Introduction to ai360 (hands-on)

The AI Fairness 360 toolkit is an extensible open-source library containing techniques developed by the research community to help detect and mitigate bias in machine learning models throughout the AI application lifecycle. The AI Fairness 360 Python package includes metrics to test for biases and algorithms to mitigate bias.

A bias detection and/or mitigation tool needs to be tailored to the particular bias of interest. More specifically, it needs to know the attribute or attributes, called protected attributes, that are of interest: race is one example of a protected attribute and age is a second.

Tutorial

This tutorial shows how to compute fairness metrics and mitigate the bias using a feature transformation. The specific mitigation technique used in this example is called a re-weighting algorithm [1]. Get started by downloading the .ipynb file and dataset locally in the following ZIP file.

AIF360-tutorial.ipynb.zip

This tutorial builds on the credit decisions demo. For a more advanced tutorial see the medical expenditure demo.

Additional Resources

- AIF360 source
- AIF360 examples

---

[1]: Faisal Kamiran and Toon Calders. Data preprocessing techniques for classification without discrimination. *Knowl. Inf. Syst.*, 33(1):1–33, October 2012. URL: https://doi.org/10.1007/s10115-011-0463-8, doi:10.1007/s10115-011-0463-8.