Fair and Accurate Regression

Anna Deza[1], Andrés Gómez[2], Alper Atamtürk[1]

[1]Department of Industrial Engineering and Operations Research,
University of California, Berkeley

[2]Department of Industrial and Systems Engineering,
University of Southern California

Sunday, October 20 2024
INFORMS Annual Meeting

Berkeley
UNIVERSITY OF CALIFORNIA

BUSINESS INSIDER

Log in Subscribe

EDUCATION

The Princeton Review's pricing method has a terrible impact on Asians, study finds

Peter Jacobs Sep 2, 2015, 8:42 AM PDT

Share Save

Asian families living in the US are almost twice as likely to be charged higher prices by college test prep service The Princeton Review, according to a new study from ProPublica.

U.S. Equal Employment Opportunity Commission

Press Release
09-11-2023

iTutorGroup to Pay $365,000 to Settle EEOC Discriminatory Hiring Suit

Settles Federal Charges Tutoring Provider Programmed its Online Software to

TECH

How a computer algorithm caused a grading crisis in British schools

PUBLISHED FRI, AUG 21 2020·7:18 AM EDT | UPDATED FRI, AUG 21 2020·8:45 AM EDT

Sam Shead
@SAM_L_SHEAD

WATCH LIVE

KEY

- Approximately 39% of A-level results were downgraded by exam regulator

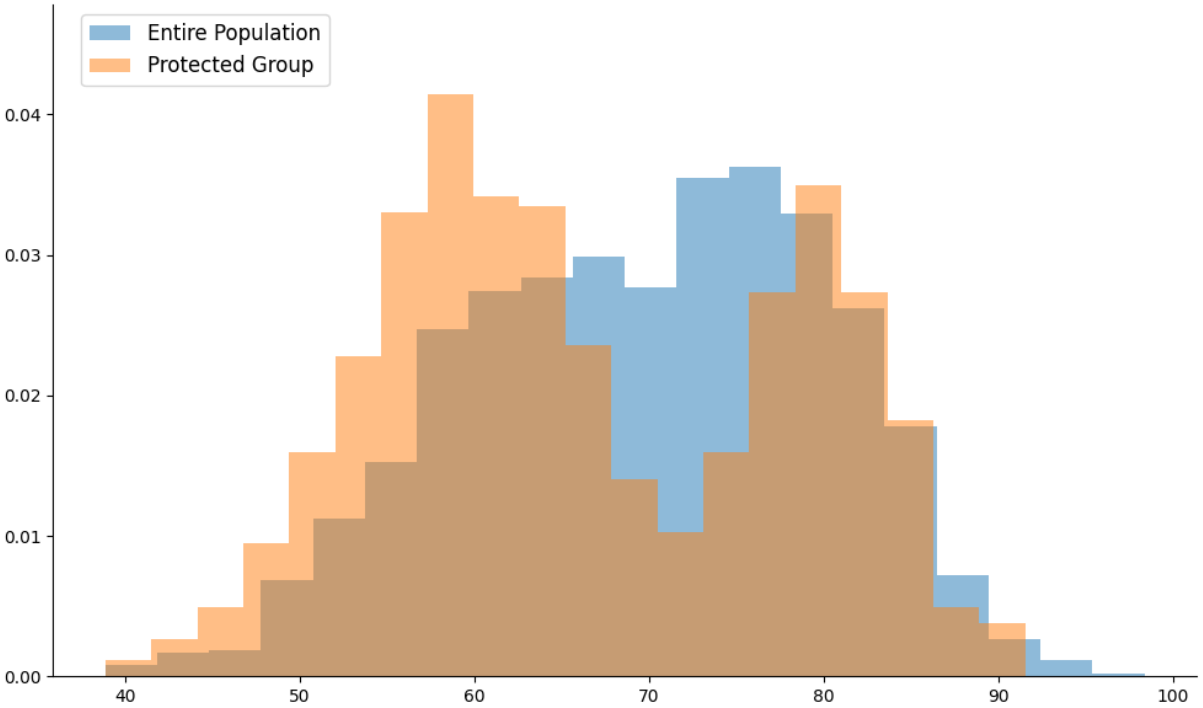
e the worst affected as the algorithm copied the
U.K.'s education system.

Racial bias found in widely used health care algorithm

An estimated 200 million people are affected each year by similar tools that are used in hospital networks

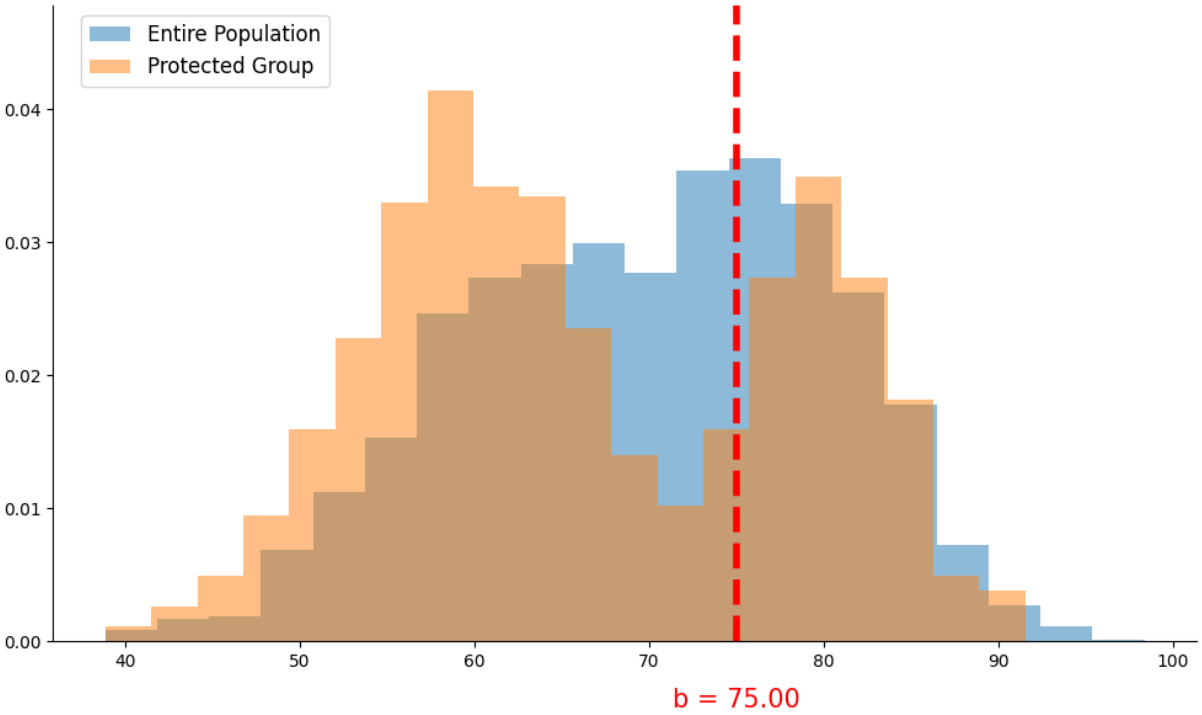
Example: College Admissions

Model predicts scores 0-100 used for college admission process. Fair?



Example: College Admissions

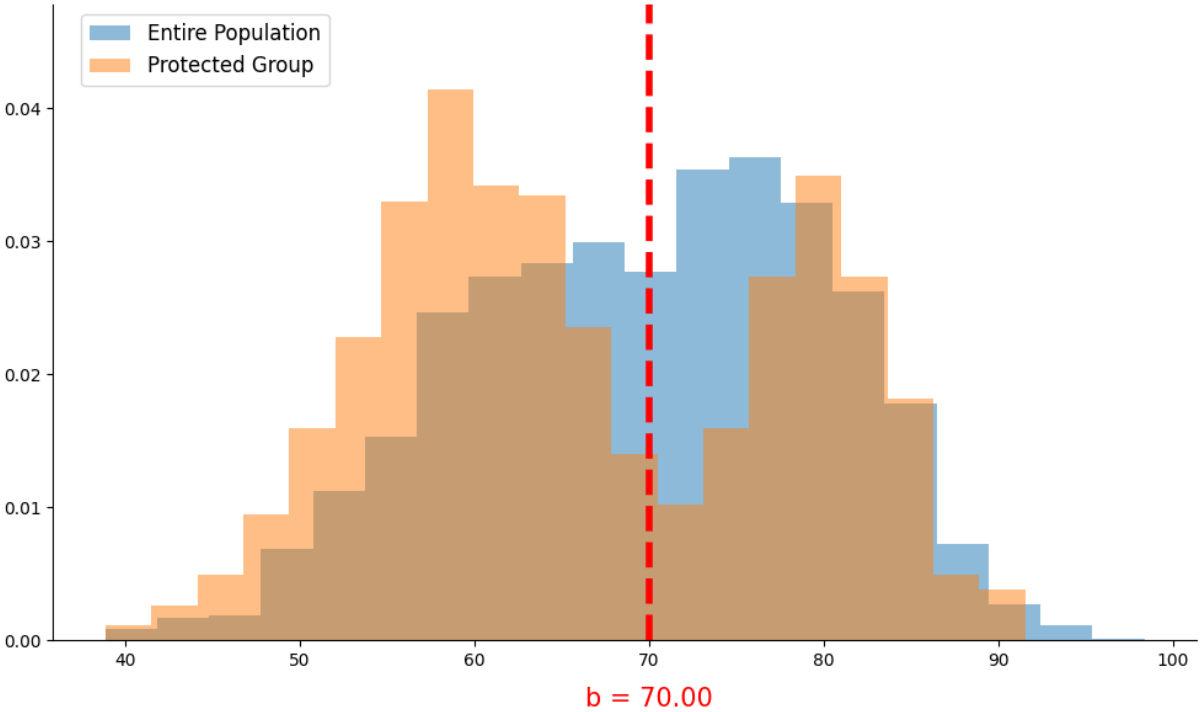
Model predicts scores 0-100 used for college admission process. Fair?



$$P(\text{score} > 75) = 0.35$$
$$P(\text{score} > 75 | \text{protected}) = 0.32$$

Example: College Admissions

Model predicts scores 0-100 used for college admission process. Fair?

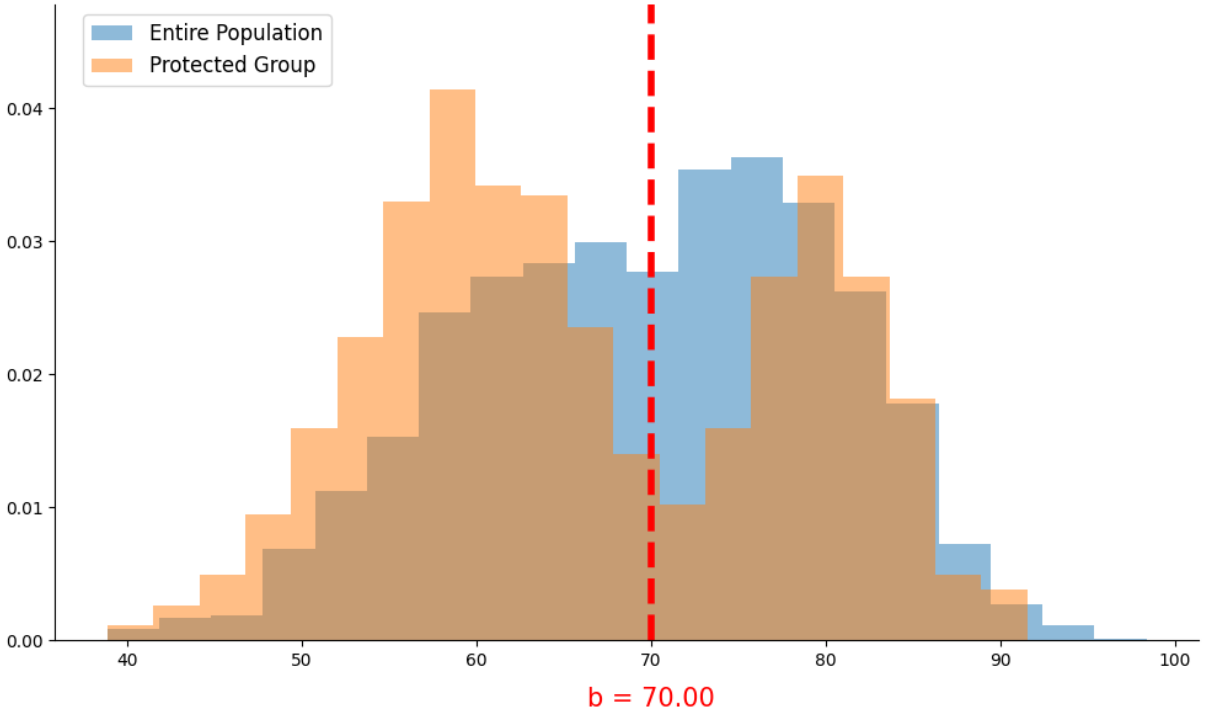


$$P(\text{score} > 75) = 0.35$$
$$P(\text{score} > 75 | \text{protected}) = 0.32$$

$$P(\text{score} > 70) = 0.52$$
$$P(\text{score} > 70 | \text{protected}) = 0.38$$

Example: College Admissions

Model predicts scores 0-100 used for college admission process. Fair?



$$P(\text{score} > 75) = 0.35$$
$$P(\text{score} > 75 | \text{protected}) = 0.32$$

$$P(\text{score} > 70) = 0.52$$
$$P(\text{score} > 70 | \text{protected}) = 0.38$$

- Fair Classification: fairness at a single b, vast existing literature
- Fair Regression: fairness at *all* b, far less work $\rightarrow$ **problem we address**

Fair Training Problem

Data: observations $i = 1, \ldots, m$, consisting of:

- $\boldsymbol{x}_i \in \mathbb{R}^n$ feature vector
- $y_i \in \mathbb{R}$ target label
- $a_i \in \{0, 1\}$ protected class status

Fair Training Problem

Data: observations $i = 1, \ldots, m$, consisting of:

- $\boldsymbol{x}_i \in \mathbb{R}^n$ feature vector
- $y_i \in \mathbb{R}$ target label
- $a_i \in \{0, 1\}$ protected class status

Training problem:

$$\min \sum_{i=1}^{m} \mathcal{L}(\hat{y}_i, y_i) + \lambda \max_{b \in \mathbb{R}} \left(\hat{P}(\hat{y} > b) - \hat{P}(\hat{y} > b | a = 1) \right)$$

Fair Training Problem

Data: observations $i = 1, \ldots, m$, consisting of:

- $\boldsymbol{x}_i \in \mathbb{R}^n$ feature vector
- $y_i \in \mathbb{R}$ target label
- $a_i \in \{0, 1\}$ protected class status

Training problem:

$$\min \sum_{i=1}^{m} \mathcal{L}(\hat{y}_i, y_i) + \lambda \max_{b \in \mathbb{R}} \left(\hat{P}(\hat{y} > b) - \hat{P}(\hat{y} > b | a = 1) \right)$$

$$P(\hat{y} > b) - P(\hat{y} > b | a = 1) = 0 \quad \forall b \in \mathbb{R} \longrightarrow \hat{y} \perp a$$

(resource allocation $\perp$ of a)

Fair Training Problem

Data: observations $i = 1, \ldots, m$, consisting of:

$\boldsymbol{x}_i \in \mathbb{R}^n$ feature vector

$y_i \in \mathbb{R}$ target label

$a_i \in \{0, 1\}$ protected class status

Training problem:

$$\min \sum_{i=1}^{m} \mathcal{L}(\hat{y}_i, y_i) + \lambda \max_{b \in \mathbb{R}} \left(\frac{1}{m} \sum_{i=1}^{m} \mathbb{1}(\hat{y}_i > b) - \frac{1}{m_1} \sum_{i=1:a_i=1}^{m} \mathbb{1}(\hat{y}_i > b) \right)$$

Fair Training Problem

Data: observations $i = 1, \ldots, m$, consisting of:

- $\boldsymbol{x}_i \in \mathbb{R}^n$ feature vector
- $y_i \in \mathbb{R}$ target label
- $a_i \in \{0, 1\}$ protected class status

Training problem:

$$\min \sum_{i=1}^{m} \mathcal{L}(\hat{y}_i, y_i) + \lambda \max_{b \in \mathbb{R}} \left(\frac{1}{m} \sum_{i=1}^{m} \mathbb{1}(\hat{y}_i > b) - \frac{1}{m_1} \sum_{i=1:a_i=1}^{m} \mathbb{1}(\hat{y}_i > b) \right)$$

Fair Training Problem

Data: observations $i = 1, \ldots, m$, consisting of:

- $\boldsymbol{x}_i \in \mathbb{R}^n$ feature vector
- $y_i \in \mathbb{R}$ target label
- $a_i \in \{0, 1\}$ protected class status

Training problem:

$$\min \sum_{i=1}^{m} \mathcal{L}(\hat{y}_i, y_i) + \lambda \max_{j \in 1, \ldots \ell} \left(\frac{1}{m} \sum_{i=1}^{m} \mathbb{1}(\hat{y}_i > b_j) - \frac{1}{m_1} \sum_{i=1:a_i=1}^{m} \mathbb{1}(\hat{y}_i > b_j) \right)$$

Discretize $\mathbb{R}$, only consider fairness at ℓ points $b_1 < b_2 < \ldots < b_\ell$

Fair Training Problem

Data: observations $i = 1, \ldots, m$, consisting of:

- $\boldsymbol{x}_i \in \mathbb{R}^n$ feature vector
- $y_i \in \mathbb{R}$ target label
- $a_i \in \{0, 1\}$ protected class status

Training problem:

$$\min \sum_{i=1}^{m} \mathcal{L}_i(\boldsymbol{w}^\top \boldsymbol{x}_i) + \lambda \max_{j \in 1, \ldots \ell} \left(\frac{1}{m} \sum_{i=1}^{m} \mathbb{1}(\boldsymbol{w}^\top \boldsymbol{x}_i > b_j) - \frac{1}{m_1} \sum_{i=1:a_i=1}^{m} \mathbb{1}(\boldsymbol{w}^\top \boldsymbol{x}_i > b_j) \right)$$

Discretize $\mathbb{R}$, only consider fairness at ℓ points $b_1 < b_2 < \ldots < b_\ell$

This presentation will focus on **linear regression**: $\hat{y}_i = \boldsymbol{w}^\top \boldsymbol{x}_i$

Literature Overview

$$\min \sum_{i=1}^{m} \mathcal{L}(\hat{y}_i, y_i) + \lambda \max_{j \in 1,\dots \ell} \left(\frac{1}{m} \sum_{i=1}^{m} \mathbb{1}(\hat{y}_i > b_j) - \frac{1}{m_1} \sum_{i=1:a_i=1}^{m} \mathbb{1}(\hat{y}_i > b_j) \right)$$

Literature Overview

$$\min \sum_{i=1}^{m} \mathcal{L}(\hat{y}_i, y_i) + \lambda \max_{j \in 1,\dots \ell} \left(\frac{1}{m} \sum_{i=1}^{m} \mathbb{1}(\hat{y}_i > b_j) - \frac{1}{m_1} \sum_{i=1:a_i=1}^{m} \mathbb{1}(\hat{y}_i > b_j) \right)$$

Existing methods:

- Convex proxies for fairness: Berk et al. (2017), Do et al. (2022)
- Reduction-based algorthims: Agarwal, Dudik & Wu (2019)
- MIO approach: Ye, Hanasusanto & Xie (2024)

A First Formulation for Fair Regression

Introduce $z_{ij} \in \{0,1\}$ to model $\mathbb{1}(\boldsymbol{w}^\top \boldsymbol{x}_i > b_j)$

$$\begin{aligned}
\min \ & \sum_{i=1}^{m} \mathcal{L}(\boldsymbol{w}^\top \boldsymbol{x}_i, y_i) + \lambda \max_{j \in 1,\dots \ell} \left(\frac{1}{m} \sum_{i=1}^{m} z_{ij} - \frac{1}{m_1} \sum_{i=1:a_i=1}^{m} z_{ij} \right) \\
\text{s.t.} \ & (\boldsymbol{w}^\top \boldsymbol{x}_i - b_j) z_{ij} \geq 0 \qquad j \in [\ell], \ i \in [m] \\
& (\boldsymbol{w}^\top \boldsymbol{x}_j - b_j)(1 - z_{ij}) \leq 0 \qquad j \in [\ell], \ i \in [m] \\
& \boldsymbol{z} \in \{0,1\}^{m \times \ell}.
\end{aligned}$$

A First Formulation for Fair Regression

Introduce $z_{ij} \in \{0,1\}$ to model $\mathbb{1}(\boldsymbol{w}^\top \boldsymbol{x}_i > b_j)$

$$\min \sum_{i=1}^{m} \mathcal{L}(\boldsymbol{w}^\top \boldsymbol{x}_i, y_i) + \lambda \max_{j \in 1,\dots \ell} \left(\frac{1}{m} \sum_{i=1}^{m} z_{ij} - \frac{1}{m_1} \sum_{i=1:a_i=1}^{m} z_{ij} \right)$$

$$\text{s.t. } (\boldsymbol{w}^\top \boldsymbol{x}_i - b_j) z_{ij} \geq 0 \qquad j \in [\ell],\ i \in [m] \qquad z_{ij} = 1 \Rightarrow \boldsymbol{w}^\top \boldsymbol{x}_i \geq b_j$$

$$(\boldsymbol{w}^\top \boldsymbol{x}_j - b_j)(1 - z_{ij}) \leq 0 \qquad j \in [\ell],\ i \in [m] \qquad z_{ij} = 0 \Rightarrow \boldsymbol{w}^\top \boldsymbol{x}_i \leq b_j$$

$$\boldsymbol{z} \in \{0,1\}^{m \times \ell}.$$

A First Formulation for Fair Regression

Relax $z_{ij} \in \{0,1\}$ to model $\mathbb{1}(\boldsymbol{w}^\top \boldsymbol{x}_i > b_j)$

$$\min \sum_{i=1}^{m} \mathcal{L}(\boldsymbol{w}^\top \boldsymbol{x}_i, y_i) + \lambda \max_{j \in 1,\dots \ell} \left(\frac{1}{m} \sum_{i=1}^{m} z_{ij} - \frac{1}{m_1} \sum_{i=1:a_i=1}^{m} z_{ij} \right)$$

$$\text{s.t. } (\boldsymbol{w}^\top \boldsymbol{x}_i - b_j) z_{ij} \geq 0 \qquad j \in [\ell],\ i \in [m] \qquad z_{ij} = 1 \Rightarrow \boldsymbol{w}^\top \boldsymbol{x}_i \geq b_j$$

$$(\boldsymbol{w}^\top \boldsymbol{x}_j - b_j)(1 - z_{ij}) \leq 0 \qquad j \in [\ell],\ i \in [m] \qquad z_{ij} = 0 \Rightarrow \boldsymbol{w}^\top \boldsymbol{x}_i \leq b_j$$

$$\boldsymbol{z} \in [0,1]^{m \times \ell}.$$

Uninformative relaxation:

- Continuous relaxation solution is vanilla regression with no fairness
- Why? The closure of the convex hull has no links between $\boldsymbol{w}$ and $\boldsymbol{z}$

Reformulation: Exploiting non-linear objective

$$\min \sum_{i=1}^{m} s_i + \lambda \max_{j \in 1,\ldots,\ell} \left(\frac{1}{m} \sum_{i=1}^{m} z_{ij} - \frac{1}{m_1} \sum_{i=1:a_i=1}^{m} z_{ij} \right)$$

$$\begin{aligned}
\text{s.t. } & \mathcal{L}(\boldsymbol{w}^\top \boldsymbol{x}_i) \leq s_i, && i \in [m] \\
& (\boldsymbol{w}^\top \boldsymbol{x}_i - b_j) z_{ij} \geq 0, && j \in [\ell],\ i \in [m] \\
& (\boldsymbol{w}^\top \boldsymbol{x}_i - b_j)(1 - z_{ij}) \leq 0, && j \in [\ell],\ i \in [m] \\
& z_{ij} \in \{0, 1\}, && j \in [\ell],\ i \in [m]
\end{aligned}$$

Reformulation: Exploiting non-linear objective

$$\min \sum_{i=1}^{m} s_i + \lambda \max_{j \in 1,\dots,\ell} \left(\frac{1}{m} \sum_{i=1}^{m} z_{ij} - \frac{1}{m_1} \sum_{i=1:a_i=1}^{m} z_{ij} \right)$$

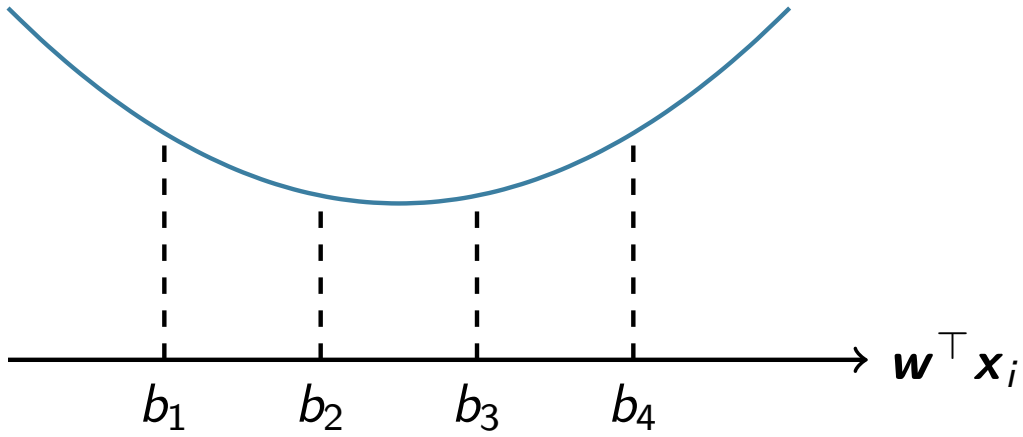
s.t. X_i

$$\mathcal{L}(\boldsymbol{w}^\top \boldsymbol{x}_i) \le s_i, \qquad i \in [m]$$
$$(\boldsymbol{w}^\top \boldsymbol{x}_i - b_j) z_{ij} \ge 0, \qquad j \in [\ell],\ i \in [m]$$
$$(\boldsymbol{w}^\top \boldsymbol{x}_i - b_j)(1 - z_{ij}) \le 0, \qquad j \in [\ell],\ i \in [m]$$
$$z_{ij} \in \{0, 1\}, \qquad j \in [\ell],\ i \in [m]$$

→ We will derive a strong reformulation by deriving a **compact extended formulation** for $\mathrm{cl\ conv}(X_i)$

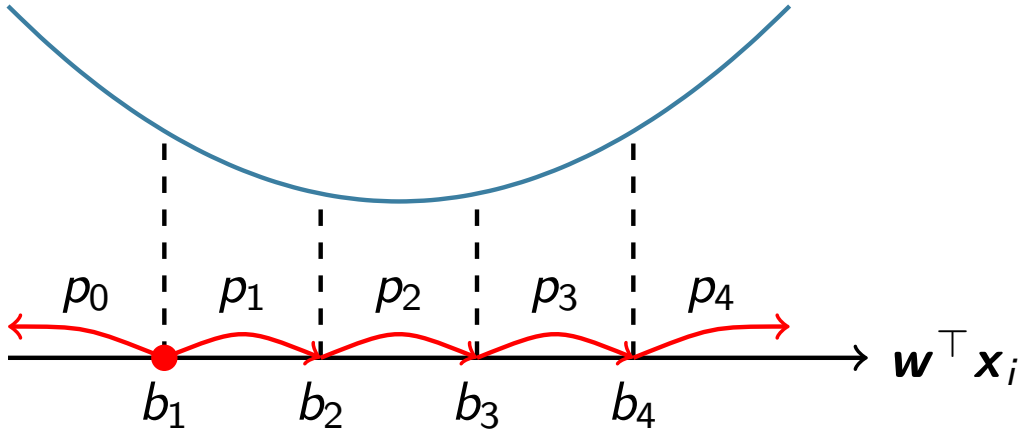
X_i: Epigraph of loss and indicators of prediction i

$$
\begin{aligned}
X_i = \{&(\boldsymbol{w}, \boldsymbol{z}, s) \in \mathbb{R}^n \times \{0,1\}^\ell \times \mathbb{R} : \\
&\mathcal{L}_i(\boldsymbol{w}^\top \boldsymbol{x}_i) \leq s \\
&(\boldsymbol{w}^\top \boldsymbol{x}_i - b_j) z_j \geq 0 \qquad j \in [\ell] \\
&(\boldsymbol{w}^\top \boldsymbol{x}_i - b_j)(1 - z_j) \geq 0 \quad j \in [\ell]\}
\end{aligned}
$$



X_i: Epigraph of loss and indicators of prediction i

$$
\begin{aligned}
X_i = \{&(\boldsymbol{w}, \boldsymbol{z}, s) \in \mathbb{R}^n \times \{0,1\}^\ell \times \mathbb{R} : \\
&\mathcal{L}_i(\boldsymbol{w}^\top \boldsymbol{x}_i) \leq s \\
&(\boldsymbol{w}^\top \boldsymbol{x}_i - b_j) z_j \geq 0 \qquad j \in [\ell] \\
&(\boldsymbol{w}^\top \boldsymbol{x}_i - b_j)(1 - z_j) \geq 0 \quad j \in [\ell]\}
\end{aligned}
$$

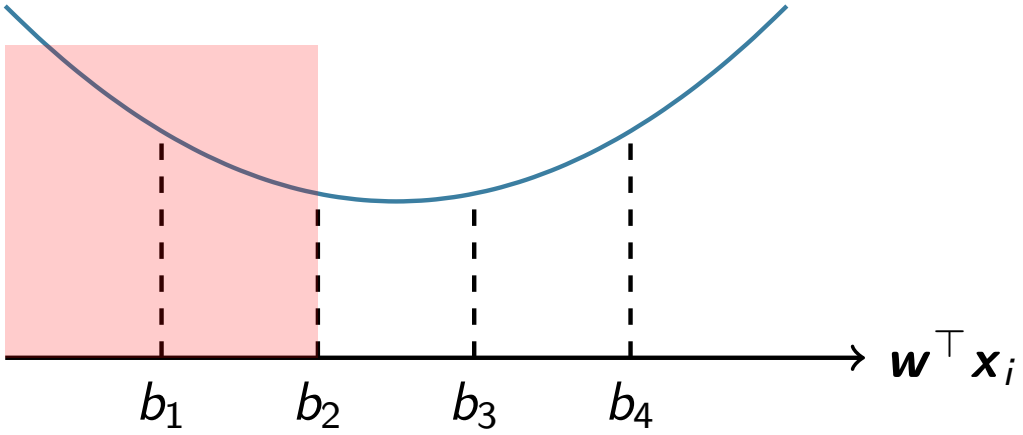


Extended set: $X_i = \{(\boldsymbol{w}, \boldsymbol{z}, s) \in \mathbb{R}^n \times \{0,1\}^\ell \times \mathbb{R} : \exists (p_0, \boldsymbol{p}) \in \mathbb{R}_+^{\ell+1} :$

$$\boldsymbol{w}^\top \boldsymbol{x}_i = b_1 + \sum_{j=1}^{\ell} p_j - p_0$$

X_i: Epigraph of loss and indicators of prediction i

$$
\begin{aligned}
X_i = \{&(\boldsymbol{w}, \boldsymbol{z}, s) \in \mathbb{R}^n \times \{0,1\}^\ell \times \mathbb{R} : \\
&\mathcal{L}_i(\boldsymbol{w}^\top \boldsymbol{x}_i) \leq s \\
&(\boldsymbol{w}^\top \boldsymbol{x}_i - b_j) z_j \geq 0 \qquad j \in [\ell] \\
&(\boldsymbol{w}^\top \boldsymbol{x}_i - b_j)(1 - z_j) \geq 0 \quad j \in [\ell]\}
\end{aligned}
$$



$z_2 = 0$

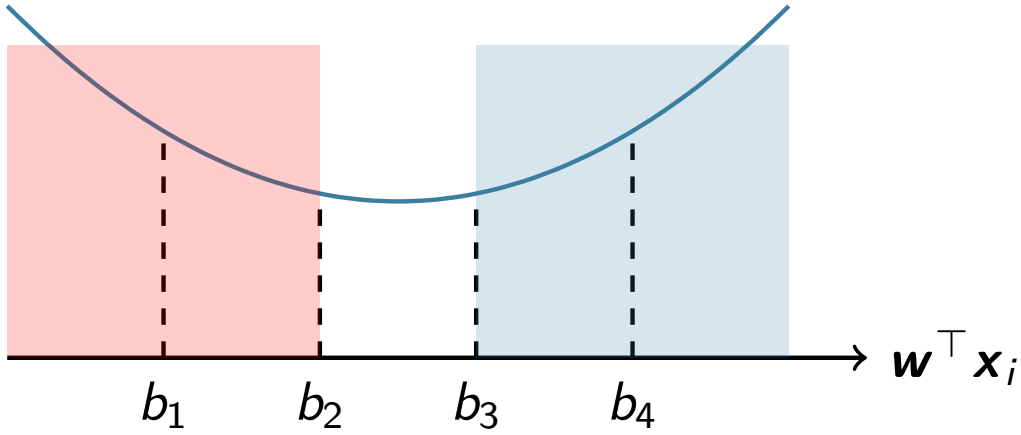
$\Rightarrow p_2 = 0$

Extended set: $X_i = \{(\boldsymbol{w}, \boldsymbol{z}, s) \in \mathbb{R}^n \times \{0,1\}^\ell \times \mathbb{R} : \exists (p_0, \boldsymbol{p}) \in \mathbb{R}_+^{\ell+1} :$

$$\boldsymbol{w}^\top \boldsymbol{x}_i = b_1 + \sum_{j=1}^{\ell} p_j - p_0$$

X_i: Epigraph of loss and indicators of prediction i

$$
\begin{aligned}
X_i = \{&(\boldsymbol{w}, \boldsymbol{z}, s) \in \mathbb{R}^n \times \{0,1\}^\ell \times \mathbb{R} : \\
&\mathcal{L}_i(\boldsymbol{w}^\top \boldsymbol{x}_i) \leq s \\
&(\boldsymbol{w}^\top \boldsymbol{x}_i - b_j) z_j \geq 0 \qquad j \in [\ell] \\
&(\boldsymbol{w}^\top \boldsymbol{x}_i - b_j)(1 - z_j) \geq 0 \quad j \in [\ell]\}
\end{aligned}
$$



$z_2 = 0$
$\Rightarrow p_2 = 0$

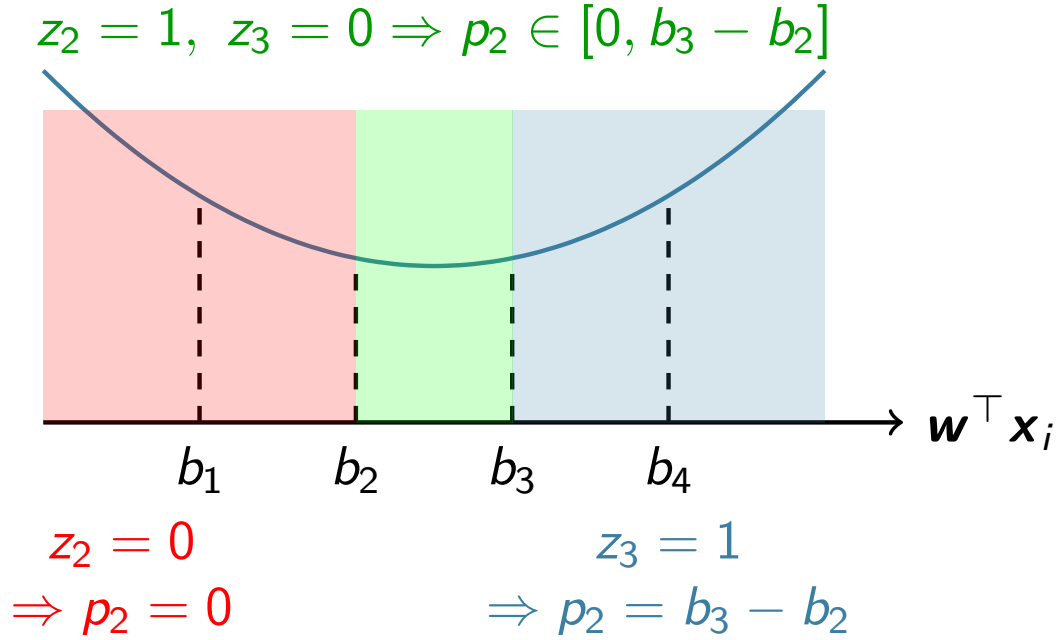
$z_3 = 1$
$\Rightarrow p_2 = b_3 - b_2$

Extended set: $X_i = \{(\boldsymbol{w}, \boldsymbol{z}, s) \in \mathbb{R}^n \times \{0,1\}^\ell \times \mathbb{R} : \exists (p_0, \boldsymbol{p}) \in \mathbb{R}_+^{\ell+1} :$

$$\boldsymbol{w}^\top \boldsymbol{x}_i = b_1 + \sum_{j=1}^{\ell} p_j - p_0$$

X_i: Epigraph of loss and indicators of prediction i

$$
\begin{aligned}
X_i = \{&(\boldsymbol{w}, \boldsymbol{z}, s) \in \mathbb{R}^n \times \{0,1\}^\ell \times \mathbb{R} : \\
&\mathcal{L}_i(\boldsymbol{w}^\top \boldsymbol{x}_i) \leq s \\
&(\boldsymbol{w}^\top \boldsymbol{x}_i - b_j) z_j \geq 0 \qquad j \in [\ell] \\
&(\boldsymbol{w}^\top \boldsymbol{x}_i - b_j)(1 - z_j) \geq 0 \quad j \in [\ell]\}
\end{aligned}
$$

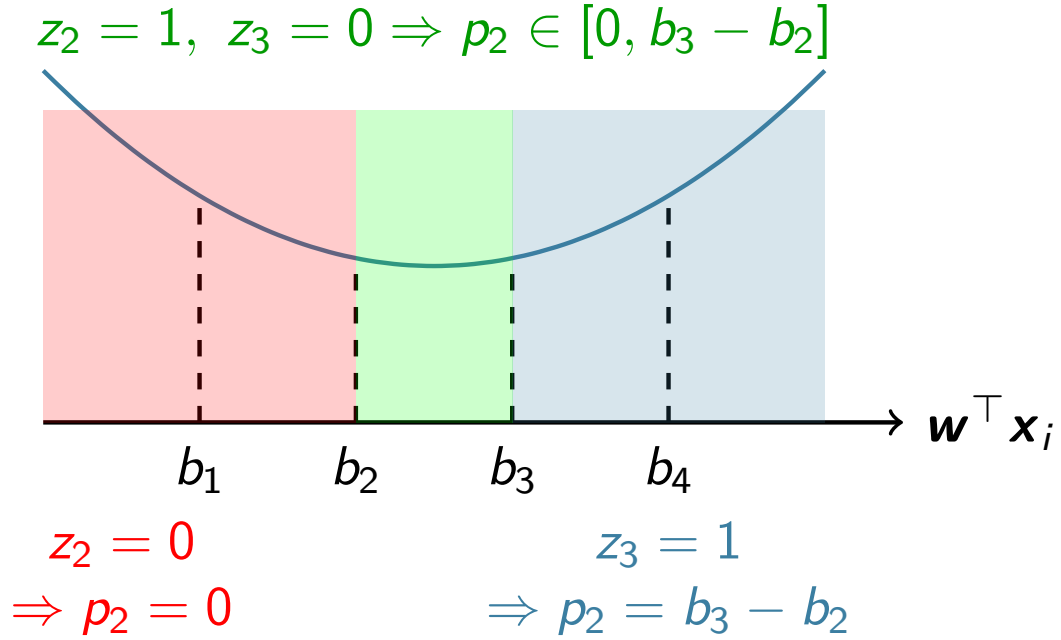


Extended set: $X_i = \{(\boldsymbol{w}, \boldsymbol{z}, s) \in \mathbb{R}^n \times \{0,1\}^\ell \times \mathbb{R} : \exists (p_0, \boldsymbol{p}) \in \mathbb{R}_+^{\ell+1} :$

$$
\boldsymbol{w}^\top \boldsymbol{x}_i = b_1 + \sum_{j=1}^{\ell} p_j - p_0
$$

X_i: Epigraph of loss and indicators of prediction i

$$
\begin{aligned}
X_i = \{&(\boldsymbol{w}, \boldsymbol{z}, s) \in \mathbb{R}^n \times \{0,1\}^\ell \times \mathbb{R} : \\
&\mathcal{L}_i(\boldsymbol{w}^\top \boldsymbol{x}_i) \leq s \\
&(\boldsymbol{w}^\top \boldsymbol{x}_i - b_j) z_j \geq 0 \qquad j \in [\ell] \\
&(\boldsymbol{w}^\top \boldsymbol{x}_i - b_j)(1 - z_j) \geq 0 \quad j \in [\ell]\}
\end{aligned}
$$



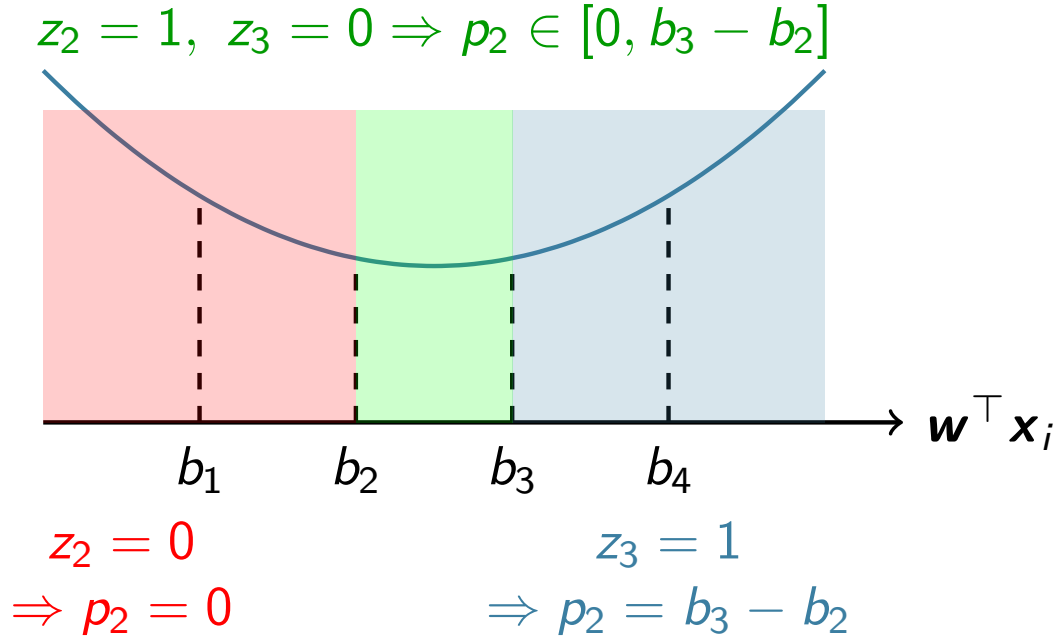
Extended set: $X_i = \{(\boldsymbol{w}, \boldsymbol{z}, s) \in \mathbb{R}^n \times \{0,1\}^\ell \times \mathbb{R} : \exists (p_0, \boldsymbol{p}) \in \mathbb{R}_+^{\ell+1} :$

$$
\boldsymbol{w}^\top \boldsymbol{x}_i = b_1 + \sum_{j=1}^{\ell} p_j - p_0
$$

$$
(b_{j+1} - b_j) z_{j+1} \leq p_j \leq (b_{j+1} - b_j) z_j,\ j \in [\ell - 1]
$$

X_i: Epigraph of loss and indicators of prediction i

$$
\begin{aligned}
X_i = \{&(\boldsymbol{w}, \boldsymbol{z}, s) \in \mathbb{R}^n \times \{0,1\}^\ell \times \mathbb{R} : \\
&\mathcal{L}_i(\boldsymbol{w}^\top \boldsymbol{x}_i) \leq s \\
&(\boldsymbol{w}^\top \boldsymbol{x}_i - b_j) z_j \geq 0 \qquad j \in [\ell] \\
&(\boldsymbol{w}^\top \boldsymbol{x}_i - b_j)(1 - z_j) \geq 0 \quad j \in [\ell]\}
\end{aligned}
$$

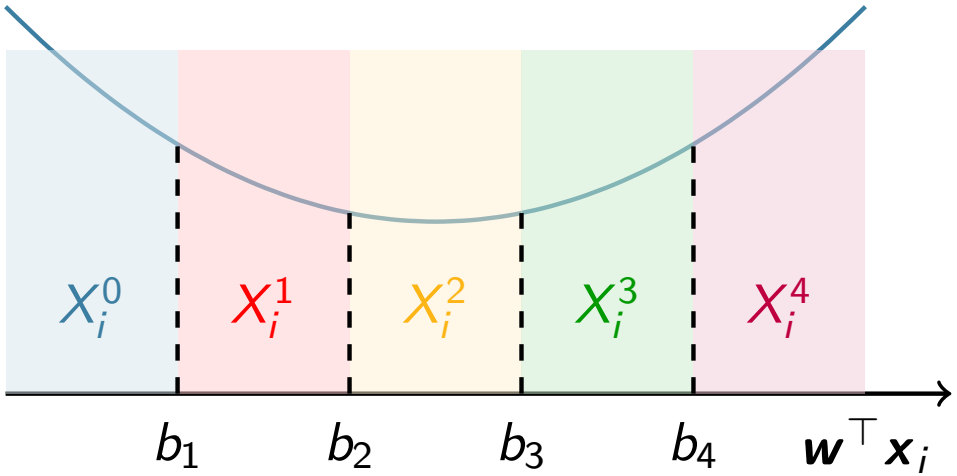


Extended set: $X_i = \{(\boldsymbol{w}, \boldsymbol{z}, s) \in \mathbb{R}^n \times \{0,1\}^\ell \times \mathbb{R} : \exists (p_0, \boldsymbol{p}) \in \mathbb{R}_+^{\ell+1} :$

$$
\begin{aligned}
&\boldsymbol{w}^\top \boldsymbol{x}_i = b_1 + \sum_{j=1}^{\ell} p_j - p_0 \\
&(b_{j+1} - b_j) z_{j+1} \leq p_j \leq (b_{j+1} - b_j) z_j,\ j \in [\ell - 1] \\
&p_0 z_1 = 0, p_\ell (1 - z_\ell) = 0 \\
&\mathcal{L}_i(\boldsymbol{w}^\top \boldsymbol{x}_i) \leq s\}
\end{aligned}
$$

cl conv(X_i)

Write as disjunction: $X_i = \bigcup_{j=0}^{\ell} X_i^j$



$X_i =$

$$\Bigg\{ (\boldsymbol{w}, \boldsymbol{z}, s) \in \mathbb{R}^n \times \{0,1\}^\ell \times \mathbb{R} : \exists (p_0, \boldsymbol{p}) \in \mathbb{R}_+^{\ell+1} :$$

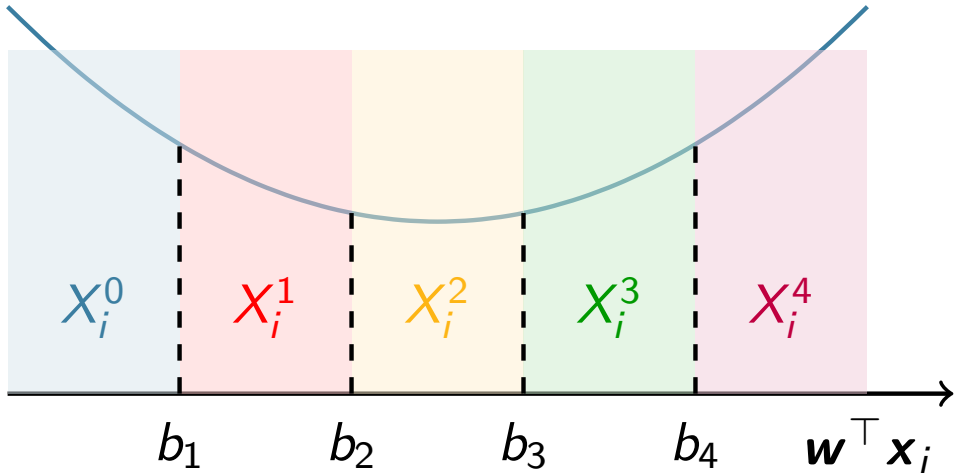
$$\boldsymbol{w}^\top \boldsymbol{x}_i = b_1 + \sum_{j=1}^{\ell} p_j - p_0$$

$$(b_{j+1} - b_j) z_{j+1} \leq p_j \leq (b_{j+1} - b_j) z_j, \; j \in [\ell - 1]$$

$$z_0 z_1 = 0 p_\ell (1 - z_\ell) = 0 \; \mathcal{L}_i(\boldsymbol{w}^\top \boldsymbol{x}_i) \leq s \Bigg\}$$

cl conv(X_i)

Write as disjunction: $X_i = \bigcup_{j=0}^{\ell} X_i^j$



cl conv$(X_i) =$

$$\Bigg\{ (\boldsymbol{w}, \boldsymbol{z}, s) \in \mathbb{R}^n \times [0,1]^\ell \times \mathbb{R} : \exists (p_0, \boldsymbol{p}) \in \mathbb{R}_+^{\ell+1} :$$

$$\boldsymbol{w}^\top \boldsymbol{x}_i = b_1 + \sum_{j=1}^{\ell} p_j - p_0$$

$$(b_{j+1} - b_j) z_{j+1} \le p_j \le (b_{j+1} - b_j) z_j, \; j \in [\ell - 1]$$

$$(1 - z_1)\mathcal{L}\left(b_1 - \frac{p_0}{1 - z_1}\right) + \sum_{j=1}^{\ell-1} (z_j - z_{j+1})\mathcal{L}\left(b_i + \frac{p_i - z_{i+1}(b_{i+1} - b_i)}{z_i - z_{i+1}}\right) + z_\ell \mathcal{L}\left(b_\ell + \frac{p_\ell}{z_\ell}\right) \le s \Bigg\}$$

Strong Reformulation

Applying the strengthening to the fair regression problem:

$$\min \sum_{i=1}^{m} s_i + \lambda \max_{j \in 1,\ldots,\ell} \left(\frac{1}{m} \sum_{i=1}^{m} z_{ij} - \frac{1}{m_1} \sum_{i=1:a_i=1}^{m} z_{ij} \right)$$

$$\text{s.t. } \boldsymbol{w}^\top \boldsymbol{x}_i = b_1 + \sum_{j=1}^{\ell} p_{ij} - p_{i0} \qquad i \in [m]$$

$$(\boldsymbol{w}, \boldsymbol{z}_i, s_i) \in X_i \qquad i \in [m]$$

$$\boldsymbol{z}_i \in \{0,1\}^\ell \qquad i \in [m]$$

Strong Reformulation

Applying the strengthening to the fair regression problem:

$$\min \sum_{i=1}^{m} s_i + \lambda \max_{j \in 1,\ldots,\ell} \left(\frac{1}{m} \sum_{i=1}^{m} z_{ij} - \frac{1}{m_1} \sum_{i=1:a_i=1}^{m} z_{ij} \right)$$

$$\text{s.t. } \boldsymbol{w}^\top \boldsymbol{x}_i = b_1 + \sum_{j=1}^{\ell} p_{ij} - p_{i0} \qquad i \in [m]$$

$$(\boldsymbol{w}, \boldsymbol{z}_i, s_i) \in \textbf{cl conv}(\mathbf{X_i}) \qquad i \in [m]$$

$$\boldsymbol{z}_i \in \{0,1\}^{\ell} \qquad i \in [m]$$

Computational Results on Synthetic Data

$n = 10$ features, $\ell = 40$, $m = \{15, 30, 50, 100\}$

λ	Big-M				Strong			
	Relax Gap	End Gap	Time	Nodes	Relax Gap	End Gap	Time	Nodes
0.01	45.3%	2.1%	487	4,010,293	1.9%	0.0%	252	223,581
0.02	60.8%	9.0%	1,654	19,895,563	5.4%	0.7%	908	1,079,684
0.04	73.2%	19.1%	2,040	20,405,258	8.3%	1.9%	950	902,741
0.05	76.6%	28.4%	2,681	29,063,625	11.3%	2.6%	1,212	1,058,541
0.06	78.8%	30.7%	2,759	28,508,133	13.1%	2.9%	1,507	1,447,340
0.08	81.5%	38.1%	3,023	32,654,651	14.0%	4.0%	1,673	1,284,442
0.10	83.4%	47.4%	3,196	35,020,673	16.0%	4.6%	1,803	1,275,695
0.20	87.7%	55.3%	3,147	28,630,313	22.9%	8.6%	1,625	683,419
0.30	89.5%	62.8%	3,295	29,468,759	31.2%	10.8%	1,555	507,315
0.50	91.7%	73.6%	3,600	31,892,472	45.5%	12.7%	1,611	437,835
Avg	**76.9%**	**36.6%**	**2588**	**25,954,974**	**16.9%**	**4.9%**	**1310**	**890,059**

Relaxation is solved in 0.25s on average.

Coordinate Descent: Optimizing a Single Coordinate

Goal: given $\boldsymbol{w}$, we want to improve upon by changing a single coordinate, w_k

$\rightarrow$ Minimize a univariate convex function $+$ linear combination of indicators

Coordinate Descent: Optimizing a Single Coordinate

Goal: given $\boldsymbol{w}$, we want to improve upon by changing a single coordinate, w_k
$\rightarrow$ Minimize a univariate convex function $+$ linear combination of indicators

---

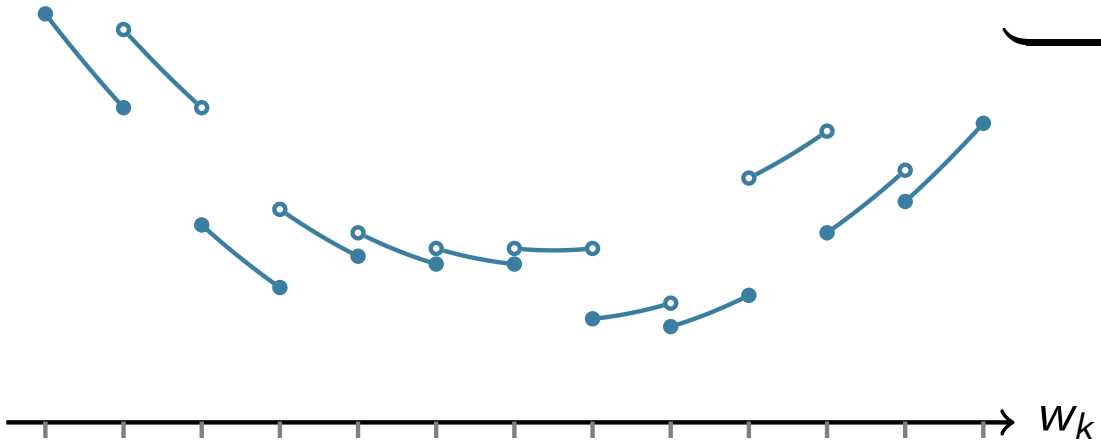
$$1(\boldsymbol{w}^\top \boldsymbol{x}_i > b_j) = 1(w_k x_{ik} + \textstyle\sum_{u \neq k} w_u x_{iu} > b_j) = 1(\text{sign}(x_{ik}) w_k > \underbrace{(b_j - \sum_{u \neq k} w_u x_{iu})/x_{ik}}_{\tilde{b}_{ij}})$$

Coordinate Descent: Optimizing a Single Coordinate

Goal: given $\boldsymbol{w}$, we want to improve upon by changing a single coordinate, w_k
$\rightarrow$ Minimize a univariate convex function $+$ linear combination of indicators

---

$$1(\boldsymbol{w}^\top \boldsymbol{x}_i > b_j) = 1(w_k x_{ik} + \textstyle\sum_{u \neq k} w_u x_{iu} > b_j) = 1(\text{sign}(x_{ik}) w_k > \underbrace{(b_j - \sum_{u \neq k} w_u x_{iu})/x_{ik}}_{\tilde{b}_{ij}})$$

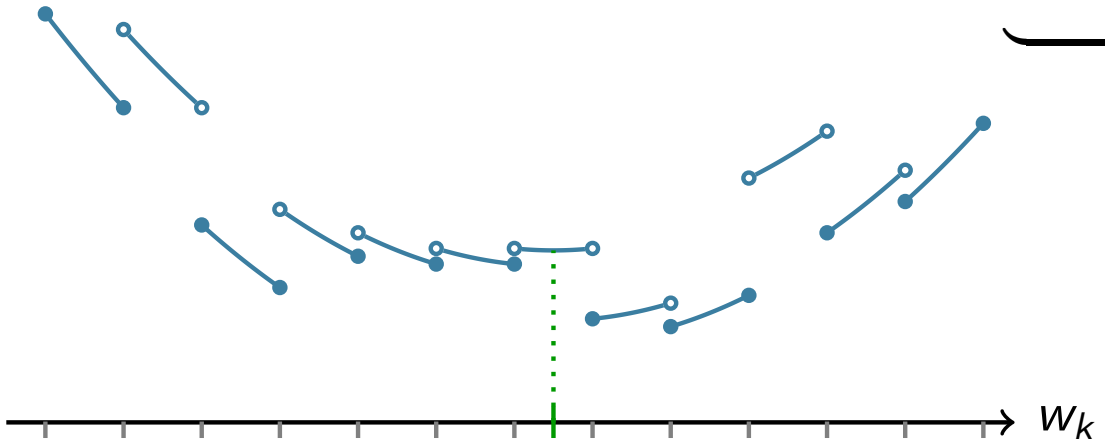


Coordinate Descent: Optimizing a Single Coordinate

Goal: given $\boldsymbol{w}$, we want to improve upon by changing a single coordinate, w_k
$\rightarrow$ Minimize a univariate convex function + linear combination of indicators

---

$$1(\boldsymbol{w}^\top \boldsymbol{x}_i > b_j) = 1(w_k x_{ik} + \textstyle\sum_{u \neq k} w_u x_{iu} > b_j) = 1(\text{sign}(x_{ik}) w_k > \underbrace{(b_j - \sum_{u \neq k} w_u x_{iu})/x_{ik}}_{\tilde{b}_{ij}})$$



Coordinate desecent update: at iteration t, choose coordinate w_k and

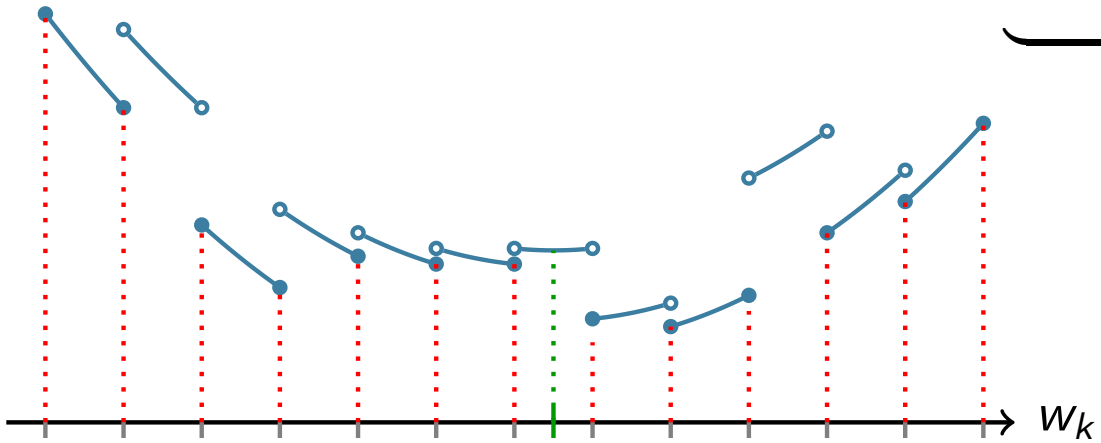
- evaluate objective at optimal solution ignoring fairness

Coordinate Descent: Optimizing a Single Coordinate

Goal: given $\boldsymbol{w}$, we want to improve upon by changing a single coordinate, w_k
$\rightarrow$ Minimize a univariate convex function + linear combination of indicators

---

$$1(\boldsymbol{w}^\top \boldsymbol{x}_i > b_j) = 1(w_k x_{ik} + \textstyle\sum_{u \neq k} w_u x_{iu} > b_j) = 1(\mathrm{sign}(x_{ik}) w_k > \underbrace{(b_j - \sum_{u \neq k} w_u x_{iu})/x_{ik}}_{\tilde{b}_{ij}})$$



Coordinate desecent update: at iteration t, choose coordinate w_k and

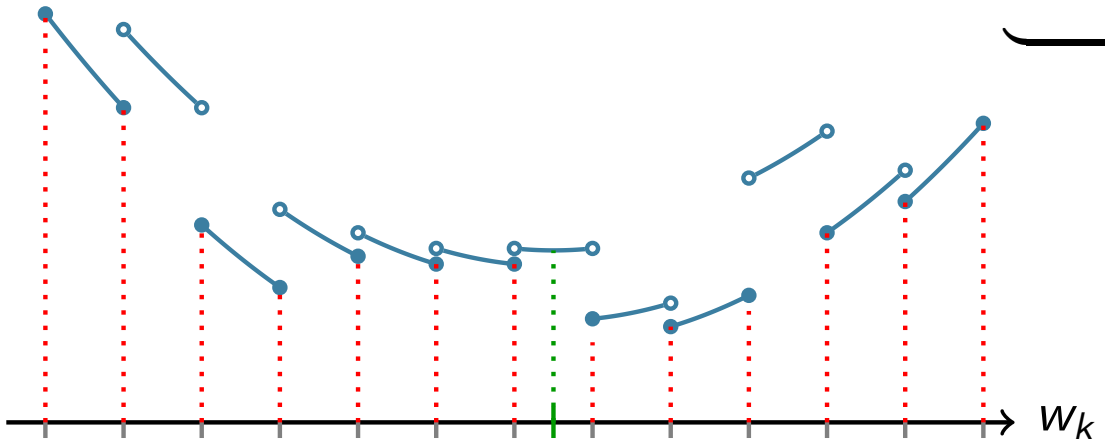
- evaluate objective at optimal solution ignoring fairness
- evaluate objective problem at $\tilde{b}_{ij}$, $\forall i \in [m],\ j \in [\ell]$

Coordinate Descent: Optimizing a Single Coordinate

Goal: given $\boldsymbol{w}$, we want to improve upon by changing a single coordinate, w_k
$\rightarrow$ Minimize a univariate convex function + linear combination of indicators

---

$$1(\boldsymbol{w}^\top \boldsymbol{x}_i > b_j) = 1(w_k x_{ik} + \textstyle\sum_{u \neq k} w_u x_{iu} > b_j) = 1(\text{sign}(x_{ik}) w_k > \underbrace{(b_j - \sum_{u \neq k} w_u x_{iu})/x_{ik}}_{\tilde{b}_{ij}})$$



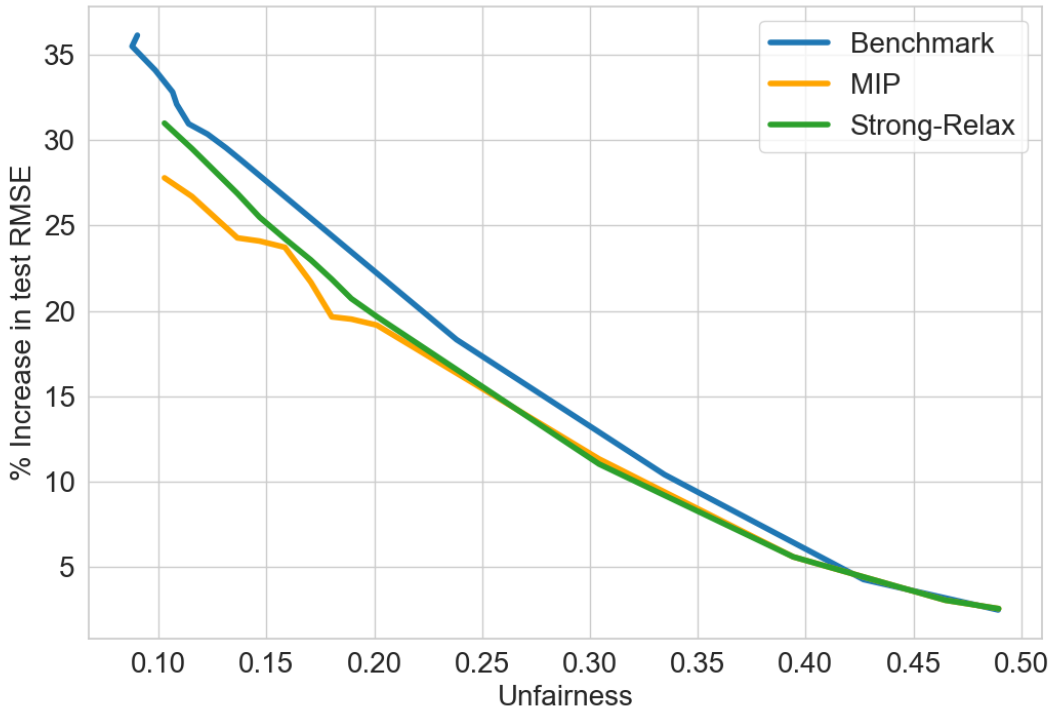
Coordinate desecent update: at iteration t, choose coordinate w_k and

- evaluate objective at optimal solution ignoring fairness
- evaluate objective problem at $\tilde{b}_{ij}$, $\forall i \in [m],\ j \in [\ell]$
- update to be the best out of $m \times \ell + 1$ options

Results on Real Data: Communities and Crime

Data: $n = 119,\ m = 1,994,\ \ell = 40$

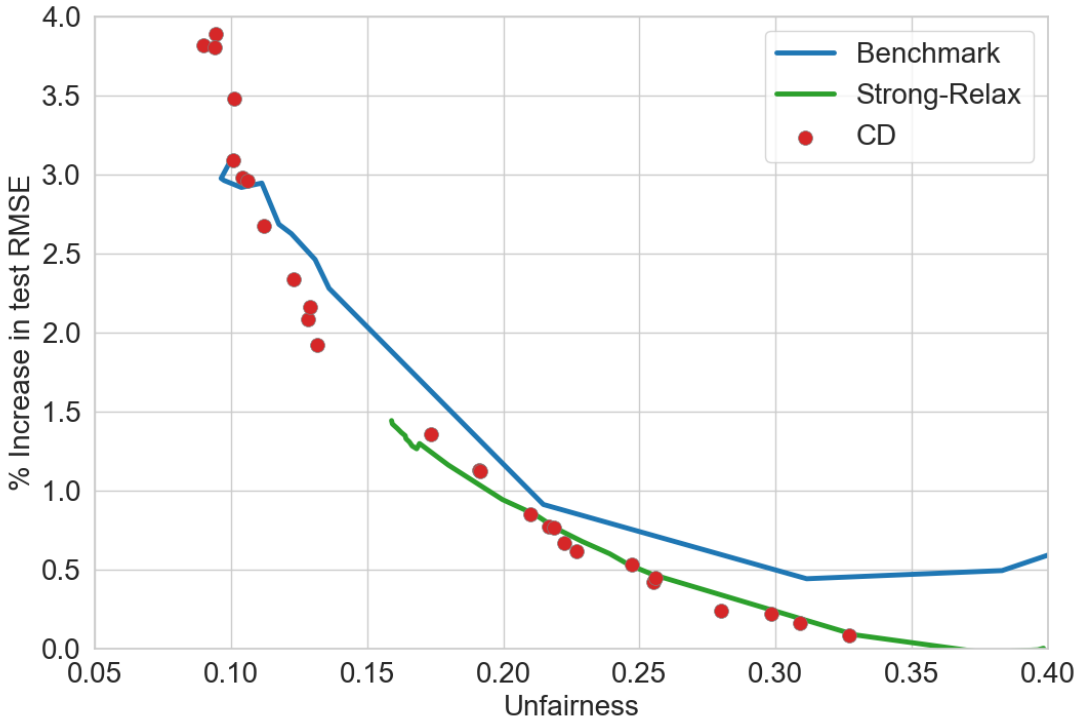
- Benchmark: Agarwal, reduction-based algorithm
- Compare model produced by the convex relaxation and best model found by MIP in three hours
- Plot fairness metric vs relative test RMSE increase compared to vanilla regression model
- Reduce relative RMSE increase by **2.5%**
- Average training time reduced from **200s** to **6s**



Results on Real Data: Law School

Data: $n = 12,\ m = 20,649,\ \ell = 40$

- Average training time reduced from **1445s** to **70s**
- Coordinate descent requires additional 2 seconds
- Coordinate descent is able to find fair solutions when relaxation fails



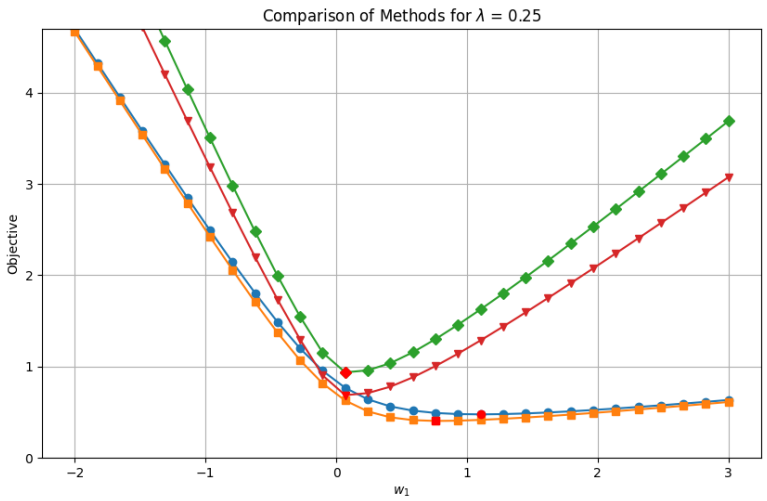
Conclusion

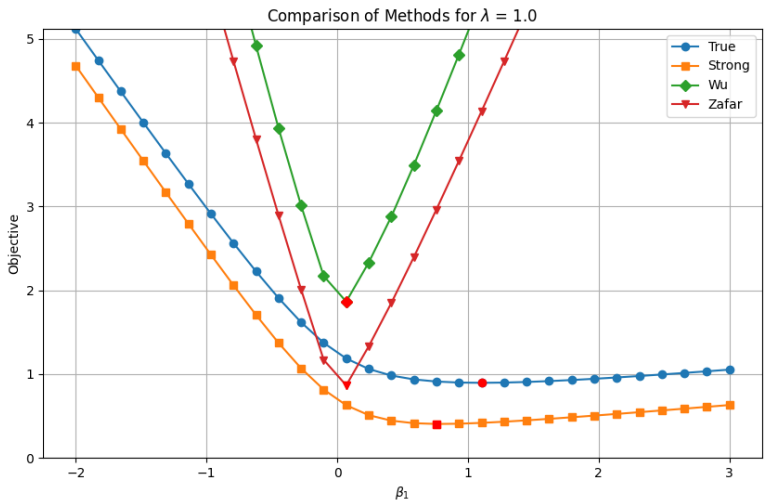
- **Versatile** framework: proposed methods can be adapted for generalized linear regression (ex: logistic regression) and for other popular fairness metrics
- **Key substructure:** piecewise convex function with indicators on intervals
- Relaxation + Coordinate Descent produce models **competitive** with state of the art at a fraction of the time ($\sim$ **30**$\times$ speed- up)

Our Convex Relaxation vs Convex Approximations

Existing approaches consider convex approximations of the fairness measure

We better capture the shape of true objective by considering the **joint structure of the loss function and fairness**



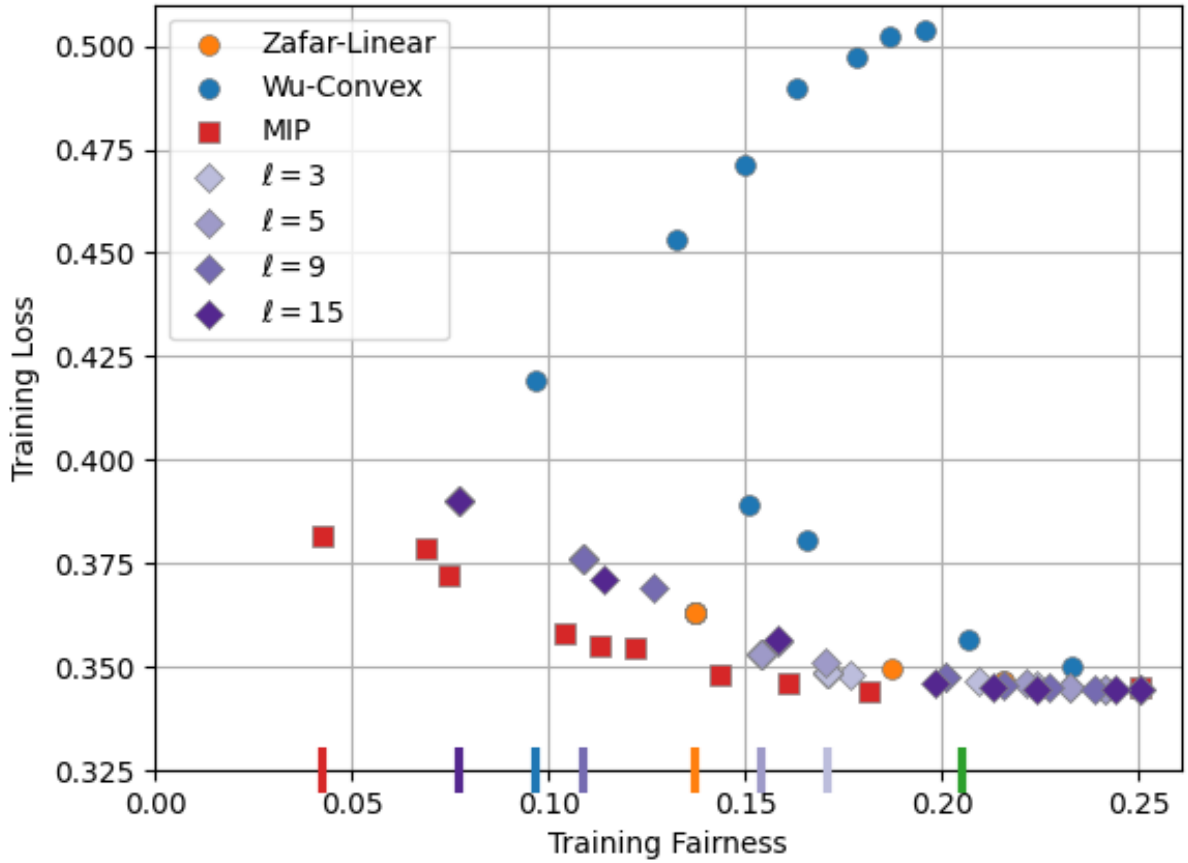


Synthetic Data: Optimality Gaps of Different Methods

λ	relax	CD-0	CD-unfair	CD-relax	MICQO
0.01	7.9	6.3	3.5	2.1	0.0
0.02	13.4	9.0	5.9	4.1	0.7
0.04	19.5	13.6	10.0	7.8	1.9
0.05	24.0	16.1	10.9	8.6	2.6
0.06	24.6	16.3	11.9	9.3	2.9
0.08	26.3	22.2	17.0	13.3	4.0
0.10	32.1	23.9	21.7	15.8	4.6
0.20	44.1	37.1	31.0	24.6	8.6
0.30	50.0	57.5	38.3	28.9	10.8
0.50	56.2	65.4	44.2	36.5	12.7
Avg	**29.8**	**26.7**	**19.4**	**15.1**	**4.9**

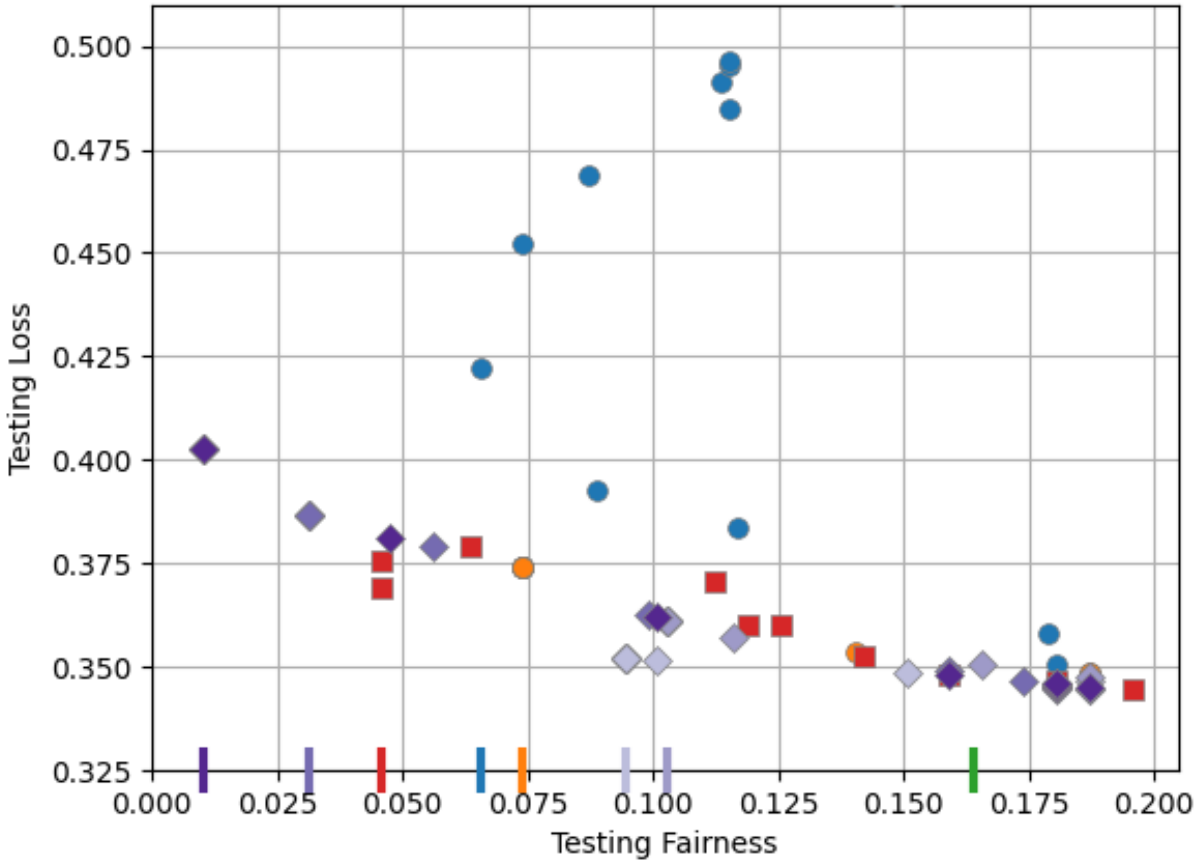
Overcoming 'too relaxed to be fair'

Idea: approximate fair classification $\ell = 1, b_1 = 0$ with a fair regression relaxation that artificially adds levels. This gives us a convex problem that can still produce fair models.



Overcoming 'too relaxed to be fair'

Idea: approximate fair classification $\ell = 1, b_1 = 0$ with a fair regression relaxation that artificially adds levels. This gives us a convex problem that can still produce fair models.



Discretized Fairness vs Exact Fairness Measures

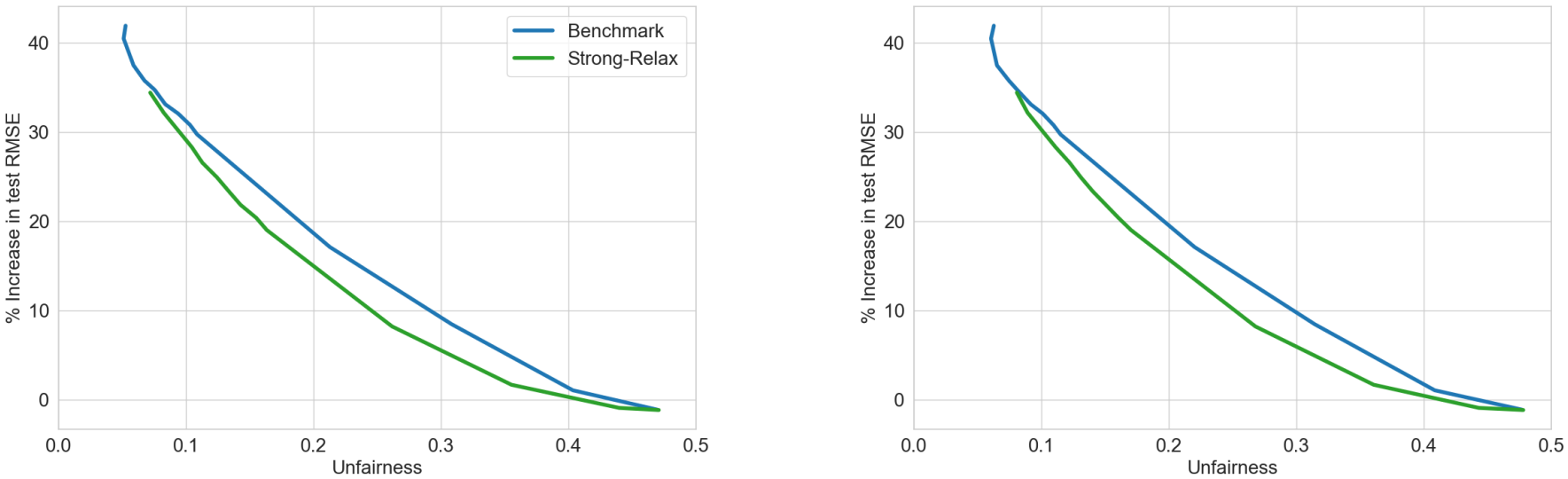


Figure: Training discretized unfairness (left) vs training exact unfairness (right).