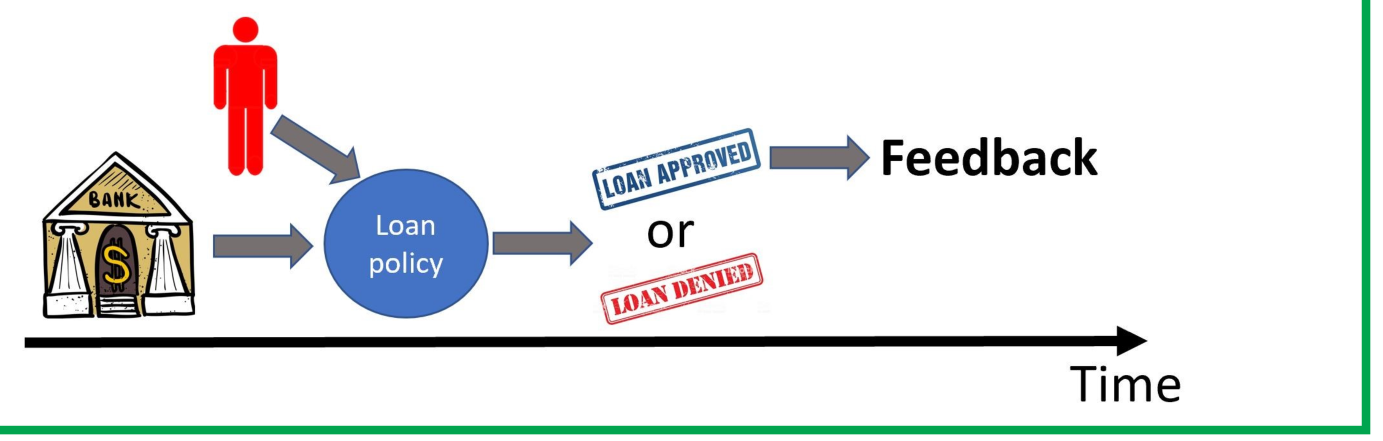
Equal Opportunity in Online Classification with Partial Feedback

{Yahav Bechavod, Katrina Ligett}, Hebrew University of Jerusalem   Aaron Roth, University of Pennsylvania
Bo Waggoner, University of Colorado   Zhiwei Steven Wu, University of Minnesota

Motivation

- Loan approvals
- Hiring decisions
- Online advertising
- Predictive policing
- ...

One-Sided Feedback!



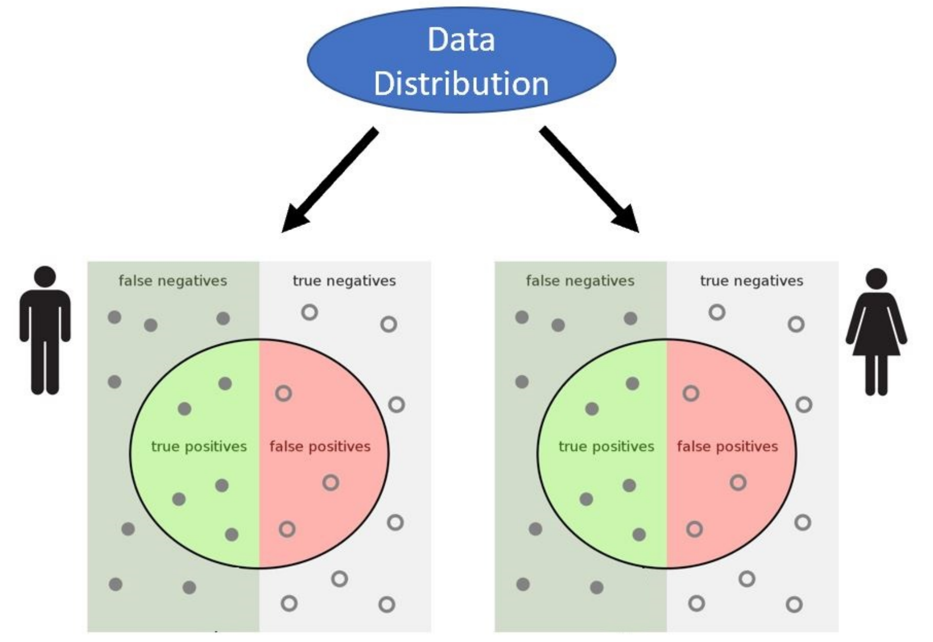
Model

- (Unknown) Distribution $\mathcal{D}$ over $\mathcal{X} \times \{\pm 1\}$
- Hypothesis class $\mathcal{H} : \mathcal{X} \to \{\pm 1\}$

Learner-Environment Interaction

for $t = 1, ..., T$ **do**
  Learner deploys a policy $\pi_t \in \Delta(\mathcal{H})$
  Environment draws $(x_t, y_t) \sim \mathcal{D}$ independently; learner observes x_t
  Learner labels the point $\hat{y}_t = h_t(x_t)$, where $h_t \sim \pi_t$
  if $\hat{y}_t = +1$ **then**
    Learner observes y_t

Fairness



Fairness constraint: Algorithm must, **on every round**, deploy a **fair policy**.
Fair policy: Similar false positive rates (or false negative rates) on both sub-populations: $\Delta_{FPR}(\pi) := FPR_{Men}(\pi) - FPR_{Women}(\pi) = 0$

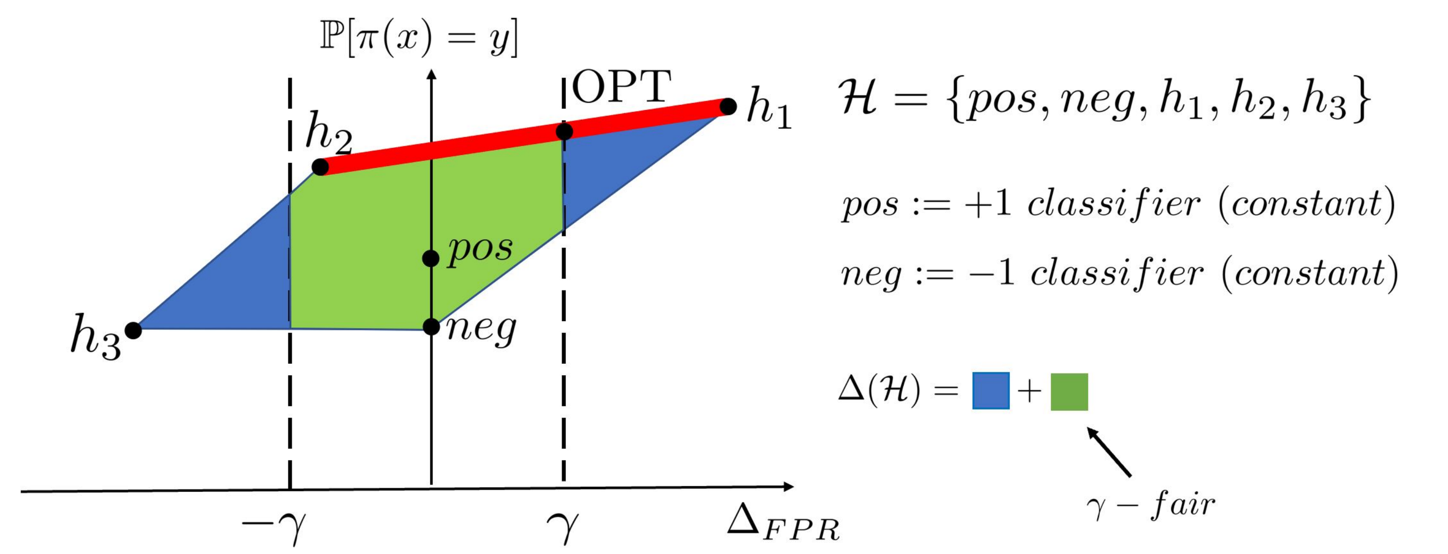
Fairness-Accuracy Trade-Off

Definition (γ-fair policy) Fix a distribution $\mathcal{D}$. A policy $\pi \in \Delta(\mathcal{H})$ satisfies the γ-*equalized false positive rate* constraint if $|\Delta_{FPR}(\pi)| \leq \gamma$.

Question: Why use a policy instead of a single hypothesis?

- Policies achieve better accuracy-fairness trade-off than single hypotheses.
- Optimal trade-off is always attained by policy of support size 2 (at most).

Example:



Objective

$$\min_A \quad Regret(A) \text{ w.r.t. } \gamma - fair \text{ policies in } \Delta(\mathcal{H})$$
$$\text{s.t.} \quad A \text{ is } \gamma' - fair \text{ online learning algorithm}$$

Question: What is the optimal trade-off between algorithm's regret and the "fairness gap" $\gamma' - \gamma \geq 0$?

Partial Feedback->Contextual Bandits

Remember: No feedback for negative predictions!

Question: How can learner minimize regret, when he cannot even measure his own regret?

Solution: Regret-preserving manipulation of the loss matrix:

$$L = \begin{array}{c|cc} & \text{Repays} & \text{Defaults} \\ \hline \text{Approve} & 0 & 1 \\ \text{Deny} & 1 & 0 \end{array} \qquad \tilde{L} = \begin{array}{c|cc} & \text{Repays} & \text{Defaults} \\ \hline \text{Approve} & 0 & 2 \\ \text{Deny} & 1 & 1 \end{array}$$

Regret-preserving: $\forall t \in [T]: \quad S_t = \{(x_i, y_i)\}_{i=1}^t$
$\forall h : \tilde{L}(h, S_t) = L(h, S_t) + \sum_{i=1}^t \mathbb{1}_{[y_i = defaults]}$

Difference between the losses of any two hypotheses remains **the same** after the transformation.

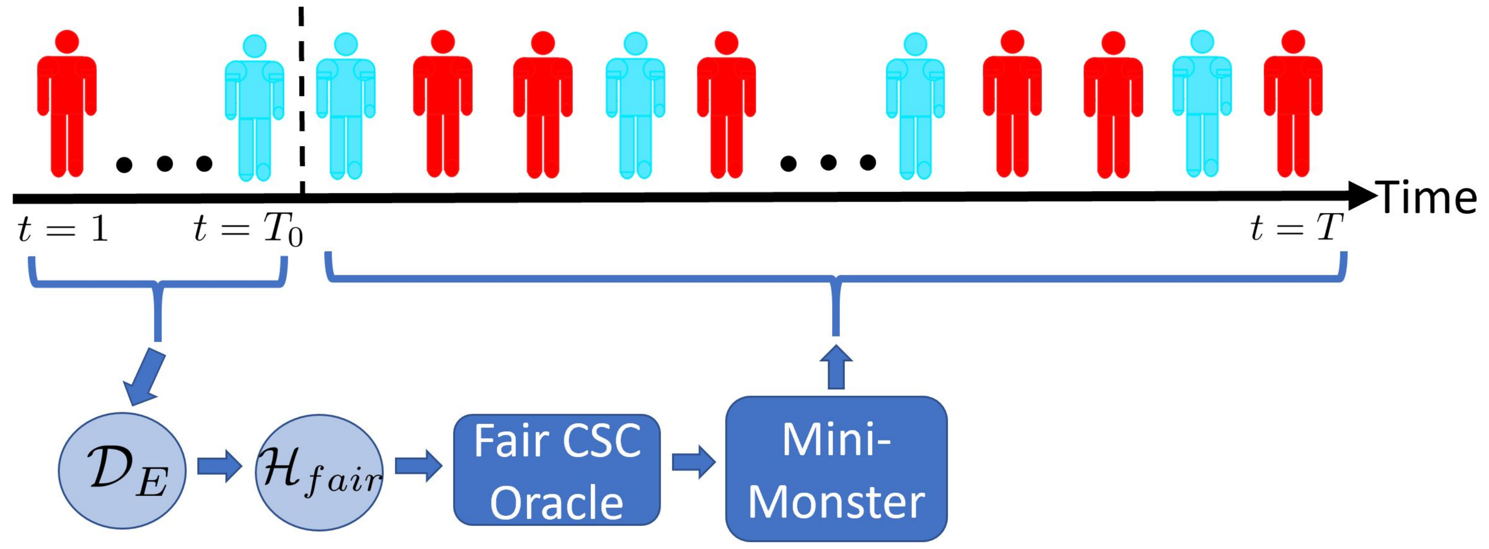
Main Result

Theorem There exists an **oracle-efficient** algorithm that takes parameters $\delta \in [0, \frac{1}{\sqrt{T}}]$ and $\gamma \geq 0$ as input and satisfies, w.p. $1 - \delta$, γ'-fairness and has an expected regret at most $\tilde{O}(\sqrt{T}\ln(|\mathcal{H}|/\delta))$ with respect to the class of γ-fair policies, where $\gamma' = \gamma + O(\sqrt{\ln(|\mathcal{H}|/\delta)}/T^{1/4})$.

Algorithm

Basic outline:

1. For the first T_0 rounds, perform *pure exploration* by always predicting $+1$ to collect labelled data.
2. Use collected data to form empirical fairness constraints, construct a fair Cost Sensitive Classification oracle based on empirical constraints.
3. Run an (oracle-efficient) adaptive contextual bandit algorithm - "Mini-Monster" by Agarwal et al. 2014 - that minimizes cumulative regret, while satisfying the empirical fairness constraint on every round.



Naive approaches: Explore-then-exploit (sub-optimal), Exploration + standard bandit algorithm (inefficient).

Optimization Oracle

1. We assume access to a Cost-Sensitive Classification oracle.
2. We adapt the reduction by Agarwal et al. 2018 to handle optimization with constraints defined only on the empirical distribution formed by the exploration data.
3. The result is an oracle that solves Cost-Sensitive Classification problems with empirical fairness constraints.

Regret Analysis

Main challenge: Unlike Agarwal et al. 2014, have to handle an **Infinite policy class.**

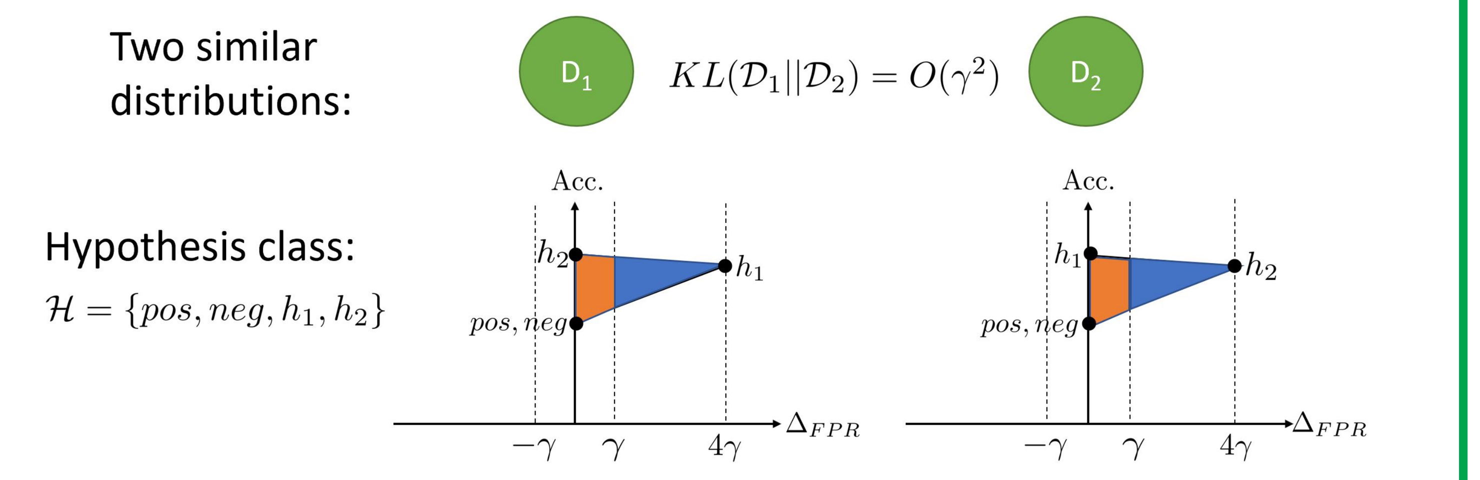
Useful fact: The set of optimal fair policies is **sparse**.

Lower Bound

Theorem Fix any $\alpha \in (0, 0.5)$ and let $T \geq \sqrt[\alpha]{16}$. Fix any $\delta \leq 0.24$. There exists a hypothesis class $\mathcal{H}$ containing the constant classifiers $\{\pm 1\}$ such that any algorithm satisfying a $T^{-\alpha}$-fairness constraint w.p. $1-\delta$ has expected regret with respect to the set of 0-fair policies of $\Omega(T^{2\alpha})$.

Intuition:

1. Define instance consisting of two very similar distributions, $\mathcal{D}_1$ and $\mathcal{D}_2$ defined as a function of our algorithm's fairness target γ.
2. Roughly, there are not enough samples to distinguish the distributions until at least $\Theta(\frac{1}{\gamma^2})$ rounds elapse
3. In order to equalize false positive rates on both distributions, an algorithm must "play it safe" and incur linear regret per round during this time.



Conclusion: The trade-off our algorithm exhibits between its regret bound and the "fairness gap" $\gamma' - \gamma$ is **optimal.**

Contact Information

Yahav Bechavod, yahav.bechavod@cs.huji.ac.il
Katrina Ligett, katrina@cs.huji.ac.il
Aaron Roth, aaroth@cis.upenn.edu
Bo Waggoner, bwag@colorado.edu
Zhiwei Steven Wu, zsw@umn.edu

Acknowledgements

YB and KL were funded in part by Israel Science Foundation (ISF) grant 1044/16, the United States Air Force and DARPA under contract FA8750-16-C-0022, and the Federmann Cyber Security Center in conjunction with the Israel national cyber directorate. AR was funded in part by NSF grant CCF-1763307 and the United States Air Force and DARPA under contract FA8750-16-C-0022. ZSW was supported in part by a Google Faculty Research Award, a J.P. Morgan Faculty Award, a Mozilla research grant, and a Facebook Research Award. Part of this work was done while KL and ZSW were visiting the Simons Institute for the Theory of Computing, and BW was a postdoc at the University of Pennsylvania's Warren Center and at Microsoft Research, New York City. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of JP Morgan, the United States Air Force and DARPA.