

Fairness in Machine Learning

Naman Goel
naman.goel@epfl.ch
<http://lia.epfl.ch/~goel/>

Recently, it has been observed that AI algorithms can learn to discriminate between people based on sensitive attributes such as gender and race. This is obviously of concern to anyone who cares about social justice. But even beyond that, this also poses a significant challenge to the deployment of AI algorithms in practice because most countries forbid discrimination based on gender and race in decision making.

In this project, you will read the literature in this area and implement new solutions for enforcing fairness in classification.

Requirement: You must be a masters student, highly proficient in Python and familiar with popular ML/data science libraries.

If you want to work on this project, please write a short email to me - attach your grade statement (undergrad and masters) and a brief CV. We can then schedule a meeting to discuss more.

References:

1. Non-Discriminatory Machine Learning through Convex Fairness Criteria. Naman Goel and Mohammad Yaghini and Boi Faltings, AAAI 2018
https://lia.epfl.ch/~goel/upload/doc/papers/2018/goel_aaai18.pdf
2. Crowdsourcing with Fairness, Diversity and Budget Constraints. Naman Goel and Boi Faltings, AAAI/ACM AIES 2019.
https://lia.epfl.ch/~goel/upload/doc/papers/2019/goel_aies19.pdf