$$\hat{\mathfrak{R}}_m(\mathcal{F} - \widehat{\mathbb{E}}_{\boldsymbol{x}}[\mathcal{F}], \boldsymbol{x}) \doteq \mathbb{E}_{\boldsymbol{\sigma}}\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{m}\sum_{i=1}^{m}\boldsymbol{\sigma}_i\big(f(\boldsymbol{x}_i) - \widehat{\mathbb{E}}_{\boldsymbol{x}}[f]\big)\right|\right]$$

An Axiomatic Theory of Provably-Fair Welfare-Centric Machine Learning

Cyrus Cousins

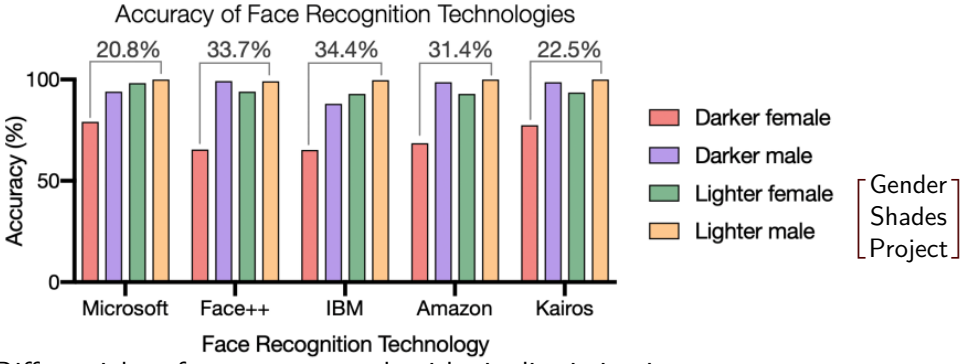
Brown University Department of Computer Science

July 2021

The Twenty-Second ACM Conference on Economics and Computation

Fairness in Machine Learning (or Lack Thereof)

- ML systems often trained on *group A*, then applied to *group B*

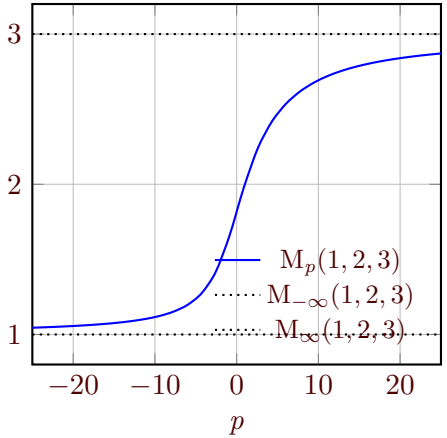


- Differential performance $\implies$ algorithmic discrimination
  - Facial recognition and policing
  - Speech recognition and accessibliity
  - Many more examples

The Power Mean

Suppose vector $\boldsymbol{\ell} = (\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_g)$ representing *utility* or *loss* across a population

$$\mathrm{M}_p(\boldsymbol{\ell}) \doteq \begin{cases} p = -\infty & \min_{i \in 1,\ldots,g} \boldsymbol{\ell}_i \\ p = 0 & \sqrt[n]{\prod_{i=1}^{n} \boldsymbol{\ell}_i} \\ p = \infty & \max_{i \in 1,\ldots,g} \boldsymbol{\ell}_i \\ \text{else} & \sqrt[p]{\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\ell}_i^p} \end{cases}$$



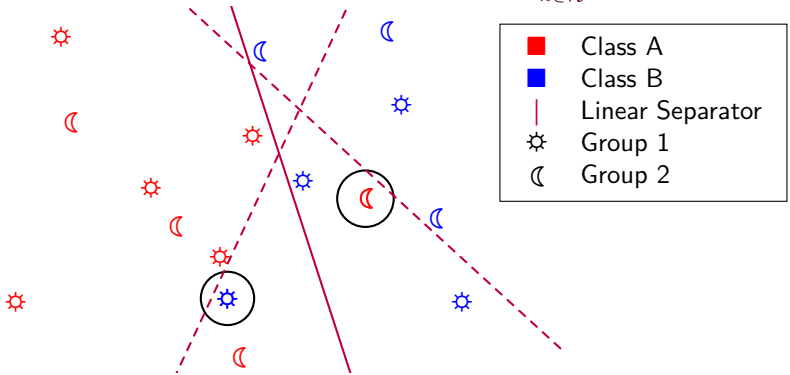
- *Smooth interpolation* between *min*, *arithmetic mean*, and *max*
- Other special cases: *geometric*, *harmonic*, and *quadratic* means

Learning Fair Linear Classifiers

- We can handle each group individually:

$$\hat{\mathrm{R}}(h; \boldsymbol{x}_i, \boldsymbol{y}_i) \doteq \frac{1}{m}\sum_{j=1}^{m} \ell\big(h(\boldsymbol{x}_{i,j}), \boldsymbol{y}_{i,j}\big); \quad \forall i: \ \hat{h}_i \doteq \operatorname*{argmin}_{h \in \mathcal{H}} \hat{\mathrm{R}}(h; \boldsymbol{x}_i, \boldsymbol{y}_i)$$

- What is the best classifier *overall?*
  - *Empirical malfare minimization* $\hat{h} \doteq \operatorname*{argmin}_{h \in \mathcal{H}} \mathrm{M}\left(\hat{\mathrm{R}}(h; \boldsymbol{x}_1, \boldsymbol{y}_1), \hat{\mathrm{R}}(h; \boldsymbol{x}_2, \boldsymbol{y}_2)\right)$



Visit my poster session for more information!

An Axiomatic Theory of Provably-Fair Welfare-Centric Machine Learning

Cyrus Cousins

Monday, July 19 @ 2:00pm - 2:30pm EST

&

Monday, July 19 @ 9:00pm - 9:30pm EST