmative Action via Metrics



Map uniform distribution on S to uniform distribution on T (via EM, L)

Fair Affirmative Action via Metrics



"Reweighting" of loss function L, taking into account the loss on S through its mapping to T.

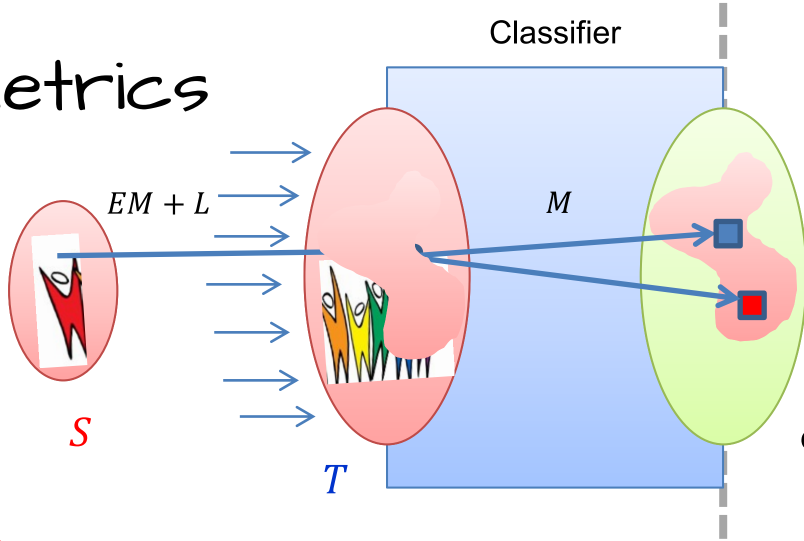
1. $\forall y \in T, o \in O : L'(y,o) = \sum_{x \in S} \mu_x(y) L(x,o) + L(y,o)$ where μ_x is from the EM+L mapping
2. Run the Fairness LP only on T, using L'

Fair Affirmative Action via Metrics

$$d_{\{EM+L\}}(S,T) \stackrel{\text{def}}{=} \min E_{x\in S} E_{y\sim\mu_x}\, d(x,y)$$

s.t.

$$D(\mu_x, \mu_{x'}) \le d(x,x') \quad \forall x, x' \in S$$
$$D_{TV}(\mu_S, U_T) \le \varepsilon$$
$$\mu_x \in \Delta(T)\ \forall x \in S$$



$$\forall y \in T\colon \mu_S(y) = E_{x\sim S}\, \mu_x(y)$$

Given $\{\mu_x\}_{x\in S}$ here and $\{\nu_x\}_{x\in T}$ from Fairnes LP,
define: $M\colon V \to \Delta(O)$:

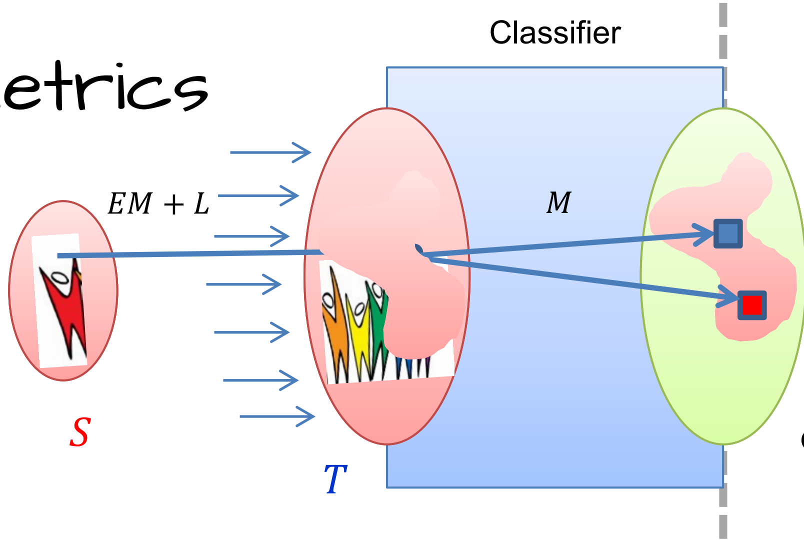
$$M(x) = \begin{cases} \nu_x & x \in T \\ E_{y\sim\mu_x}\nu_y & x \in S \end{cases}$$

Fair Affirmative Action via Metrics

$$d_{\{EM+L\}}(S,T) \stackrel{\text{def}}{=} \min E_{x\in S} E_{y\sim\mu_x}\, d(x,y)$$

s.t.

$$D(\mu_x, \mu_{x'}) \le d(x,x') \quad \forall x, x' \in S$$
$$D_{TV}(\mu_S, U_T) \le \varepsilon$$
$$\mu_x \in \Delta(T)\ \forall x \in S$$



Given $\{\mu_x\}_{x\in S}$ here and $\{\nu_x\}_{x\in T}$ from Fairnes LP, define: $M: V \to \Delta(O)$:

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

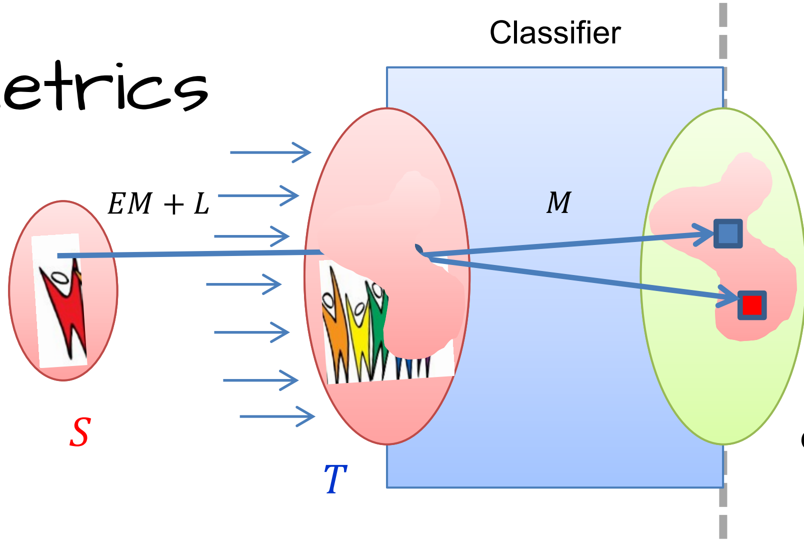
Minimizes loss AND disruption of $S \times T$ Lipschitz requirement, subject to parity and the within-group Lipschitz constraints

Fair Affirmative Action via Metrics

$$d_{\{EM+L\}}(S,T) \stackrel{\text{def}}{=} \min E_{x\in S} E_{y\sim\mu_x}\, d(x,y)$$

s.t.

$$D(\mu_x, \mu_{x'}) \le d(x,x') \quad \forall x, x' \in S$$
$$D_{TV}(\mu_S, U_T) \le \varepsilon$$
$$\mu_x \in \Delta(T)\ \forall x \in S$$



Given $\{\mu_x\}_{x\in S}$ here and $\{\nu_x\}_{x\in T}$ from Fairnes LP, define: $M: V \to \Delta(O)$:

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

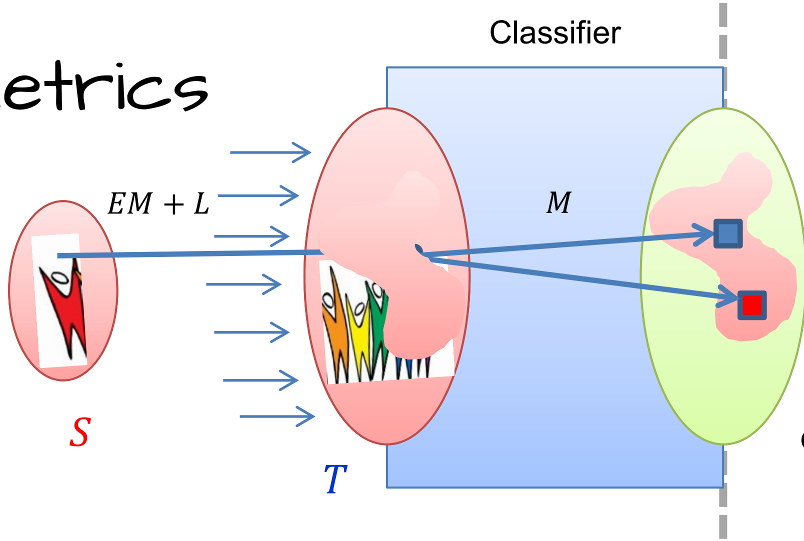
More flexibility still: can eliminate the re-weighting, prohibiting expression of opinions on the fate of elements in S. May make sense if vendor has done no market research on S

Fair Affirmative Action via Metrics

$$d_{\{EM+L\}}(S,T) \stackrel{\text{def}}{=} \min E_{x\in S} E_{y\sim\mu_x}\, d(x,y)$$

s.t.

$$D(\mu_x, \mu_{x'}) \le d(x,x') \quad \forall x, x' \in S$$
$$D_{TV}(\mu_S, U_T) \le \varepsilon$$
$$\mu_x \in \Delta(T)\ \forall x \in S$$



Given $\{\mu_x\}_{x\in S}$ here and $\{\nu_x\}_{x\in T}$ from Fairnes LP, define: $M: V \rightarrow \Delta(O)$:

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

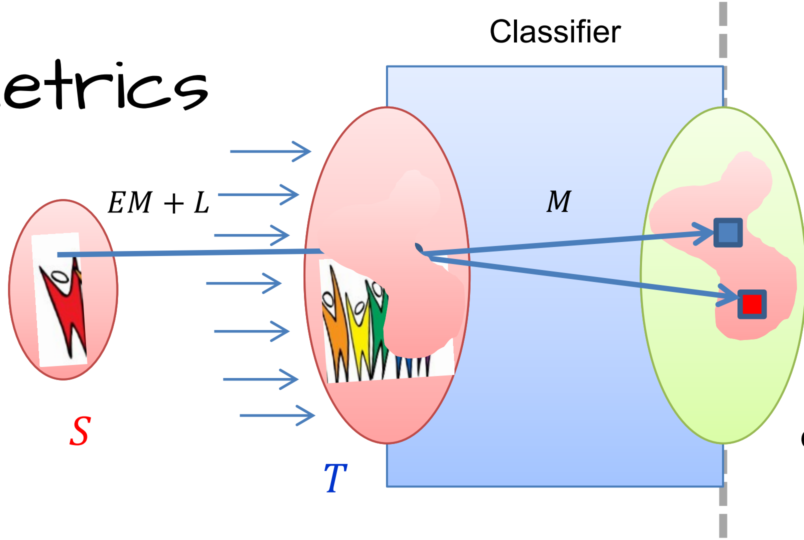
Compare to just adding statistical parity the Fairness LP, and eliminating the cross-group Lipschitz constraints: the approach here is more faithful to the $S \times T$ distances, providing protection against the "self-fulfilling prophecy" evil in which one deliberately selects the "wrong" subset of S

Fair Affirmative Action via Metrics

$$d_{\{EM+L\}}(S,T) \stackrel{\text{def}}{=} \min E_{x\in S} E_{y\sim\mu_x}\, d(x,y)$$

s.t.

$$D(\mu_x, \mu_{x'}) \le d(x,x') \quad \forall x, x' \in S$$
$$D_{TV}(\mu_S, U_T) \le \varepsilon$$
$$\mu_x \in \Delta(T)\ \forall x \in S$$



Given $\{\mu_x\}_{x\in S}$ here and $\{\nu_x\}_{x\in T}$ from Fairnes LP, define: $M: V \rightarrow \Delta(O)$:

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

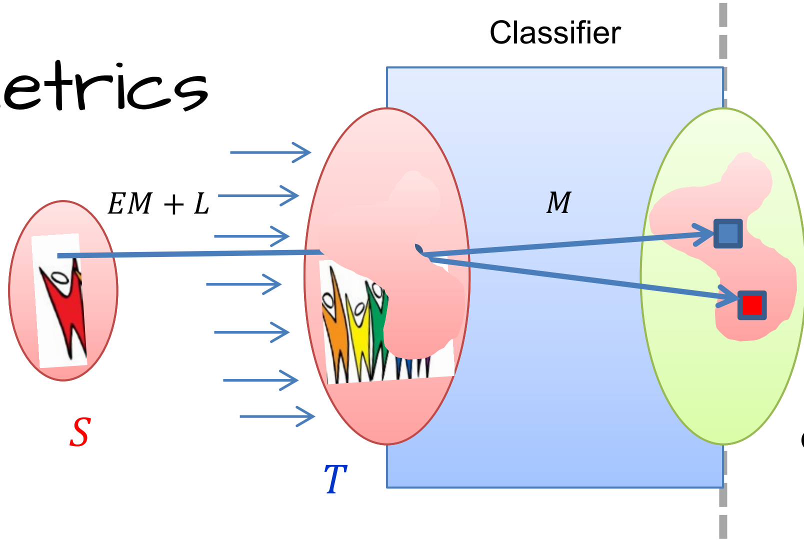
The metric is everything.
In this view, one can adjust the metric in such a way that the Lipschitz condition will imply statistical parity; makes sense if one believes that the metric does not fully reflect potential that may be undeveloped because of unequal access to resources. Reflected in the *ranking* approach discussed below.

Fair Affirmative Action via Metrics

$$d_{\{EM+L\}}(S,T) \overset{\text{def}}{=} \min E_{x\in S} E_{y\sim\mu_x}\, d(x,y)$$

s.t.

$$D(\mu_x,\mu_{x'}) \le d(x,x') \quad \forall x,x' \in S$$
$$D_{TV}(\mu_S, U_T) \le \varepsilon$$
$$\mu_x \in \Delta(T)\ \forall x \in S$$



Given $\{\mu_x\}_{x\in S}$ here and $\{\nu_x\}_{x\in T}$ from Fairnes LP, define: $M: V \to \Delta(O)$:

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

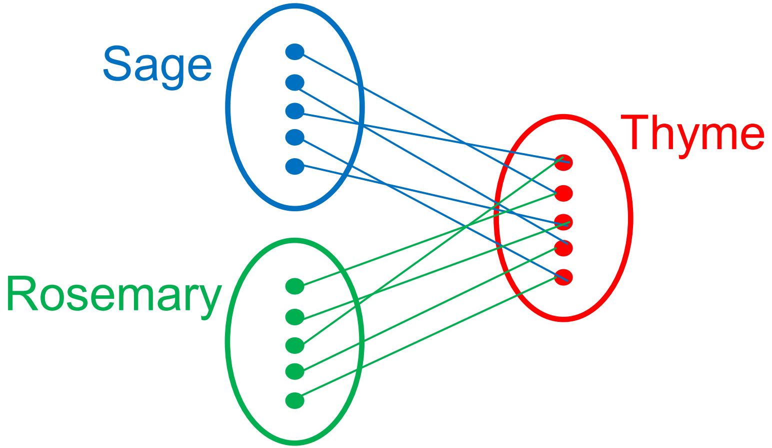
Claim: $M(x)$ satisfies

(1) statistical parity between S and T up to bias ε; and
(2) the Lipschitz condition for every within-group pair.

$$D_{TV}\big(M(S), M(T)\big) = D_{TV}\big(E_{x\in S}E_{y\sim\mu_x}\nu_y, E_{x\in T}\nu_x\big)$$
$$\le D_{TV}(\mu_S, U_T) \le \varepsilon$$

Fair Affirmative Action via Metrics

- We know how to handle multiple disjoint groups / strata / ZIP+4s
  - With a metric
- The intersectional case?



Fair Affirmative Action via Rankings

- Example: Universities of Texas and California
  - Top 10% of students in each high school class
- Example [John Roemer]:
  - Stratify students according to education level of mother
  - Rank students within each stratum by number of hours spent on homework per week
  - Admit to university top k% from each stratum
- Example [Danielle Allen, “Talent is Everywhere”]
  - Stratify students according to SAT/GPA and discard all below a fixed threshold
  - Admit randomly so as to maximize geographic diversity

Metric-Fair Affirmative Action

- We know how to handle multiple disjoint groups / strata / ZIP+4s
  - With a metric
  - Without a metric, from a "fair ranking"
- The intersectional case?
- *Can address intersectionality via Evidence-Based Ranking*
  - *Hold that thought*