Active Sampling for Min-Max Fairness

Jacob Abernethy [1] Pranjal Awasthi [2] Matthäus Kleindessner [3]
Jamie Morgenstern [4] Chris Russell [3] Jie Zhang [4]

[1] Georgia Tech, USA

[2]Google, USA

[3]Amazon Web Services, Germany

[4]University of Washington, USA

Motivation for Min-Max Fairness and Related Work

- "Level-down", "Unnecessary harm": (Ustun et al., 2019)
  Unnecessarily reduce performance on better-off groups and overall performance

- "Level-up" - Min-max Fairness
  Only degrade performance of a group if it improves on the worst-off group

  - Martinez et al., 2020
    Min-max Pareto optimal classifier, no convergence guarantee

  - Diana et al, 2021
    Re-solve on all data in each iteration

  - **Advantage of Our Work**
    Simplicity, $1/\sqrt{T}$ and $1/T$ convergence guarantee

Min-Max Fair Model

Define: ℓ is a loss function; f_θ is a model; $\theta \in \Theta \subset \mathbb{R}^d$;
g is a set of demographic groups; D_i is a group specific distribution;
$v(\theta; D) := \mathbb{E}_{z \sim D}\, \ell(f_\theta, z)$.

Goal: learn $\theta^\star$ that satisfies:

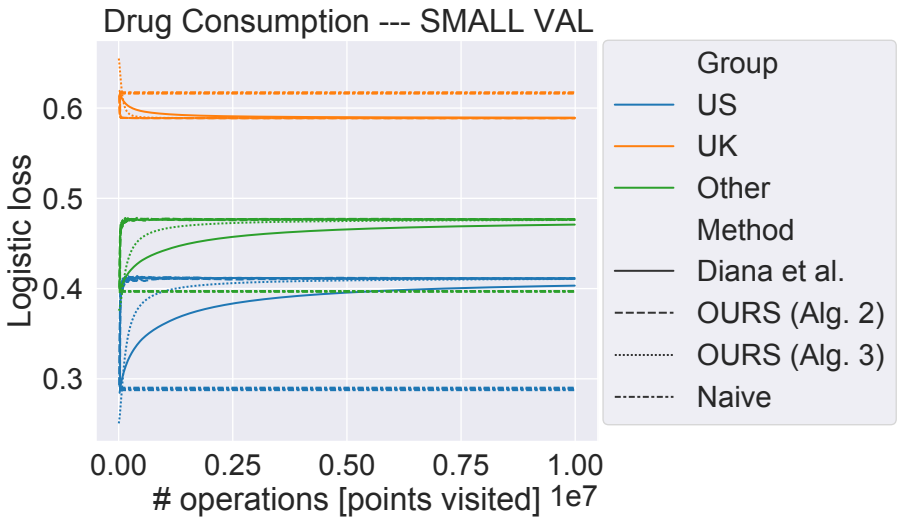
$$\max_{i \in [g]} v\left(\theta^\star; D_i\right) = \inf_{\theta \in \Theta} \max_{i \in [g]} v\left(\theta; D_i\right).$$

Our Algorithm (convergence rate $1/\sqrt{T}$)

Algorithm 1 Min-max Stochastic Gradient Descent

Init: $\theta_1 \in \Theta$ arbitrary
for $t = 1 \dots T-1$ **do**
  compute $i_t = \underset{i \in [g]}{\text{argmax}}\, v\left(\theta_t; \hat{D}_i\right)$
  sample $z_t \sim D_{i_t}$
  compute $\nabla_t \leftarrow \nabla_\theta \ell(f_{\theta_t}; z_t)$
  update $\theta_{t+1} \leftarrow \text{PROJ}_\Theta(\theta_t - \eta \nabla_t)$
end for
return $\bar{\theta}_T = \frac{\sum_{t=1}^{T} \theta_t}{T}$

Experiments - Drug Consumption Data



Experiments - Drug Consumption Data

	US	UK	Other	Overall	
Diana et al. (2021)	0.4035	**0.5894**	0.4710	0.5166	Log loss
Martinez et al. (2020)	0.4122	**0.5889**	0.4771	0.5198	
OURS (Alg. 2 with TRAIN=VAL; Avg. over 5 runs)	0.4114	**0.5889**	0.4766	0.5195	
OURS (Alg. 2 with SMALL VAL; Avg. over 5 runs)	0.4112	**0.5889**	0.4766	0.5195	
OURS (Algorithm 3)	0.4108	**0.5889**	0.4761	0.5193	
Diana et al. (2021)	0.1566	**0.2964**	0.2173	0.2432	01 loss
Martinez et al. (2020)	0.1598	**0.2950**	0.2183	0.2435	
OURS (Alg. 2 with TRAIN=VAL; Avg. over 5 runs)	0.1634	**0.2971**	0.2218	0.2463	
OURS (Alg. 2 with SMALL VAL; Avg. over 5 runs)	0.1601	**0.2960**	0.2211	0.2446	
OURS (Algorithm 3)	0.1598	**0.2960**	0.2183	0.2440	

Algorithm 2. (Average) Per-group losses and errors as well as overall losses and errors from the final iteration are shown in the table. For every method, the maximum loss / error among the groups is shown in bold.