INFO251 – Applied Machine Learning

Lab 9

Emily Aiken

Announcements

- Due dates for remaining problem sets
  - PS5 (Trees, forests, networks) – Monday April 4
  - PS6 (Fairness) – Monday April 18
  - PS7 (Unsupervised and deep learning) – Monday May 2
- Lowest problem set grade will be dropped
- Remaining labs
  - Lab 10 (April 6) -- Neural network regularization, CNNs, RNNs
  - Lab 11 (April 13) -- TBD
  - Lab 12 (April 20) – Unsupervised learning
  - Lab 13 (April 27) -- End-to-end machine learning (guest lecture)

Fairness for ML: Formal Criteria

- R: Predicted value of target variable
- A: Sensitive (protected) characteristic
- Y: True value of target variable

Also called "demographic parity" or "statistical parity"

Non-discrimination criteria		
Independence	Separation	Sufficiency
$R \perp A$	$R \perp A \mid Y$	$Y \perp A \mid R$

Also called "calibration"

Source: Barocas et al. (2020), fairmlbook.org

Tools for testing for fairness criteria

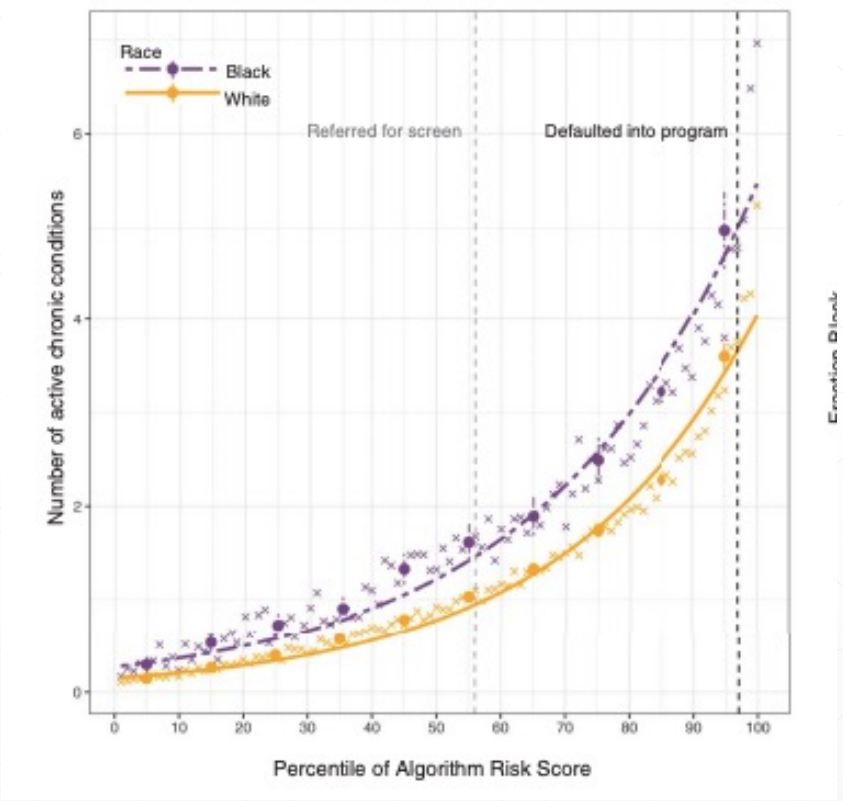
- Visual evaluation
- “Within-epsilon” comparisons
- Hypothesis testing
  - T-test
  - Regression analysis

Obermeyer et al. (2019)

Obermeyer et al. (2019). “Dissecting racial bias in an algorithm used to manage the health of populations.” *Science* Vol. 336 No. 6464. [Link]

- **Data**
  - ***Prediction***: Algorithmic “risk score” for “complex health needs” from the previous year
  - ***True value***: Realized health outcomes in the current year (chronic conditions)
  - ***Sensitive characteristic*****:** Race (white and black)
- **Results**
  - At given risk score, blank patients have substantially worse “true” health than white patients
  - Algorithm uses total medical expenditures to predict health needs; at given level of health black patients generate lower costs than white patients

Obermeyer et al. (2019)



- Ziad Obermeyer
  - Public Health 196: Artificial Intelligence for Medicine and Health Policy
  - Public Health 293: Data Science in Health Policy