Towards Equitable Machine Learning Driven Housing Price Prediction

Guillermo Arredondo; Álvaro López
Bachelor in Applied Mathematics and Data Science, ITAM, Mexico

Abdalwahab Almajed; Dr Maryam Tabar
UTSA Mentors

Abstract

- A well informed public policy is needed.
- Artificial Intelligence Models are popular in production, however there is an increased bias concern.
- Fair housing price prediction may be accomplished through bias mitigation.
- How effective is Adversarial Training mitigating race bias?

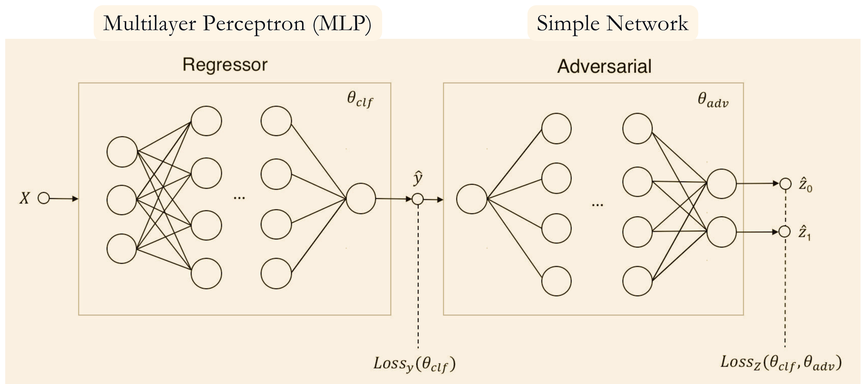
Background

- Housing plays an important role in human life, yet it has become barely accessible.
- People rely blindly on AI algorithms, forgetting they are flawed.
- Researchers have raised concerns towards race and ethnicity unfairness on AI models.
- An equitable price prediction could lead to a best informed urban development.
- There have been attempts on mitigating bias using Correlation Remover and Reduction-based Algorithms with good results.

Main Goal

- Provide an accurate prediction of house pricing in San Antonio.
- Improve the regressor model fairness on ethnicity and race criterion.

Methods



MLP price predictions → Adversarial training for bias mitigation → Training update and fairness comparison

Adversarial Training Objective Function

$$min_{\theta_{reg}} \left[Loss_y(\theta_{reg}) - \alpha Loss_z(\theta_{reg}, \theta_{adv}) \right]$$

Dataset
232,057 houses across San Antonio from publicly-available sources.

Protected attribute
Two race groups are considered (white vs non-white).

Results

Model	MLP	MLP+AT (α=3)
Independence	1.326	1.188
Separation	1.274	1.208
Sufficiency	1.018	1.15

Table 1. Fairness criterion values for original model and adversarial trained model

Independence: Equality of outcomes/selection
Separation: Equality of Errors (Equality of outcomes given a threshold)
Sufficiency: Choices reflect same accuracy per group (calibration)

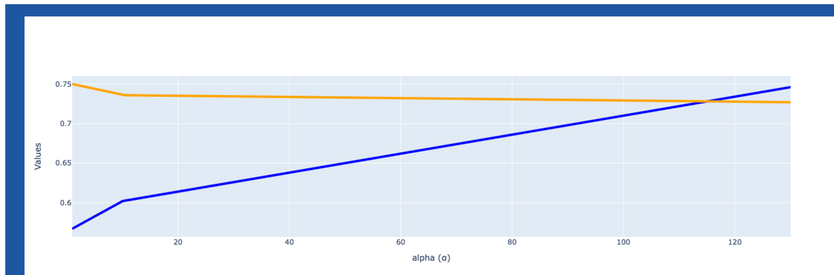


Fig 1. Fairness (blue) and performance (orange) through different α values.

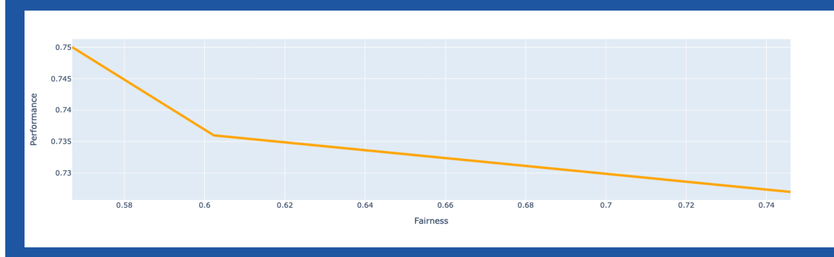


Fig 2. Fairness (x axis) against performance (y axis)

- The initial model presents significant bias on race.
- There is some fairness improvement after Adversarial Training.
- When increasing the adversarial loss weight (α), fairness improves but accuracy of the model is sacrificed.

Conclusions

- Adversarial training proves to be an alternative for bias mitigation in ML models.
- However, fairness improvement is not significant and requires to sacrifice the model performance.
- For future work it would be recommended to modify the adversarial model and verify the impact on bias mitigation.

References

Almajed, A., Tabar, M. & Najafirad, P. (2024). *Modeling Diversity in Urban Growth: A Machine Learning Case Study on Housing Price Prediction in America's Expanding Metropolises* [Unpublished manuscript]. School of Data Science, The University of Texas at San Antonio.

Tonk, S., & Tonk, S. (2024). *Towards fairness in ML with adversarial networks.* Xebia. https://xebia.com/blog/towards-fairness-in-ml-with-adversarial-networks/