Fair Generalized Linear Models with a Convex Penalty

Hyungrok Do

Joint work with Preston Putzel, Axel Martin, Padhraic Smyth, Judy Zhong

2022 ICML

https://github.com/hyungrok-do/fair-glm-cvx

Background

- Machine learning is now widely used for supporting crucial decision making.
- However, it has been demonstrated that data-driven models can often retain biases in the underlying data which may result in inequalities of predictions.
- It has motivated many researchers to devote themselves to algorithmic fairness.
- Most of the existing work has focused on binary classification and regression tasks.

Overview

- Developing the first in-process fairness framework for generalized linear models (GLMs) which can handle various types of response variables, including less-studied multinomial and count.
- Providing its theoretical properties and optimization guarantees.
- Demonstrating that the proposed fair GLM can improve prediction parities for variety of outcomes, including multinomial and count outcomes.

Generalized Linear Models (GLMs)

Definition: Generalized Linear Model

$$\mathbb{E}[Y|\mathbf{X}] = g^{-1}(\mathbf{X}\boldsymbol{\beta}), \tag{1}$$

- Y: response variable, $\mathbf{X}$: predictor variables, $\boldsymbol{\beta}$: regression coefficient
- g: link function - different link function leads to different types of response variables

Table: Representative Examples of Link Functions

Distribution	Link	Support	$\mu = g^{-1}(\mathbf{X}\boldsymbol{\beta})$	Task
Bernoulli(μ)	Logit	$\{0, 1\}$	$\frac{1}{1+\exp(-\mathbf{X}\boldsymbol{\beta})}$	Binary Classification
Categorical($\boldsymbol{\mu}$)	Logit	$\{0, 1\}$	$\frac{\exp(\mathbf{X}\boldsymbol{\beta}_i)}{1+\sum_{j \neq i} \exp(\mathbf{X}\boldsymbol{\beta}_j)}$	Multiclass Classification
Normal(μ, σ^2)	Identity	$\mathbb{R}$	$\mathbf{X}\boldsymbol{\beta}$	Regression
Poisson(μ)	Log	$\{0\} \cup \mathbb{Z}_+$	$\exp(\mathbf{X}\boldsymbol{\beta})$	Count Regression

Fair Generalized Linear Model

We propose a *fair generalized linear model* which is defined:

$$\widehat{\boldsymbol{\beta}}_{\mathsf{FGLM}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \; -\mathbb{E}[\ell(\boldsymbol{\beta}; \mathbf{X}, Y)] + \lambda \mathcal{D}_{\mathsf{LC}}(\boldsymbol{\beta}), \tag{2}$$

Fair Generalized Linear Model

We propose a *fair generalized linear model* which is defined:

$$\widehat{\boldsymbol{\beta}}_{\mathsf{FGLM}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \; -\mathbb{E}[\ell(\boldsymbol{\beta}; \mathbf{X}, Y)] + \lambda \mathcal{D}_{\mathsf{LC}}(\boldsymbol{\beta}), \tag{2}$$

where

- ℓ: log-likelihood
- $\mathcal{D}_{\mathsf{LC}}$: fairness encouraging penalty which is convex in $\boldsymbol{\beta}$

Fair Generalized Linear Model

We propose a *fair generalized linear model* which is defined:

$$\widehat{\boldsymbol{\beta}}_{\mathsf{FGLM}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \; -\mathbb{E}[\ell(\boldsymbol{\beta}; \mathbf{X}, Y)] + \lambda \mathcal{D}_{\mathsf{LC}}(\boldsymbol{\beta}), \tag{2}$$

where

- ℓ: log-likelihood
- $\mathcal{D}_{\mathsf{LC}}$: fairness encouraging penalty which is convex in $\boldsymbol{\beta}$
- The penalty term encourages the GLM to generate similar predicted values for the same true y across all the groups.

Fair Generalized Linear Model

We propose a *fair generalized linear model* which is defined:

$$\widehat{\boldsymbol{\beta}}_{\mathsf{FGLM}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \; -\mathbb{E}[\ell(\boldsymbol{\beta}; \mathbf{X}, Y)] + \lambda \mathcal{D}_{\mathsf{LC}}(\boldsymbol{\beta}), \tag{2}$$

where

- ℓ: log-likelihood
- $\mathcal{D}_{\mathsf{LC}}$: fairness encouraging penalty which is convex in $\boldsymbol{\beta}$
- The penalty term encourages the GLM to generate similar predicted values for the same true y across all the groups.
- Moreover, it also encourages the GLM to have similar log-likelihoods for the same true y across all the groups.

Fair Generalized Linear Model

- Our penalty term has the following form

$$\mathcal{D}_{\mathsf{LC}}(\boldsymbol{\beta}) = \sum_{k,l \in \mathcal{A}} \sum_{y \in \mathcal{Y}} \mathbb{E}[(\mathbf{X}^{ky}\boldsymbol{\beta} - \mathbf{X}^{ly}\boldsymbol{\beta})^2] \tag{3}$$

Fair Generalized Linear Model

- Our penalty term has the following form

$$\mathcal{D}_{\mathsf{LC}}(\boldsymbol{\beta}) = \sum_{k,l \in \mathcal{A}} \sum_{y \in \mathcal{Y}} \mathbb{E}[(\mathbf{X}^{ky}\boldsymbol{\beta} - \mathbf{X}^{ly}\boldsymbol{\beta})^2] \tag{3}$$

- We have

$$\underbrace{\sum_{k,l \in \mathcal{A}} \sum_{y \in \mathcal{Y}} \Big(\mathbb{E}[\mu(\mathbf{X}^{ky}\widehat{\boldsymbol{\beta}}_{\mathsf{FGLM}})] - \mathbb{E}[\mu(\mathbf{X}^{ly}\widehat{\boldsymbol{\beta}}_{\mathsf{FGLM}})] \Big)^2}_{\text{Sum of squared differences of expected predictions}} \leq C_\mu \mathcal{D}_{\mathsf{LC}}(\widehat{\boldsymbol{\beta}}_{\mathsf{FGLM}}), \tag{4}$$

Fair Generalized Linear Model

- Our penalty term has the following form

$$\mathcal{D}_{\mathsf{LC}}(\boldsymbol{\beta}) = \sum_{k,l \in \mathcal{A}} \sum_{y \in \mathcal{Y}} \mathbb{E}[(\mathbf{X}^{ky}\boldsymbol{\beta} - \mathbf{X}^{ly}\boldsymbol{\beta})^2] \tag{3}$$

- We have

$$\underbrace{\sum_{k,l \in \mathcal{A}} \sum_{y \in \mathcal{Y}} \left(\mathbb{E}[\mu(\mathbf{X}^{ky}\widehat{\boldsymbol{\beta}}_{\mathsf{FGLM}})] - \mathbb{E}[\mu(\mathbf{X}^{ly}\widehat{\boldsymbol{\beta}}_{\mathsf{FGLM}})] \right)^2}_{\text{Sum of squared differences of expected predictions}} \leq C_\mu \mathcal{D}_{\mathsf{LC}}(\widehat{\boldsymbol{\beta}}_{\mathsf{FGLM}}), \tag{4}$$

$$\underbrace{\sum_{k,l \in \mathcal{A}} \sum_{y \in \mathcal{Y}} \left(\mathbb{E}[\ell(\widehat{\boldsymbol{\beta}}_{\mathsf{FGLM}}; \mathbf{X}^{ky}, y)] - \mathbb{E}[\ell(\widehat{\boldsymbol{\beta}}_{\mathsf{FGLM}}; \mathbf{X}^{ly}, y)] \right)^2}_{\text{Sum of squared differences of expected log-likelihoods}} \leq C_\ell \mathcal{D}_{\mathsf{LC}}(\widehat{\boldsymbol{\beta}}_{\mathsf{FGLM}}). \tag{5}$$

- These inequalities support that our penalty term encourages the fairness.

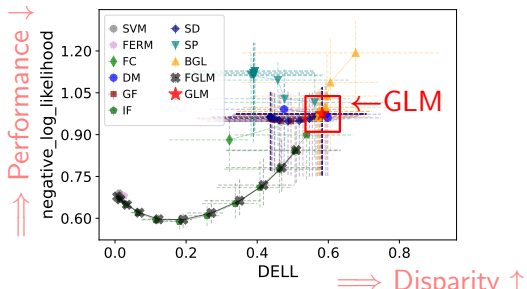
Experiments

Benchmark Datasets (11 Datasets, 12 Experiments)

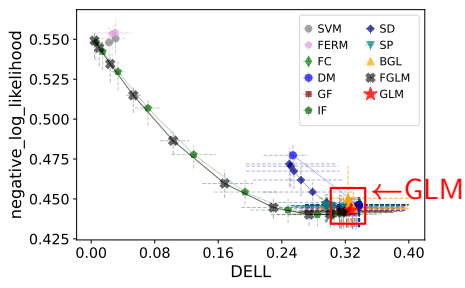
Outcome Type	Dataset	Group (K)
Binary	Adult	Gender (2)
	Arrhythmia	Gender (2)
	COMPAS	Race (4)
	Drug	Race (2)
	German	Gender (2)
Continuous	Communities	Race (3)
	Law School	Race (5)
	Parkinsons Tel.	Gender (2)
	Student	Gender (2)
Count	HRS	Race (4)
Multiclass	Drug	Race (2)
	Obesity	Gender (2)

Results

- Binary Classification



(a) Arrhythmia



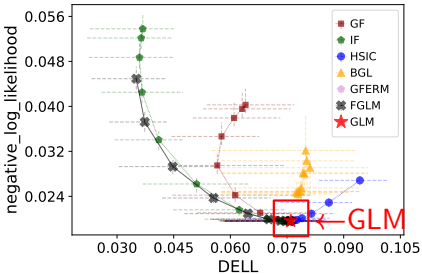
(b) Drug Consumption (Binarized)

* Repeated 20 times, 70% training / 30% testing

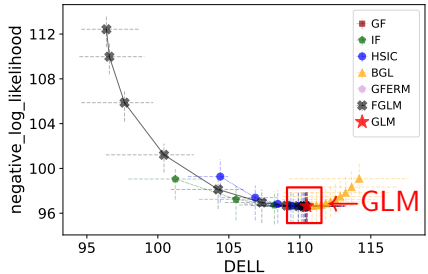
* Markers for the average across the testing set / dotted lines for deviations (IQRs) across the testing set

Results

- Regression



(c) Communities and Crimes (Binarized)



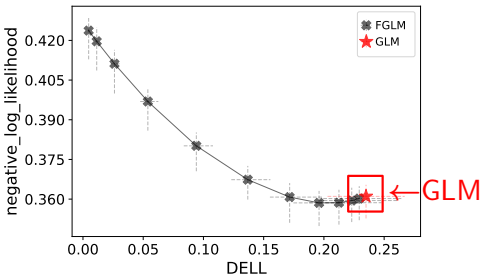
(d) Parkinson's Telemonitoring

* Repeated 20 times, 70% training / 30% testing

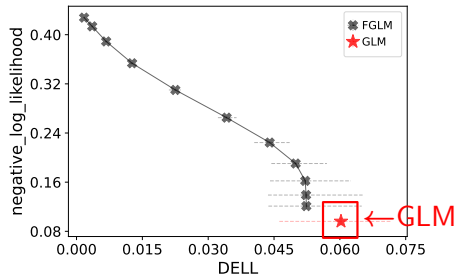
* Markers for the average across the testing set / dotted lines for deviations (IQRs) across the testing set

Results

- Mutliclass Classification



(e) Drug Consumption



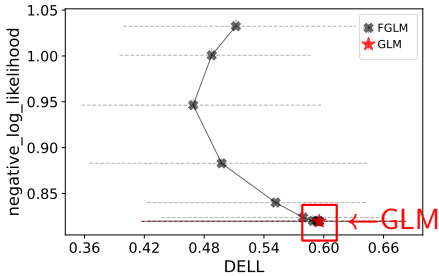
(f) Obesity

* Repeated 20 times, 70% training / 30% testing

* Markers for the average across the testing set / dotted lines for deviations (IQRs) across the testing set

Results

- Count Regression



(g) Health & Retirement Study

* Repeated 20 times, 70% training / 30% testing

* Markers for the average across the testing set / dotted lines for deviations (IQRs) across the testing set

Summary

- To the best of our knowledge, this is the first study on the algorithmic fairness of GLMs.
- We have developed a new convex penalty function that encourages fairness in terms of *equalized expected outcomes* and *equalized expected log-likelihoods*.
- We proved the F-GLM estimator is $\sqrt{n}-$consistent.
- Through an extensive experiment on benchmark datasets, we have demonstrated that our method can provide decent performance-disparity trade-offs for various types of outcomes, including multinomial and count outcomes.
- Code for reproducing the results is available at `https://github.com/hyungrok-do/fair-glm-cvx`.

Thank You!