

A clear description of the problem and well-defined setting is here

<https://compass.onlinelibrary.wiley.com/doi/10.1111/phc3.12891>

We can define noise term as the difference between decisions and outcomes in the given setting. The noise is dependent on sensitive attributes and we are going to remove it.

Some free datasets with bias are collected for example here

<https://arxiv.org/abs/1802.04422>

The aim reminds denoising auto-encoders but with noisy targets (outcomes) instead of inputs

<https://dl.acm.org/doi/abs/10.1145/1390156.1390294>

Except denoising that is proposed in this project, other ways of preprocessing of biased data are relabeling and reweighing

<https://ieeexplore.ieee.org/document/5360534>

The preprocessing techniques are also discusses in

<https://link.springer.com/article/10.1007/s10115-012-0584-8>

The paper also provides a simple model, where discrimination impact is constant.

Individual component analysis (ICA) also can be useful.

<https://www.sciencedirect.com/science/article/pii/S0893608000000265>

Another measures of independence (in terms of ICA) that can be maximized are in

<https://www.sciencedirect.com/science/article/pii/S0165188923000362>