Deadline 31 Mar 2025 23:59 (CPH time).
This is a group submission (2-3 person group).

What to hand-in: A report as pdf summarizing the main findings, max. 3 pages, including plots. A jupyter notebook detailing the process. Upload the two files as a single zip file on learnIT.

Where to start: This assignment consists of one tasks. You can find a Jupyter template to get started in the assignment on learnIT.

Dataset: US Census data from https://github.com/zykls/folktables. We use data of individuals from the state California in 2018, as detailed in the template. The template also details which attributes we use as feature vector. More details on the dataset can be found in the accompanying paper at https://arxiv.org/pdf/2108.04884.pdf.

Using the Folktales dataset (se Jupyter notebooks on Github) use the two techniques presented in lecture 7 to create fairer representations of the dataset. For this assignment you will work with 2 protected attributes: gender and race (in the dataset they are called `SEX` and `RAC1P`).

The Task: Train multiple binary classifiers to predict income (label = 1 or True if income>$25k, otherwise label = 0 or False).

1. Train one logistic regression model without any fairness constraing on the Folktales dataset and calculate its general accuracy, and respectively the accuracies for men and women (feature `SEX`) and for different races (feature `RAC1P`). Remember to evaluate the model using cross validation.

2. Use the group fairness constrain from Berk et al. (see lecture 6) to ensure a fair model is trained. Here you have to train two 'fair'-logistic regression models. models (see more below). Subtasks:

    - Build one logistic regression model with where `SEX` splits the data into two groups. Calculate: a) the model's overall accuracy, b) separate accuracies for each of the different values of `SEX`. Plot the results and discuss what they mean. Is the model fair(er)?

    - Build one logistic regression model with where `RAC1P` splits the data into two groups. Calculate: a) the model's overall accuracy, b) separate accuracies for each of the different values of `RAC1P`. Plot

the results and discuss what they mean. Is the model fair(er)?

Please submit your report (in PDF format) & jupyter notebook on learnIT.

Checklist

To avoid surprises, please make sure that your hand-in covers the following parts:

Overall

- □ Assignment: Concise discussion of the results
- □ Plots: Labels and titles clear and readable?
- □ Code: can run the whole notebook? Modular, concise, and documented code?
- □ Page limit: At most 3 pages. Plots should still be readable!

Task 1.1

- □ Discussion of feature engineering and scaling steps
- □ Build and train relevant models
- □ Evaluate model accuracy for different values of `SEX` and `RAC1P`
- □ Plot(s) comparing the metrics
- □ Discussion on whether the model is fair or unfair, and why

Task 1.2

- □ Code to implement the group fairness constraint
- □ Build and train relevant models
- □ Evaluate model accuracy for different values of `SEX` and `RAC1P`
- □ Plot(s) comparing the metrics
- □ Discussion on whether the model is more fair or unfair, compare to previous results where fairness constraint was not implemented (i.e. from Task 1.1)