# Machine Learning Fairness

**Group Member:**
**Gexin Chen**
**Noah Love**
**Rhea Sablani**
**Jiachen Liu**
**Xubo Wang**
**Yixuan Zhang**

# Evaluation Metrics

- **Accuracy** (Proportion of Correct Prediction)
- **Individual Fairness** (Similar Individuals Receive Similar Classification)
- **Group Fairness** (Similar or Equivalent Classification Distribution)
- **Calibration** (Difference of Accuracy Between Groups)

# Algorithms

1. Learning Fair Representations
2. Maximizing Accuracy under Fairness Constraints (C-SVM & C-LR)
3. Maximizing Fairness under Accuracy Constraints (gamma & fine-gamma)

# Learning Fair Representation
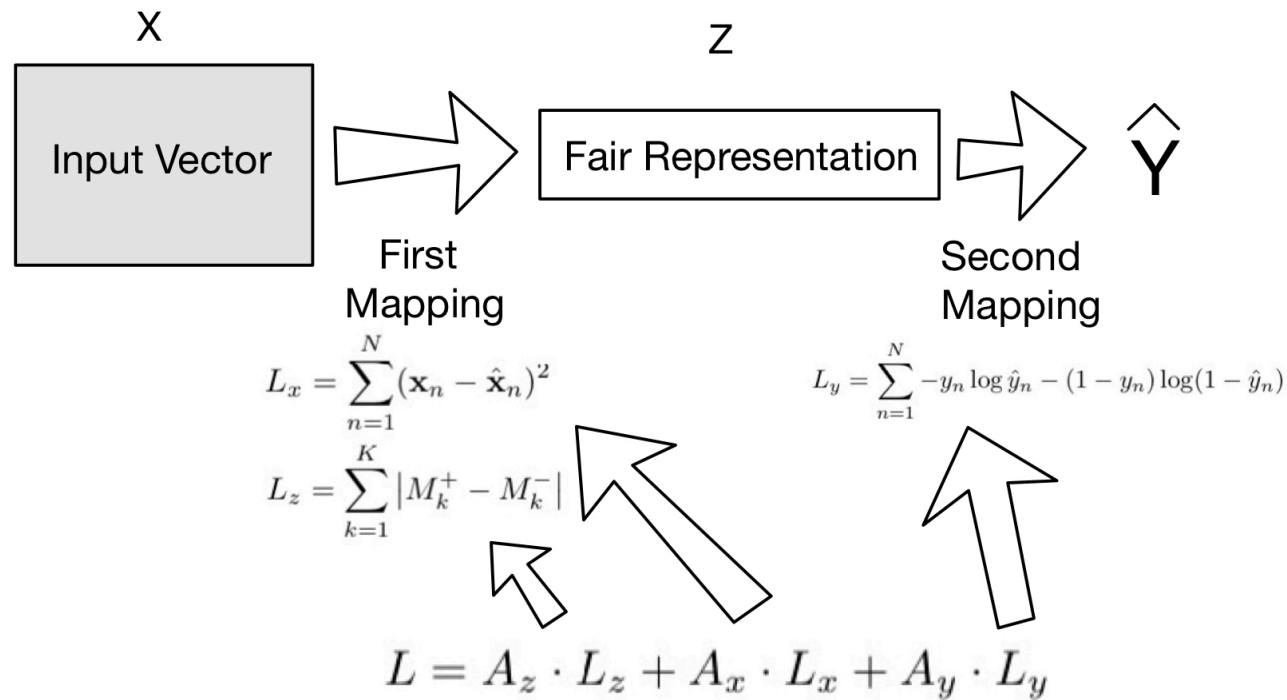


A here is a hyper-parameter controlling the trade-off between accuracy and fairness.

# Maximize Accuracy under Fairness Constraint

## C-SVM & C-LR

$$
\begin{aligned}
\text{minimize} \quad & L(\boldsymbol{\theta}) \\
\text{subject to} \quad & \frac{1}{N} \sum_{i=1}^{N} (\mathbf{z}_i - \bar{\mathbf{z}}) \, d_{\boldsymbol{\theta}}(\mathbf{x}_i) \leq \mathbf{c}, \\
& \frac{1}{N} \sum_{i=1}^{N} (\mathbf{z}_i - \bar{\mathbf{z}}) \, d_{\boldsymbol{\theta}}(\mathbf{x}_i) \geq -\mathbf{c},
\end{aligned}
$$

**Letter c here is a hyper-parameter controlling the trade-off between accuracy and fairness.**

# Maximize Fairness under Accuracy Constraint

## gamma & fine-gamma

$$\begin{aligned} \text{minimize} \quad & \left| \tfrac{1}{N} \sum_{i=1}^{N} \left( \mathbf{z}_i - \bar{\mathbf{z}} \right) d_{\boldsymbol{\theta}}(\mathbf{x}_i) \right| \\ \text{subject to} \quad & L(\boldsymbol{\theta}) \leq (1 + \gamma) L(\boldsymbol{\theta}^*), \end{aligned}$$

**Gamma here is a hyper-parameter controlling the trade-off between accuracy and fairness.**
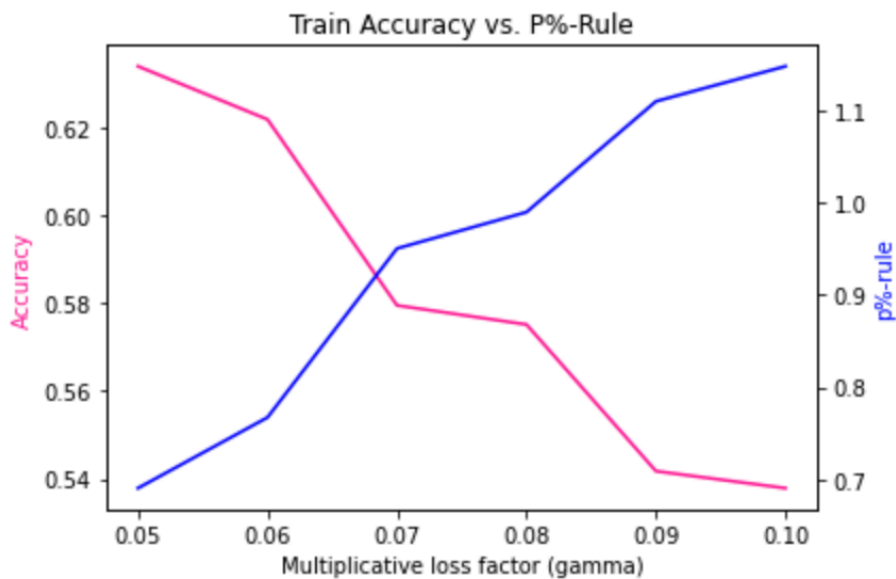
# Evaluation

| | A1 (LFR) | A2 (C-SVM & C-LR) | A3 (gamma & fine-gamma) |
|---|---|---|---|
| **Accuracy** | 53.66% | 66% | 56.39% |
| **Calibration** | 11.77% | 4.94% | 1.38% |
| **Protected Accuracy** | 60.69% | 63.96% | NA |
| **Non-protected Accuracy** | 48.91% | 68.91% | NA |
| **Individual Fairness** | 0.8118 | NA | NA |
| **p-rule** | NA | NA | 95% |

A1 Under A_x = 0.001, A_y = 0.1 A_z = 1000
A2 Under  c = 1
A3 Under gamma = 0.07

# Hyper-Tuning Visualization



Negative Correlation Between Fairness and Accuracy

# Reference

- Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rodriguez, and Krishna P. Gummadi. Fairness Constraints: Mechanisms for Fair Classification. https://arxiv.org/abs/1507.05259

- Rjcbard Zemel, Yu Wu, Kevin Swersky, and Toniann Pitassi. Learning Fair Representations. http://proceedings.mlr.press/v28/zemel13.pdf