Fair-MAML offers a route to distribute fair ML models that can quickly be trained on new tasks.

Fair Meta-Learning: Learning How to Learn Fairly

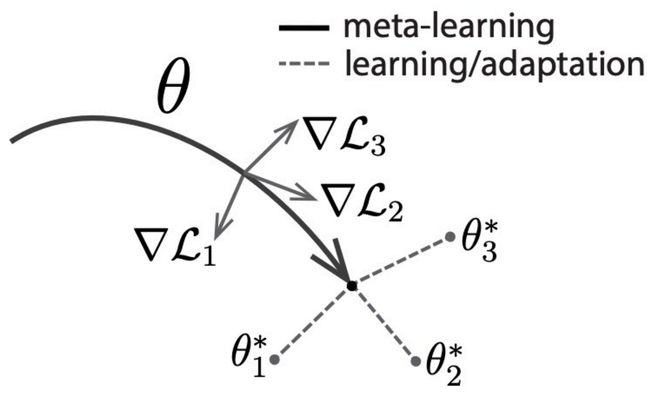
Dylan Slack, Sorelle Friedler, and Emile Givental

MOTIVATION

- Minor changes in test distribution can have significant effects on fairness (see Fairness Warnings). How can we train a model that copes?

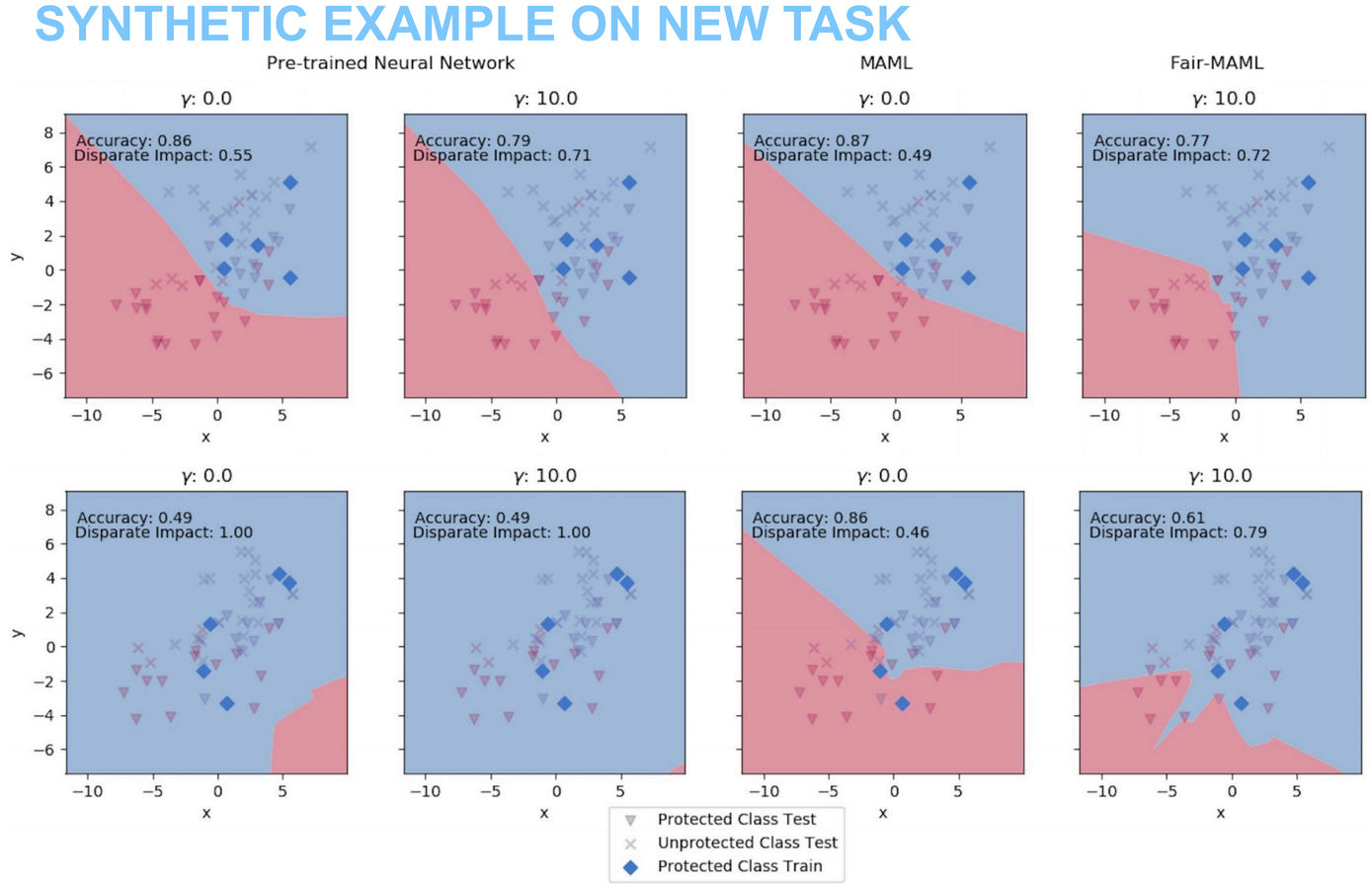
METHODS

- We can train *fair meta-model* that contains general features relating to both fairness and accuracy using *model agnostic meta-learning* with added fairness objective (Fair-MAML).
- Fair-MAML can be fine-tuned to new fairness tests to achieve high degrees of accuracy with minimal data.

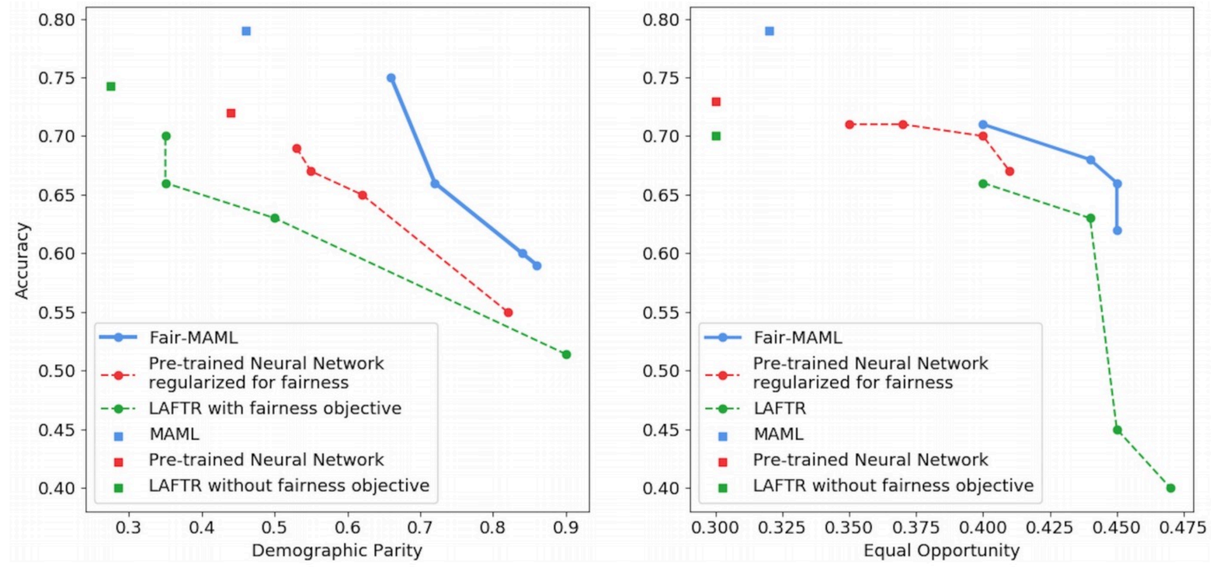


[Finn et. al., 2017]

SYNTHETIC EXAMPLE ON NEW TASK



COMMUNITIES AND CRIME EXAMPLE



HAVERFORD COLLEGE

UCI University of California, Irvine

Check us out at FAT* 2020:

Fairness Warnings & Fair-MAML: Learning Fairly from Minimal Data