Learning Fair Scoring Functions

Robin Vogel[1] Aurélien Bellet[3] Stephan Clémençon[2]

[1] University of Edinburgh, [2] Télécom Paris, [3] Inria

27/04/2021

Outline

Interest in fair algorithms

Vox recode

Why algorithms can be racist and sexist

A computer can make a decision faster. That doesn't make it fair.

By Rebecca Heilweil | Feb 18, 2020, 12:20pm EST

NBER | NATIONAL BUREAU of ECONOMIC RESEARCH

‹ Working Papers

Consumer-Lending Discrimination in the FinTech Era

Sections | The Washington Post — Democracy Dies in Darkness | Get one year for €20 | Sign in

Health

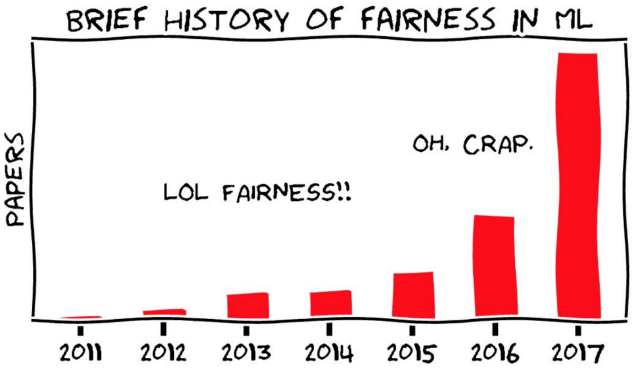
Racial bias in a medical algorithm favors white patients over sicker black patients

WIRED BACKCHANNEL BUSINESS CULTURE GEAR IDEAS MORE SIGN IN SUBSCRIBE

TOM SIMONITE BUSINESS 07.22.2019 07:00 AM

The Best Algorithms Struggle to Recognize Black Faces Equally

US government tests find even top-performing facial recognition systems misidentify blacks at rates five to 10 times higher than they do whites.



Fair algorithmic decisions

Algorithmic decisions are increasingly used in many domains:

Banking (*e.g.* loans) Recruting (*e.g.*, hiring)

Insurance (*e.g.* cars) Judiciary (*e.g.*, bail)

Recently, the fairness of algorithms has gathered lots of attention.

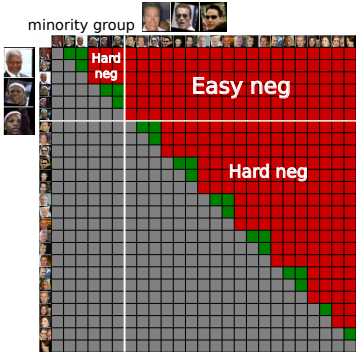
e.g. May 2016: The COMPAS system assesses the likelihood of recidivism of a defendant for U.S. courts.

While algorithms are usually designed for the interest of some user, fair algorithms suggests confronting those to the law.

"Predictive models are really just opinions embedded in math."

Cathy O'Neil. (Weapons of Math Destruction, 2016)

Case Study: Fair Facial Recognition (FR)



(LFW, Huang et al.,2007)

In FR, we compare pairs of images with a similarity $s(x, x')$.

The similarity is **thresholded** $s(x, x') > t$ to distinguish **similar** from **dissimilar** examples.
$\rightarrow$ obtain true and false positive rates.

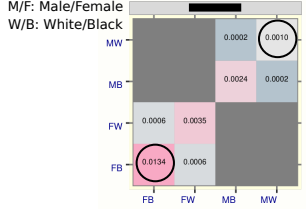
FR algorithms are trained with celebrities (*e.g.* LFW, CelebA), with **uneven distribution over ethnicities/gender**.

In 1:1 verification (*e.g.* border control), the NIST has shown differences in FPR.

Ongoing Face Recognition Vendor Test (FRVT)
Part 1: Verification

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

FPR for t s.t. $FPR_{MW} = 10^{-3}$



(NIST FRVT 1:1, Grother and Ngan, 2019)

Outline

Fairness in binary classification (1/2)

Probabilistic framework:

· *Binary classification:* $(X, Y) \sim P$ and $(X, Y) \in \mathcal{X} \times \{-1, 1\}$,
learn a classifier $g : \mathcal{X} \to \{-1, 1\}$ from data $\{(X_i, Y_i)\}_{i=1}^{n} \overset{i.i.d.}{\sim} P$.

· *Fairness:* Sensitive information $Z \in \{0, 1\}$, a Z_i for each (X_i, Y_i).
e.g. gender, ethnicity, …

Fairness? → no universal definition.
Example of **discrimination**:

> [. . .] ***wrongfully*** *impose a* ***relative disadvantage*** *on persons based on their membership in some* ***salient social group*** *e.g. race or gender.*
>
> *Altman et al. (2016)*

Different real-world scenarios → diff notions of fairness → diff measurements.
[Zafar et al., 2019]

Fairness in binary classification (2/2)

Fairness without ground truth: Parity in …

· Treatment: $g(X, Z) = g(X)$ almost surely.
i.e. the decision does not depend on the sensitive attribute.

· Impact: $\mathbb{P}\{g(X) = +1|Z = 0\} = \mathbb{P}\{g(X) = +1|Z = 1\}$.

Fairness with ground truth: Parity in …

· Error: $\mathbb{P}\{g(X) \neq Y \mid Z = 0\} = \mathbb{P}\{g(X) \neq Y \mid Z = 1\}$,

· TPR: $\mathbb{P}\{g(X) = 1 \mid Z = 0, Y = +1\} = \mathbb{P}\{g(X) = 1 \mid Z = 1, Y = +1\}$,

· FPR: $\mathbb{P}\{g(X) = 1 \mid Z = 0, Y = -1\} = \mathbb{P}\{g(X) = 1 \mid Z = 1, Y = -1\}$,

[Zafar et al., 2019]

Related work

Fairness in binary classification has gathered lots of attention recently.
In binary classification:
· A flexible approach for relaxed constraints [Zafar et al., 2019],
· ERM guarantees [Donini et al., 2018],
· Textbook (WIP) on fairness in ML [Barocas et al., 2019].

Fairness in ranking became only recently a research topic,
mostly tackled by the information retrieval (IR) community.
Some authors:
· modify a fixed score to induce a notion of fairness
[Zehlike et al., 2017, Biega et al., 2018],
· introduce fairness in exposure over several rankings
[Singh and Joachims, 2018, Singh and Joachims, 2019],
· use a notion of fairness based on the AUC
[Borkan et al., 2019, Beutel et al., 2019].

Outline

Bipartite ranking (1/2)

Scoring: $(X, Y) \sim P$ and $(X, Y) \in \mathcal{X} \times \mathcal{Y}$ with $\mathcal{Y} = \{-1, 1\}$,
learn a score $s : \mathcal{X} \to \mathbb{R}$ from data $\{(X_i, Y_i)\}_{i=1}^n \overset{i.i.d.}{\sim} P$.

Objective: Order new elements $X'_1, \ldots, X'_m$ by relevance,
i.e. by decreasing posterior probability $\eta(x) := \mathbb{P}\{Y = +1 \mid X = x\}$.

Perf. measure: The ROC curve: the true positive rate (TPR) for any false positive rate (FPR) for testing $Y = +1$ with $s(X) > t$.

Introduce the distributions (cdf) of $s(X)|Y = -1$ and $s(X)|Y = +1$ as:

$$H_s(t) = \mathbb{P}\{s(X) \leq t \mid Y = -1\} \quad \text{and} \quad G_s(t) = \mathbb{P}\{s(X) \leq t \mid Y = +1\}.$$

Let $\bar{F} = 1 - F$ and define the pseudo-inverse of F as:

$$F^{-1} : u \mapsto \inf\{t \mid F(t) > u\}.$$

Bipartite ranking (2/2)

The FPR (resp. TPR) of s at threshold t is equal to $\overline{H}_s(t)$ (resp. $\overline{G}_s(t)$).
Formally, the ROC and AUC write:

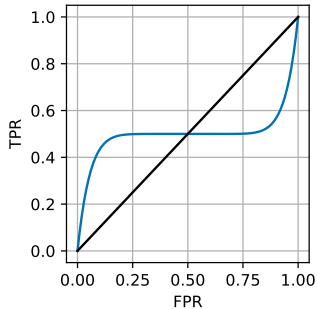
$$\mathrm{ROC}_{H_s,G_s}(\alpha) = \overline{G}_s \circ \overline{H}_s^{-1}(\alpha) \quad \text{and} \quad \mathrm{AUC}_{H_s,G_s} = \int_0^1 \mathrm{ROC}_{H_s,G_s}(\alpha)\, d\alpha.$$

The ROC **measures the difference** between two cdfs in $\mathbb{R}$.

Specifically, given two distributions F, F' on $\mathbb{R}$:

$$\forall \alpha \in [0,1], \quad \mathrm{ROC}_{F,F'}(\alpha) = \alpha \quad \Leftrightarrow \quad F = F'.$$

The AUC is a **scalar summary** of the ROC.



Fair bipartite ranking

We adopt the **same probabilistic model** as fair binary classification.

Denote by:

$$H_s^{(z)} \text{ the cdf of } s(X)|Y=-1, Z=z,$$
$$G_s^{(z)} \text{ the cdf of } s(X)|Y=+1, Z=z,$$

for any $z \in \mathcal{Z}$ with $\mathcal{Z} = \{0, 1\}$.

Our contributions

We provide:

· A **general formulation** for AUC-based fairness constraints,

· A new, restrictive type of **constraint**: ROC-based constraints,

· **Guarantees** and a gradient descent (GD) **method** for learning under both AUC and ROC-based constraints.

Outline

Examples of AUC constraints in the literature

Intra-group pairwise and BNSP AUC fairness ([Borkan et al., 2019, Beutel et al., 2019])

$$\mathrm{AUC}_{H_s^{(0)},G_s^{(0)}} = \mathrm{AUC}_{H_s^{(1)},G_s^{(1)}},$$
$$\mathrm{AUC}_{H_s,G_s^{(0)}} = \mathrm{AUC}_{H_s,G_s^{(1)}}.$$

BPSN AUC fairness and zero Average Equality Gap (AEG)
[Borkan et al., 2019, Kallus and Zhou, 2019]

$$\mathrm{AUC}_{H_s^{(0)},G_s} = \mathrm{AUC}_{H_s^{(1)},G_s},$$
$$\mathrm{AUC}_{G_s,G_s^{(0)}} = \mathrm{AUC}_{G_s,G_s^{(1)}}.$$

$x\mathrm{AUC}$ parity ([Beutel et al., 2019, Kallus and Zhou, 2019])

$$\mathrm{AUC}_{H_s^{(0)},G_s^{(1)}} = \mathrm{AUC}_{H_s^{(1)},G_s^{(0)}}.$$

A general AUC constraint

Introduce all relevant distributions as $D(s) = (H_s^{(0)}, H_s^{(1)}, G_s^{(0)}, G_s^{(1)})$.
Any known AUC constraint writes as:

$$\mathrm{AUC}_{\alpha^\top D(s), \beta^\top D(s)} = \mathrm{AUC}_{\alpha'^\top D(s), \beta'^\top D(s)}, \tag{1}$$

with $\alpha, \alpha', \beta, \beta' \in [0,1]^4$ and any of those sums to 1.

Theorem 1

The following propositions are equivalent:

1. *Eq.* (1) *is verified when* $X|Y = y, Z = 0$ *and* $X|Y = y, Z = 1$ *have same distribution for any* $y \in \mathcal{Y}$ *and* $\eta(X)$ *is not almost surely constant.*
2. $(e_1 + e_2)^\top [(\alpha - \alpha') - (\beta - \beta')] = 0$.
3. *Eq.* (1) *is equivalent to* $\Gamma^\top C(s) = 0$ *where* $\Gamma \in \mathbb{R}^5$ *and* $C(s) \in \mathbb{R}^5$ *are 5 elementary measures.*

Learning with AUC constraints

Let $\mathcal{S}$ be a proposal family of scores. We consider L_λ, where $\lambda > 0$ is fixed:

$$\max_{s \in \mathcal{S}} L_\lambda(s) \quad \text{with} \quad L_\lambda(s) = \mathrm{AUC}_{H_s, G_s} - \lambda \left| \mathrm{AUC}_{H_s^{(0)}, G_s^{(0)}} - \mathrm{AUC}_{H_s^{(1)}, G_s^{(1)}} \right|,$$

and its solution is written s^*_λ.

Using the sample $\mathcal{D}_n = \{(X_i, Y_i, Z_i)\}_{i=1}^n$, we replace the AUC's above by estimators, using counterparts of H_s, $H_s^{(z)}$, G_s, $G_s^{(z)}$. It gives $\widehat{L}_\lambda$ whose maximizer is written $\widehat{s}_\lambda$.

Theorem 2

Assume that $\mathcal{S}$ is VC-major with VC-dim $V < +\infty$,
and there exists $\epsilon > 0$, $\epsilon \leq \mathbb{P}\{Y = y, Z = z\}$ for any $y \in \mathcal{Y}, z \in \mathcal{Z}$.
Then, for any $\delta > 0$ and $n > 1$, with probability $\geq 1 - \delta$:

$$\epsilon^2 \left[L_\lambda(s^*_\lambda) - L_\lambda(\widehat{s}_\lambda) \right] \leq C \sqrt{\frac{V}{n}} + (8\lambda + 2\epsilon) \sqrt{\frac{\log(13/\delta)}{n-1}} + O(n^{-1}).$$

Sketch of Proof for Theorem 2

Recall that:

$$L_\lambda(s) = \mathrm{AUC}_{H_s, G_s} - \lambda \left| \mathrm{AUC}_{H_s^{(0)}, G_s^{(0)}} - \mathrm{AUC}_{H_s^{(1)}, G_s^{(1)}} \right|.$$

Empirical AUC's are "almost U-statistics". Given a sample $(X_i)_{i=1}^n$, consider

$$U_n(h) := \frac{2}{n(n-1)} \sum_{i<j} h(X_i, X_j) = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \frac{1}{\lfloor n/2 \rfloor} \sum_{i=1}^{\lfloor n/2 \rfloor} h(X_{\sigma(i)}, X_{\sigma(i+\lfloor n/2 \rfloor)}), \qquad (2)$$

where the second equality is the **first Hoeffding decomposition**.

Eq. (2), Jensen's inequality and usual results for empirical processes,
we can control the deviations of empirical AUC's.

Optimizing L_λ by gradient descent

$\widehat{L}_\lambda$ is not continuous, to relax it:

We replace $x \mapsto \mathbb{I}\{x \geq 0\}$ by a logistic $\sigma : x \mapsto 1/(1+e^{-x})$.

Introduce a parameter $c \in [-1, +1]$, our relaxed objective writes:

$$\widetilde{L}_\lambda(s) := \widetilde{\mathrm{AUC}}_{H_s, G_s} - \lambda \cdot c \left(\widetilde{\mathrm{AUC}}_{H_s^{(1)}, G_s^{(1)}} - \widetilde{\mathrm{AUC}}_{H_s^{(0)}, G_s^{(0)}} \right).$$

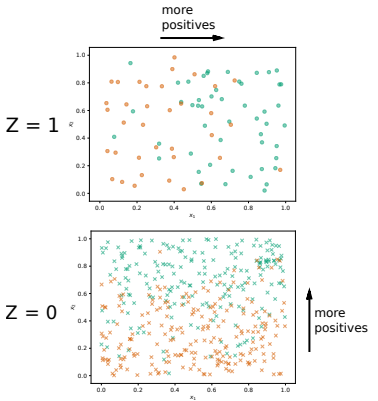
We modify c every n_{adapt} iterations, based on stats computed on a validation dataset:
$\rightarrow$ if the term in the constraint is positive then $c = \min(c + \Delta c, 1)$ with $\Delta c > 0$,
$\rightarrow$ otherwise $c = \max(c - \Delta c, -1)$,

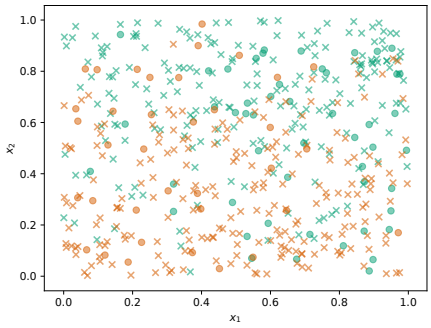
We normalize s with moving means and averages (like BatchNorm).

Toy example: fairness with AUC constraints (1/2)

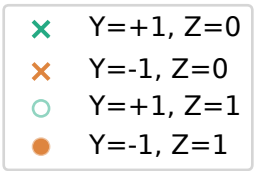
Set $\mathcal{X} = [0,1]^2$, with $X|Z = z$ uniform, and for $x = (x_1\ x_2)^\top \in \mathcal{X}$,
$\eta^{(0)}(x) = x_1$ and $\eta^{(1)}(x) = x_2$ where $\eta^{(z)}(x) = \mathbb{P}\{Y = +1 | Z = z, X = x\}$.

Fix $\mathbb{P}\{Z = 1\} = 17/20$, *i.e.* $Z = 1$ is the majority group.



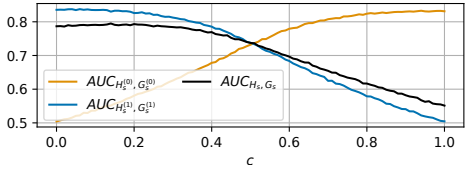


Example sample from P

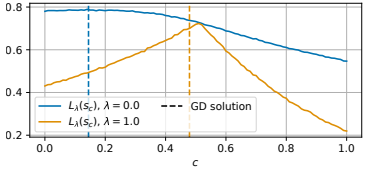


Toy example: fairness with AUC constraints (2/2)

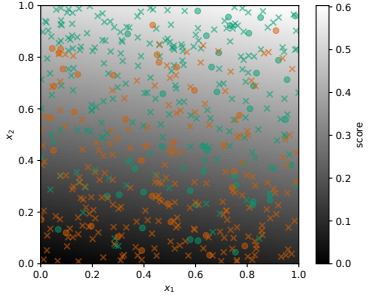
Consider a family of scores $\{s_c\}_{c\in[0,1]}$, with $s_c(x) = cx_1 + (1-c)x_2$.



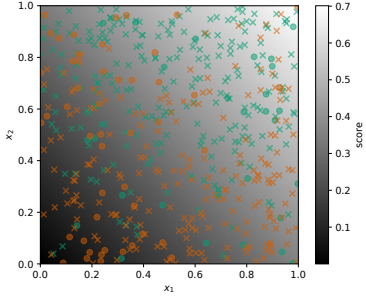
Values of the AUC's for any c.



Solutions s_c with AUC fairness.



Solution s_c with $\lambda = 0$.



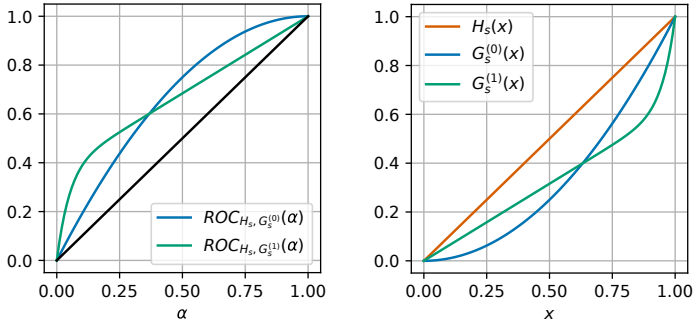
Solution s_c with $\lambda = 1$.

Outline

Limitations of AUC constraints

Below, with $s \in [0, 1]$, an AUC constraint is verified.

However, $\sup_{t\in[0,1]} |G_s^{(0)}(t) - G_s^{(1)}(t)| \approx 0.10$.



Mean value theorem: Let h, g, h', g' cdfs on $\mathbb{R}$ s.t. $\mathrm{ROC}_{h,g}$ and $\mathrm{ROC}_{h',g'}$ continuous. If $\mathrm{AUC}_{h,g} = \mathrm{AUC}_{h',g'}$, $\exists \alpha \in (0, 1)$ s.t. $g \circ h^{-1}(\alpha) = g' \circ h'^{-1}(\alpha)$.

Conclusion: An AUC constraint imposes a “pointwise constraint”.

Learning with pointwise ROC constraints

To measure the difference between cdfs for $Z = 0$ and $Z = 1$, let:

$$\Delta_{H,\alpha}(s) = \mathrm{ROC}_{H_s^{(0)}, H_s^{(1)}}(\alpha) - \alpha \quad \text{and} \quad \Delta_{G,\alpha}(s) = \mathrm{ROC}_{G_s^{(0)}, G_s^{(1)}}(\alpha) - \alpha.$$

Introduce a sum of m_H pointwise constraints for $\Delta_{H,\cdot}$ and m_G for $\Delta_{G,\cdot}$, maximize L_Λ in $\mathcal{S}$.

$$L_\Lambda(s) := \mathrm{AUC}_{H_s, G_s} - \sum_{k=1}^{m_H} \lambda_H^{(k)} |\Delta_{H,\alpha_H^{(k)}}(s)| - \sum_{k=1}^{m_G} \lambda_G^{(k)} |\Delta_{G,\alpha_G^{(k)}}(s)|$$

which gives the score s_Λ^*. The empirical counterpart of L_Λ is $\widehat{L}_\Lambda$, its maximizer is $\widehat{s}_\Lambda$.

Theorem 3

Assume that $\exists M, \kappa > 0$ *s.t.* $M \leq D_k'(s) \leq M \cdot \kappa$ *for all* $k \in [\![1, 4]\!], s \in \mathcal{S}$.
Under the assumptions of Theorem 2,

$$\epsilon^2 \cdot [L_\Lambda(s_\Lambda^*) - L_\Lambda(\widehat{s}_\Lambda)] \leq C_{\lambda,\epsilon,\kappa} \sqrt{\frac{V}{n}} + C'_{\Lambda,\epsilon,\kappa} \sqrt{\frac{\log(19/\delta)}{n-1}} + O(n^{-1}).$$

Sketch of proof for Theorem 3

Recall that:

$$L_\Lambda(s) := \mathrm{AUC}_{H_s,G_s} - \sum_{k=1}^{m_H} \lambda_H^{(k)} |\Delta_{H,\alpha_H^{(k)}}(s)| - \sum_{k=1}^{m_G} \lambda_G^{(k)} |\Delta_{G,\alpha_G^{(k)}}(s)|$$

The AUC term is already dealt with, see Theorem 2. From [Hsieh and Turnbull, 1996],

$$\begin{aligned}\mathrm{ROC}_{\widehat{G}_s^{(0)},\widehat{G}_s^{(1)}}(\alpha) - \mathrm{ROC}_{G_s^{(0)},G_s^{(1)}}(\alpha) = &\left[G_s^{(1)} \circ (G_s^{(0)})^{-1} - G_s^{(1)} \circ (\widehat{G}_s^{(0)})^{-1}\right](1-\alpha) \\ &+ \left[G_s^{(1)} \circ (\widehat{G}_s^{(0)})^{-1} - \widehat{G}_s^{(1)} \circ (\widehat{G}_s^{(0)})^{-1}\right](1-\alpha),\end{aligned}$$

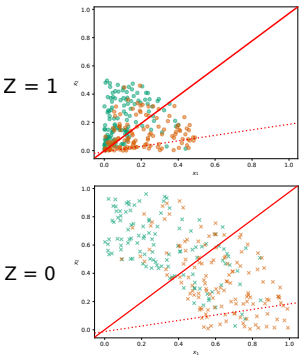
A uniform bound over α bounds easily the second term.

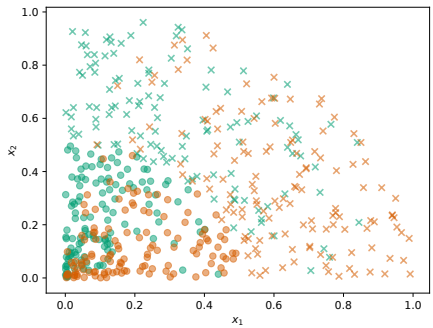
The first term uses the condition on the derivative, followed by the equality between the deviation of the standard and quantile uniform empirical processes.
[Shorack and Wellner, 1989]

Toy example: fairness with ROC constraints (1/2)

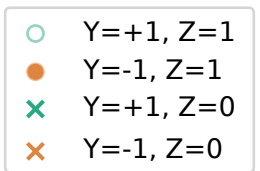
Set $\mathcal{X} = [0,1]^2$ and $\eta^{(0)}(x) = \eta^{(1)}(x) = (2/\pi) \cdot \arctan(x_2/x_1)$.

$$\mu^{(0)}(x) = \frac{16}{\pi}\mathbb{I}\left\{x^2 + y^2 \leq \frac{1}{2}\right\} \quad \text{and} \quad \mu^{(1)}(x) = \frac{16}{3\pi}\mathbb{I}\left\{\frac{1}{2} \leq x^2 + y^2 \leq 1\right\},$$



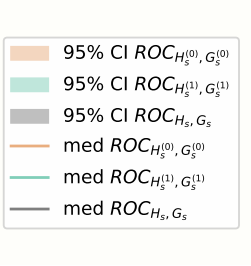


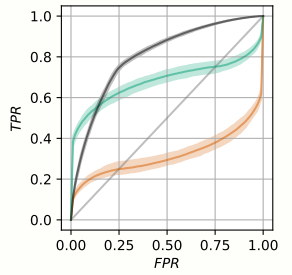
Example sample from P

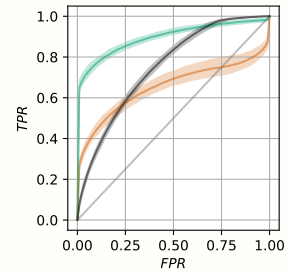


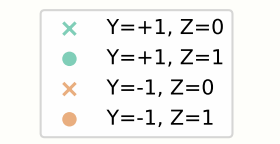
Toy example: fairness with ROC constraints (2/2)

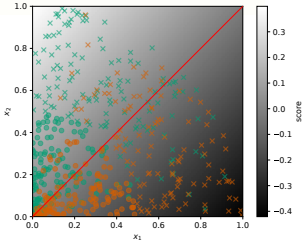
We chose for $\mathcal{S}$ a family of linear scores. The constraint we impose is $\Delta_{H,3/4} = 0$.

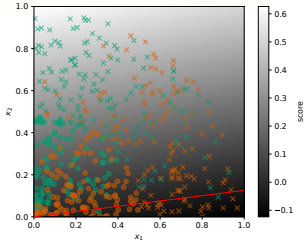












$\lambda = 0$ $\lambda = 1$

Outline

Datasets and parameters

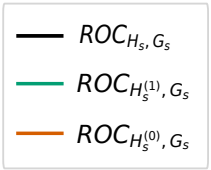
Here, we report on two datasets from the fairness literature:
· *Adult Income Dataset*, featured *e.g.* in [Donini et al., 2018],
Prediction: salary $\geq \$50K$ / sensitive group: gender.
· *Compas Dataset*, featured *e.g.* in [Donini et al., 2018],
Prediction: recidivist or not / sensitive group: ethnicity.

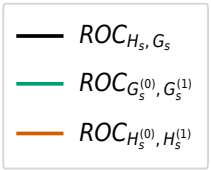
AUC-based constraints:
Different constraints are used, depending on the dataset.

ROC-based constraints:
To align the dist. of low FPR's and TPR's between $Z = 0$ and $Z = 1$,
we penalize high $|\Delta_{H,1/8}(s)|$, $|\Delta_{H,1/4}(s)|$, $|\Delta_{G,1/8}(s)|$ and $|\Delta_{G,1/4}(s)|$.

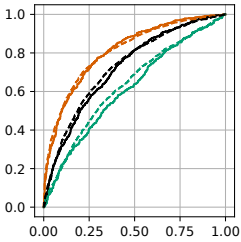
Results - Compas

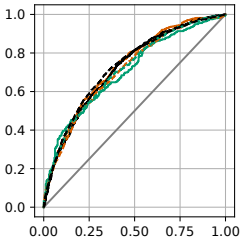
0: caucasian / 1: ethnic minority



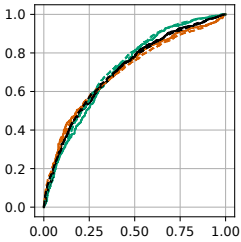


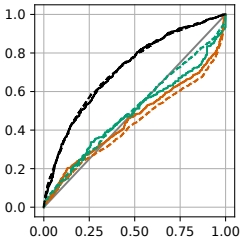
No constraint



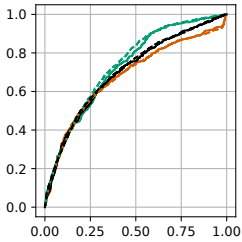


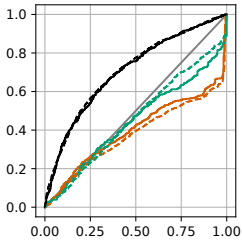
AUC Fairness





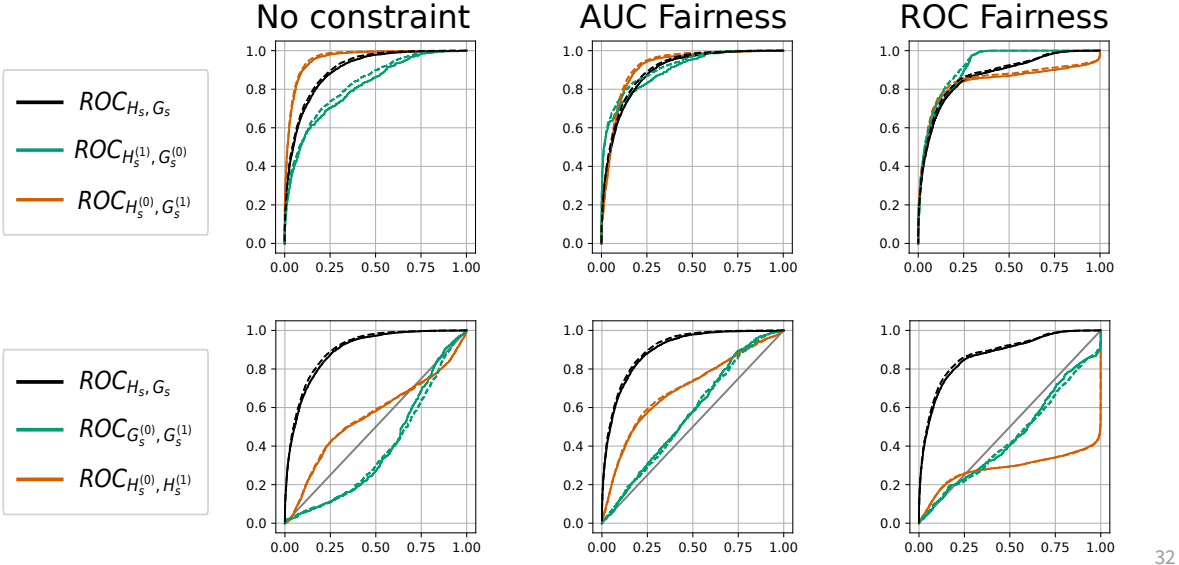
ROC Fairness





Results - Adult

0: woman / 1: man



Outline

Discussion

In this presentation, we have:

· proposed a **generalization of AUC-based** constraints.

· shown they imply a notion of **pointwise fairness**,

· proposed new **ROC-based** constraints.

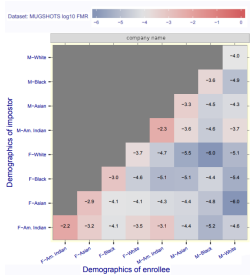
Limitations:

· No theoretical characterization of fairness/accuracy trade-offs.
Can we derive a notion of optimal fair scorer?
As in [Menon and Williamson, 2018, Chzhen et al., 2020]

· Limited optimization technique.
Tested on tabular data.

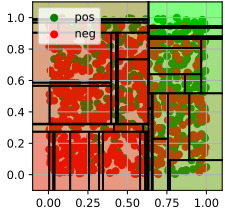
Future work

Extension:
This work extends to **similarity ranking**,
i.e. ranking pairs of items by similarity, see [Vogel et al., 2018].
Generalize results for $s(x, x')$? Efficient algorithms?



FRVT: FPR by ethnicity at fixed t.

Future work:
Fair constraints for ROC optimization,
based on recursive partitioning, see [Clémençon et al., 2010].



Thank you !

References I


Barocas, S., Hardt, M., and Narayanan, A. (2019).
Fairness and Machine Learning.
fairmlbook.org.
`http://www.fairmlbook.org`.


Beutel, A., Chen, J., Doshi, T., Qian, H., Wei, L., Wu, Y., Heldt, L., Zhao, Z., Hong, L., Chi, E. H., and Goodrow, C. (2019).
Fairness in recommendation ranking through pairwise comparisons.
In *KDD*.


Biega, A. J., Gummadi, K. P., and Weikum, G. (2018).
Equity of attention: Amortizing individual fairness in rankings.
In *SIGIR*.


Borkan, D., Dixon, L., Sorensen, J., Thain, N., and Vasserman, L. (2019).
Nuanced metrics for measuring unintended bias with real data for text classification.
arXiv:1903.04561.


Chzhen, E., Denis, C., Hebiri, M., Oneto, L., and Pontil, M. (2020).
Fair Regression via Plug-in Estimator and Recalibration With Statistical Guarantees.
HAL, archives ouvertes.


Clémençon, S., Depecker, M., and Vayatis, N. (2010).
Adaptive partitioning schemes for bipartite ranking.
Machine Learning.

References II


Donini, M., Oneto, L., Ben-David, S., Shawe-Taylor, J. S., and Pontil, M. (2018).
Empirical risk minimization under fairness constraints.
In *NeurIPS*.


Hsieh, F. and Turnbull, B. W. (1996).
Nonparametric and semiparametric estimation of the receiver operating characteristic curve.
The Annals of Statistics, 24(1):25–40.


Kallus, N. and Zhou, A. (2019).
The fairness of risk scores beyond classification: Bipartite ranking and the XAUC metric.
In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019*, pages 3433–3443.


Menon, A. K. and Williamson, R. C. (2018).
The cost of fairness in binary classification.
In *Conference on Fairness, Accountability and Transparency, FAT 2018*, volume 81 of *Proceedings of Machine Learning Research*, pages 107–118. PMLR.


Shorack, G. and Wellner, J. a. (1989).
Empirical Processes with applications to Statistics.
SIAM.

References III


Singh, A. and Joachims, T. (2018).
Fairness of exposure in rankings.
In *KDD*.


Singh, A. and Joachims, T. (2019).
Policy learning for fairness in ranking.
In *NeurIPS*.


Vogel, R., Bellet, A., and Clémençon, S. (2018).
A probabilistic theory of supervised similarity learning for pointwise ROC curve optimization.
In *ICML*. PMLR.


Zafar, M. B., Valera, I., Gomez-Rodriguez, M., and Gummadi, K. P. (2019).
Fairness constraints: A flexible approach for fair classification.
Journal of Machine Learning Research, 20(75):1–42.


Zehlike, M., Bonchi, F., Castillo, C., Hajian, S., Megahed, M., and Baeza-Yates, R. (2017).
FA*IR: A Fair Top-k Ranking Algorithm.
In *CIKM*.