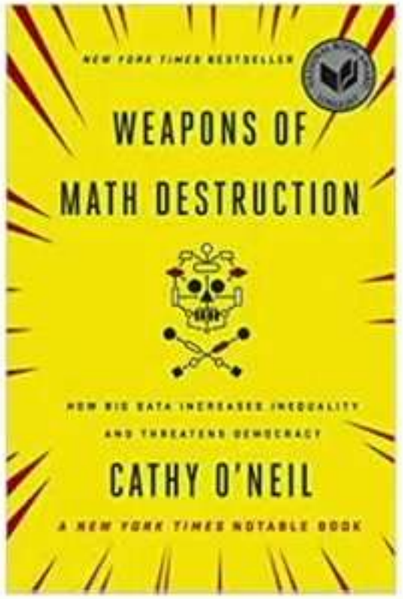
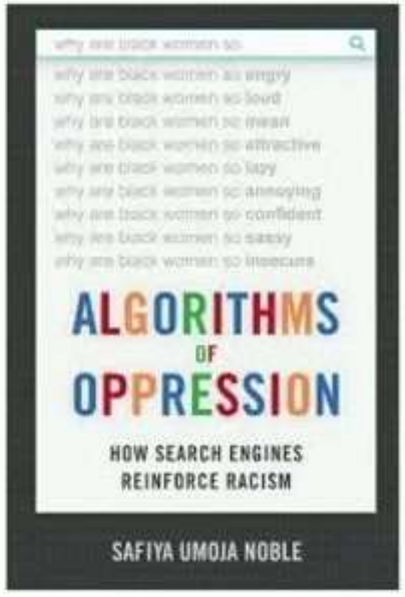
michael kearns + aaron roth

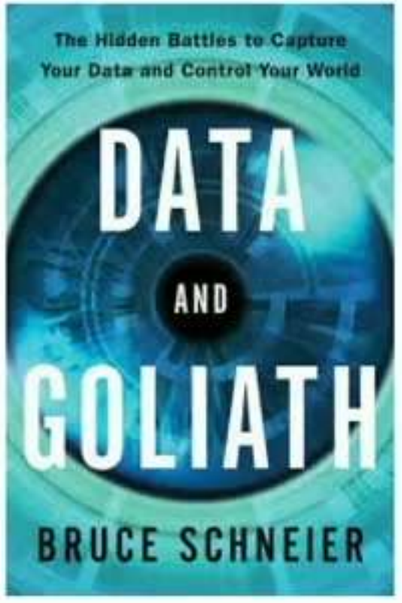
MICROSOFT RESEARCH AI
NOVEMBER 12, 2019

the ethical algorithm

the science of
socially aware algorithm design


NEW YORK TIMES BESTSELLER
WEAPONS OF
MATH DESTRUCTION
HOW BIG DATA INCREASES INEQUALITY
AND THREATENS DEMOCRACY
CATHY O'NEIL
A NEW YORK TIMES NOTABLE BOOK


why are black women so
why are black women so angry
why are black women so loud
why are black women so mean
why are black women so attractive
why are black women so lazy
why are black women so annoying
why are black women so confident
why are black women so sassy
why are black women so insecure
ALGORITHMS
OF
OPPRESSION
HOW SEARCH ENGINES
REINFORCE RACISM
SAFIYA UMOJA NOBLE

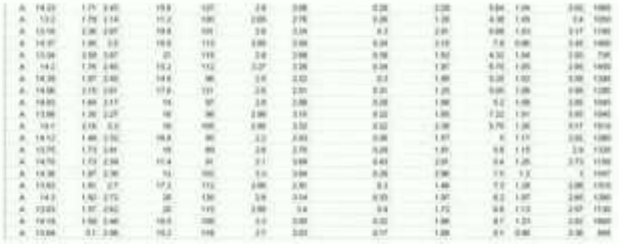

The Hidden Battles to Capture
Your Data and Control Your World
DATA
AND
GOLIATH
BRUCE SCHNEIER

Ethical Algorithms?

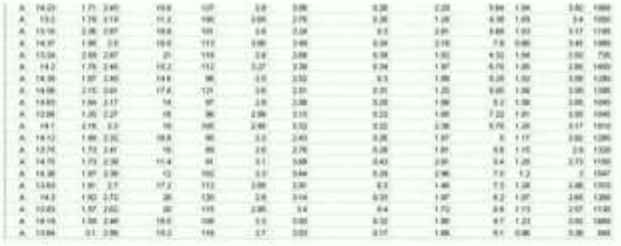
Ethical Algorithms?

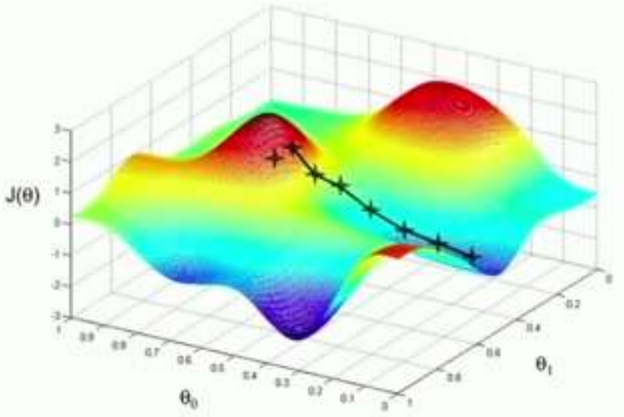
Algorithms: Hard to Assign Blame

Algorithms: Hard to Assign Blame

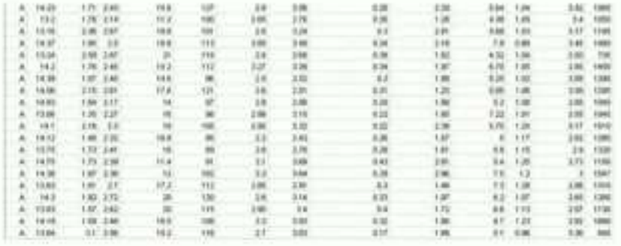


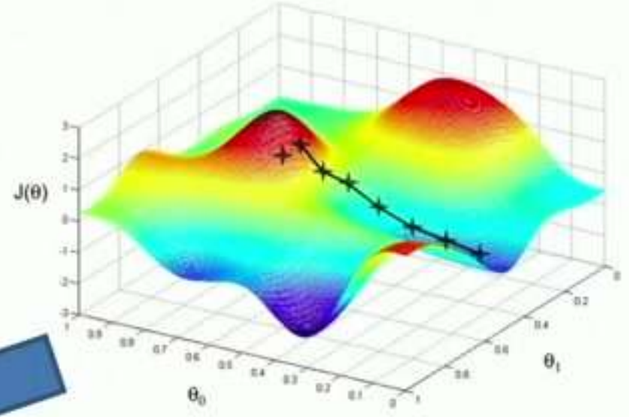
Algorithms: Hard to Assign Blame

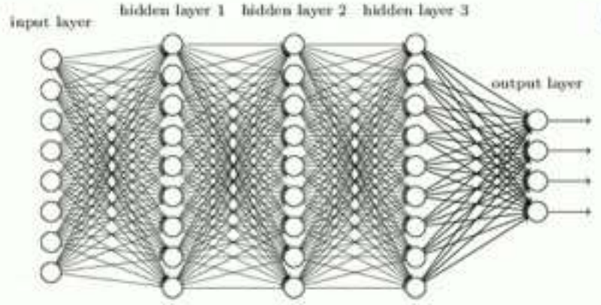




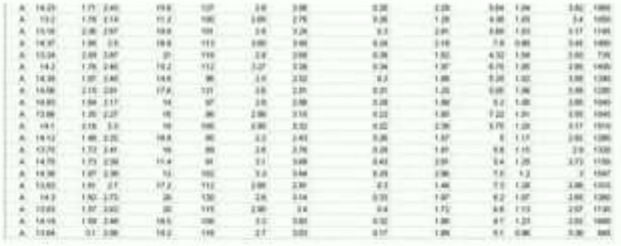
Algorithms: Hard to Assign Blame

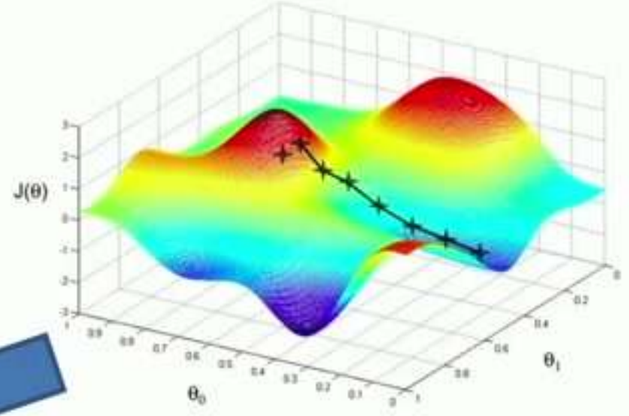


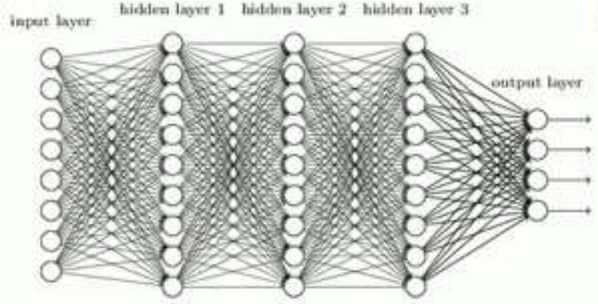




Algorithms: Hard to Assign Blame







Need to Embed Social Values in Algorithms

- Requires being precise about definitions, developing their consequences.
  - Privacy
  - Fairness
  - Accountability
  - Interpretability
  - Morality

"Anonymized Data Isn't"

Name	Age	Gender	Zip Code	Smoker	Diagnosis
*	60-70	Male	191**	Y	Heart disease
*	60-70	Female	191**	N	Arthritis
*	60-70	Male	191**	Y	Lung cancer
*	60-70	Female	191**	N	Crohn's Disease
*	60-70	Male	191**	Y	Lung Cancer
*	50-60	Female	191**	N	HIV
*	50-60	Male	191**	Y	Lyme Disease
*	50-60	Male	191**	Y	Seasonal Allergies
*	50-60	Female	191**	N	Ulcerative Colitis

Name	Age	Gender	Zip Code	Diagnosis
*	50-60	Female	191**	HIV
*	50-60	Female	191**	Lupus
*	50-60	Female	191**	Hip Fracture
*	60-70	Male	191**	Pancreatic Cancer
*	60-70	Male	191**	Ulcerative Colitis
*	60-70	Male	191**	Flu Like Symptoms

BRITISH MEDICAL JOURNAL

LONDON SATURDAY SEPTEMBER 30 1950

SMOKING AND CARCINOMA OF THE LUNG

PRELIMINARY REPORT

BY

RICHARD DOLL, M.D., M.R.C.P.

Member of the Statistical Research Unit of the Medical Research Council

AND

A. BRADFORD HILL, Ph.D., D.Sc.

Professor of Medical Statistics, London School of Hygiene and Tropical Medicine; Honorary Director of the Statistical Research Unit of the Medical Research Council

In England and Wales the phenomenal increase in the number of deaths attributed to cancer of the lung provides one of the most striking changes in the pattern of mortality recorded by the Registrar-General. For example, in the quarter of a century between 1922 and 1947 the annual number of deaths recorded increased from 612 to 9,287, or roughly fifteenfold. This remarkable increase is, of course, out of all proportion to the increase of population—both in total and, particularly, in its older age groups. Stocks (1947), using standardized death rates to allow for these population changes, shows the following trend: rate per 100,000 in 1901–20, males 1.1, females 0.7; rate per 100,000 in 1936–9, males 10.6, females 2.5. The rise seems to have been particularly rapid since the end of the first world war; between 1921–30 and 1940–4 the death rate of men at ages 45 and over increased sixfold and of women of the same ages approximately threefold. This increase is still continuing. It has occurred, too, in Switzerland, Denmark, the U.S.A., Canada, and Australia, and has been reported from Turkey and Japan.

Many writers have studied these changes, considering whether they denote a real increase in the incidence of the disease or are due merely to improved standards of diagnosis. Some believe that the latter factor can be regarded as wholly, or at least mainly, responsible—for example, Willis (1948), Clemmesen and Busk (1947), and Steiner (1944). On the other hand, Kennaway and Kennaway (1947) and Stocks (1947) have given good reasons for believing that the rise is at least partly real. The latter, for instance, has pointed out that "the increase of certified respiratory cancer mortality during the past 20 years has been as rapid in country districts as in the cities with the best diagnostic facilities, a fact which does not support the view that such increase merely reflects improved diagnosis of cases previously certified as bronchitis or other respiratory affections." He also draws attention to differences in mortality between some of the large cities of England and Wales, differences which it is difficult to explain in terms of diagnostic standards.

The large and continued increase in the recorded deaths even within the last five years, both in the national figures and in those from teaching hospitals, also makes it hard to believe that improved diagnosis is entirely responsible. In short, there is sufficient reason to reject that factor as the whole explanation, although no one would deny that it may well have been contributory. As a corollary, it is right and proper to seek for other causes.

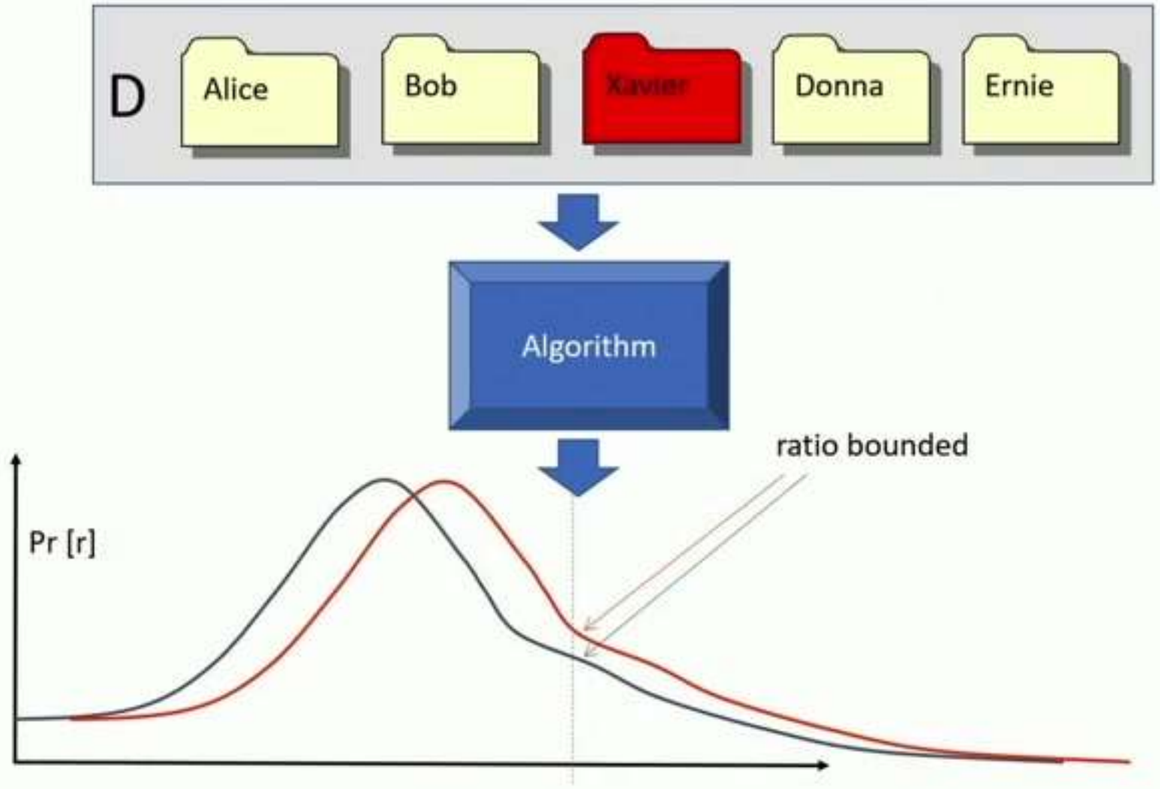
Possible Causes of the Increase

Two main causes have from time to time been put forward: (1) a general atmospheric pollution from the exhaust fumes of cars, from the surface dust of tarred roads, and from gas-works, industrial plants, and coal fires; and (2) the smoking of tobacco. Some characteristics of the former have certainly become more prevalent in the last 50 years, and there is also no doubt that the smoking of cigarettes has greatly increased. Such associated changes in time can, however, be no more than suggestive, and until recently there has been singularly little more direct evidence. That evidence, based upon clinical experience and records, relates mainly to the use of tobacco. For instance, in Germany, Müller (1939) found that only 3 out of 86 male patients with cancer of the lung were non-smokers, while 56 were heavy smokers, and, in contrast, among 86 "healthy men of the same age groups" there were 14 non-smokers and only 31 heavy smokers. Similarly, in America, Schrek and his co-workers (1950) reported that 14.6% of 82 male patients with cancer of the lung were non-smokers, against 23.9% of 522 male patients admitted with cancer of sites other than the upper respiratory and digestive tracts. In this country, Thelwall Jones (1949—personal communication) found 8 non-smokers in 82 patients with proved carcinoma of the lung, compared with 11 in a corresponding group of patients with diseases other than cancer; this difference is slight, but it is more striking that there were 28 heavy smokers in the cancer group, against 14 in the comparative group.

Clearly none of these small-scale inquiries can be accepted as conclusive, but they all point in the same direction. Their evidence has now been borne out by the results of a large-scale inquiry undertaken in the U.S.A. by Wynder and Graham (1950).

Wynder and Graham found that of 605 men with epidermoid, undifferentiated, or histologically unclassified types of bronchial carcinoma only 1.3% were "non-smokers"—that is, had averaged less than one cigarette a day for the last 20 years—whereas 51.2% of them had smoked more than 20 cigarettes a day over the same

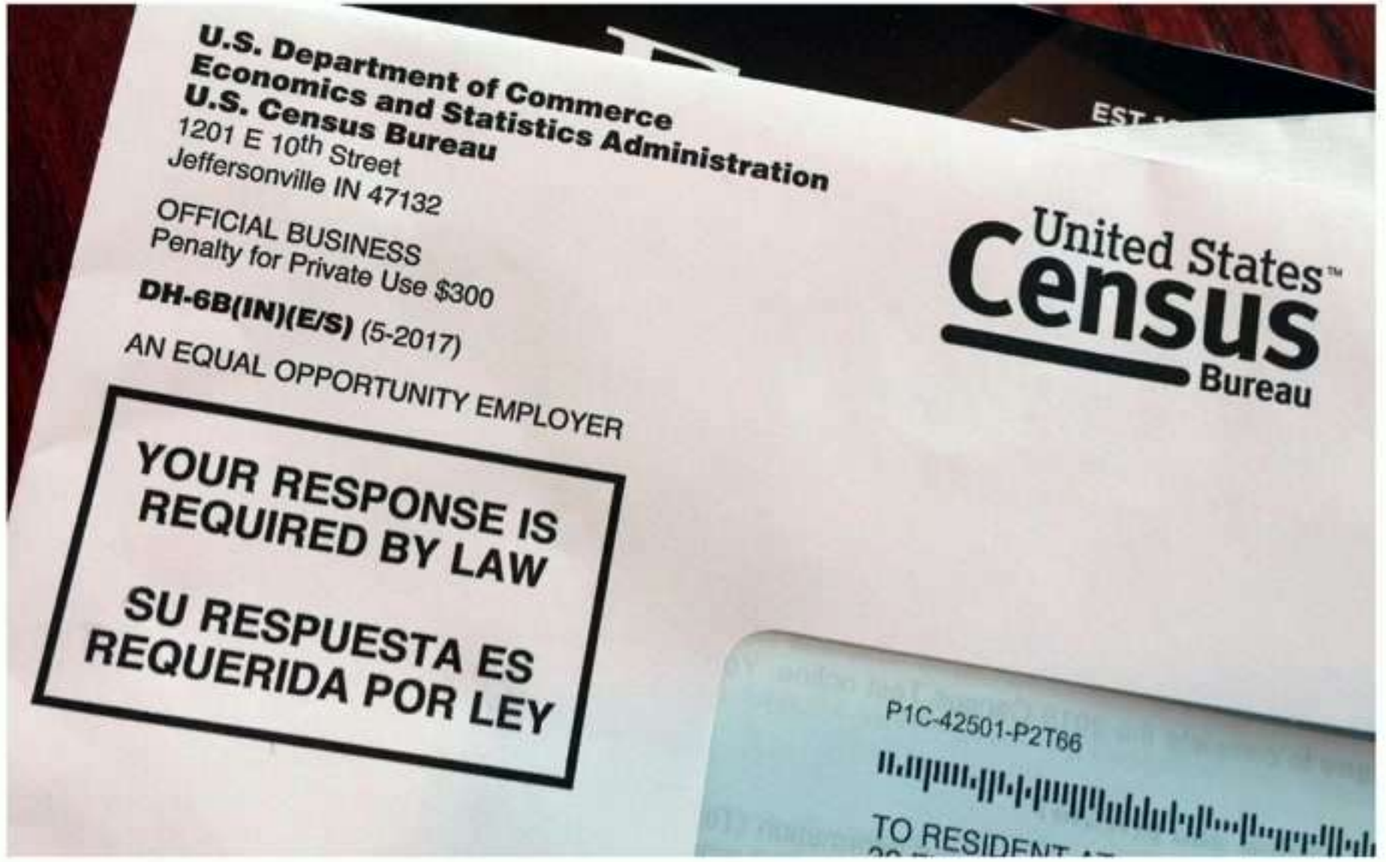
Differential Privacy



TheUpshot

To Reduce Privacy Risks, the Census Plans to Report Less Accurate Data

Guaranteeing people's confidentiality has become more of a challenge, but some scholars worry that the new system will impede research.



A 2018 census test letter mailed to a resident in Providence, R.I. The nation's test run of the 2020 Census is in Rhode Island. Michelle R. Smith/Associated Press

Fairness: A Work in Progress

- Don't agree on the definitions.
- Only beginning to understand *tradeoffs* between different kinds of fairness, and between fairness and accuracy.

Fairness: A Work in Progress

- Don’t agree on the definitions.
- Only beginning to understand *tradeoffs* between different kinds of fairness, and between fairness and accuracy.

Why might machine learning be “unfair”?


THE WALL STREET JOURNAL.

Home World U.S. Politics Economy Business Tech Markets Opinion Life & Arts Real Estate WSJ Magazine

U.S. | NEW YORK

New York Regulator Probes UnitedHealth Algorithm for Racial Bias

Financial Services Department is investigating whether algorithm violates state antidiscrimination law


Bloomberg

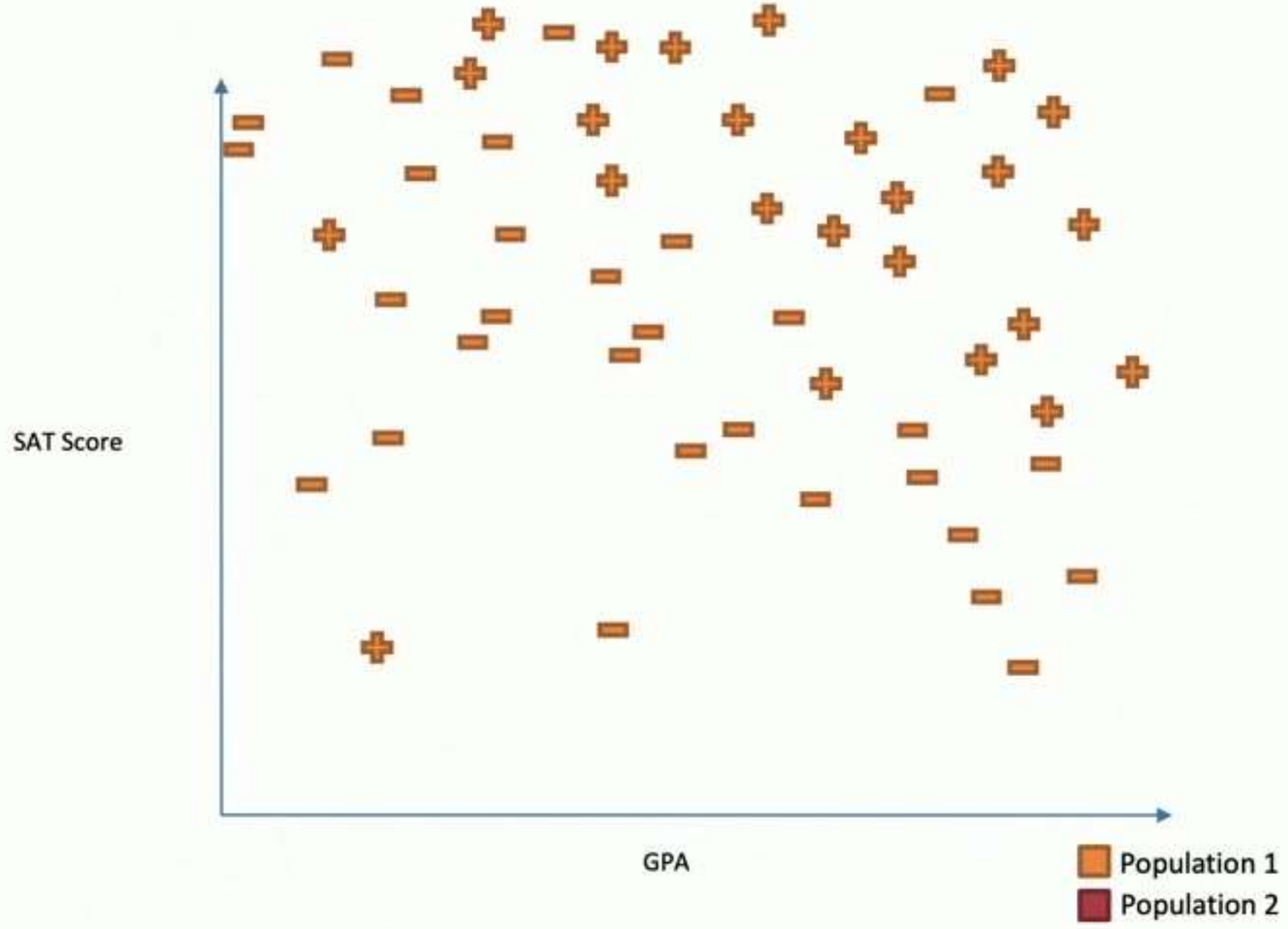
Business

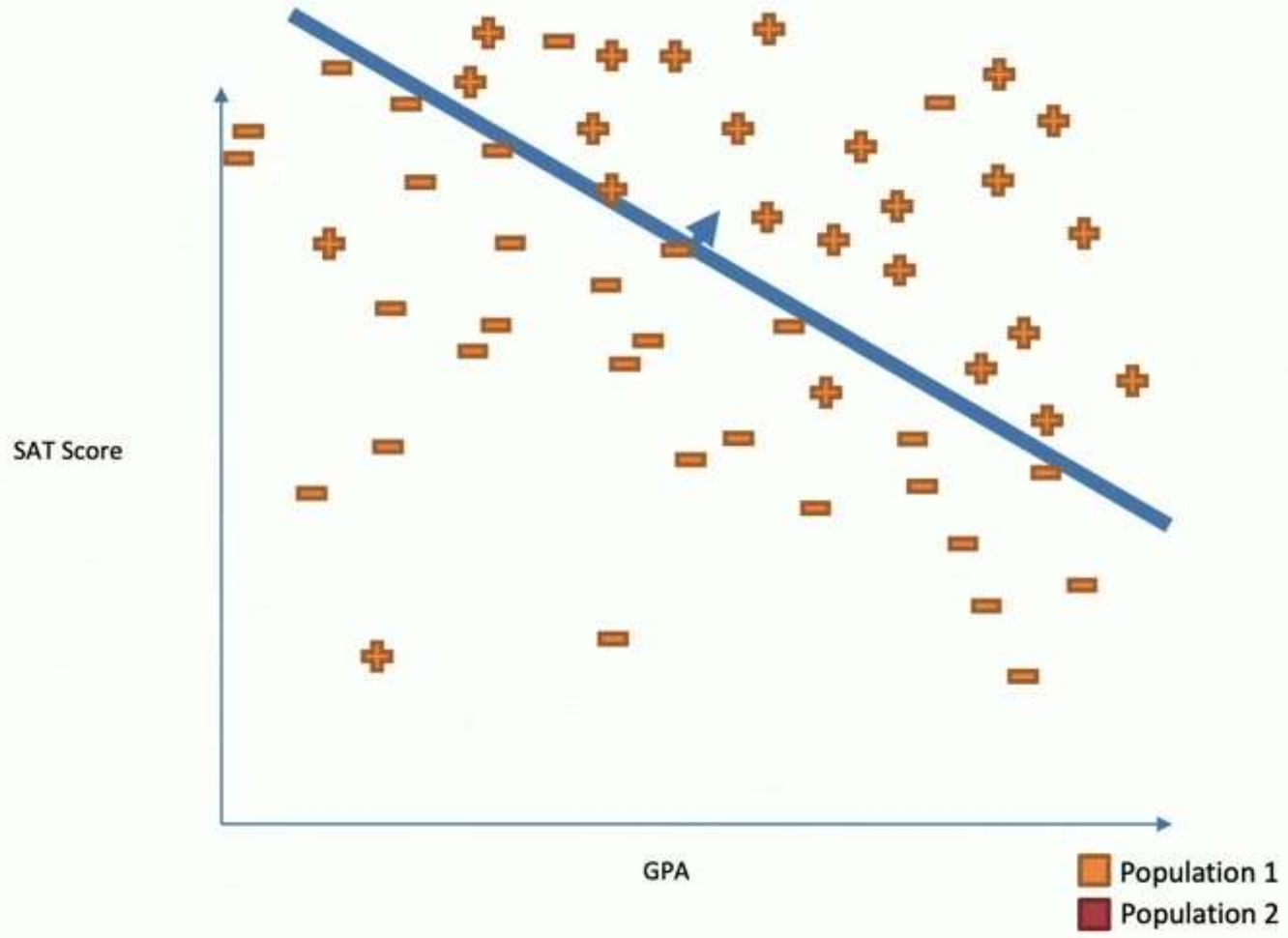
Viral Tweet About Apple Card Leads to Goldman Sachs Probe

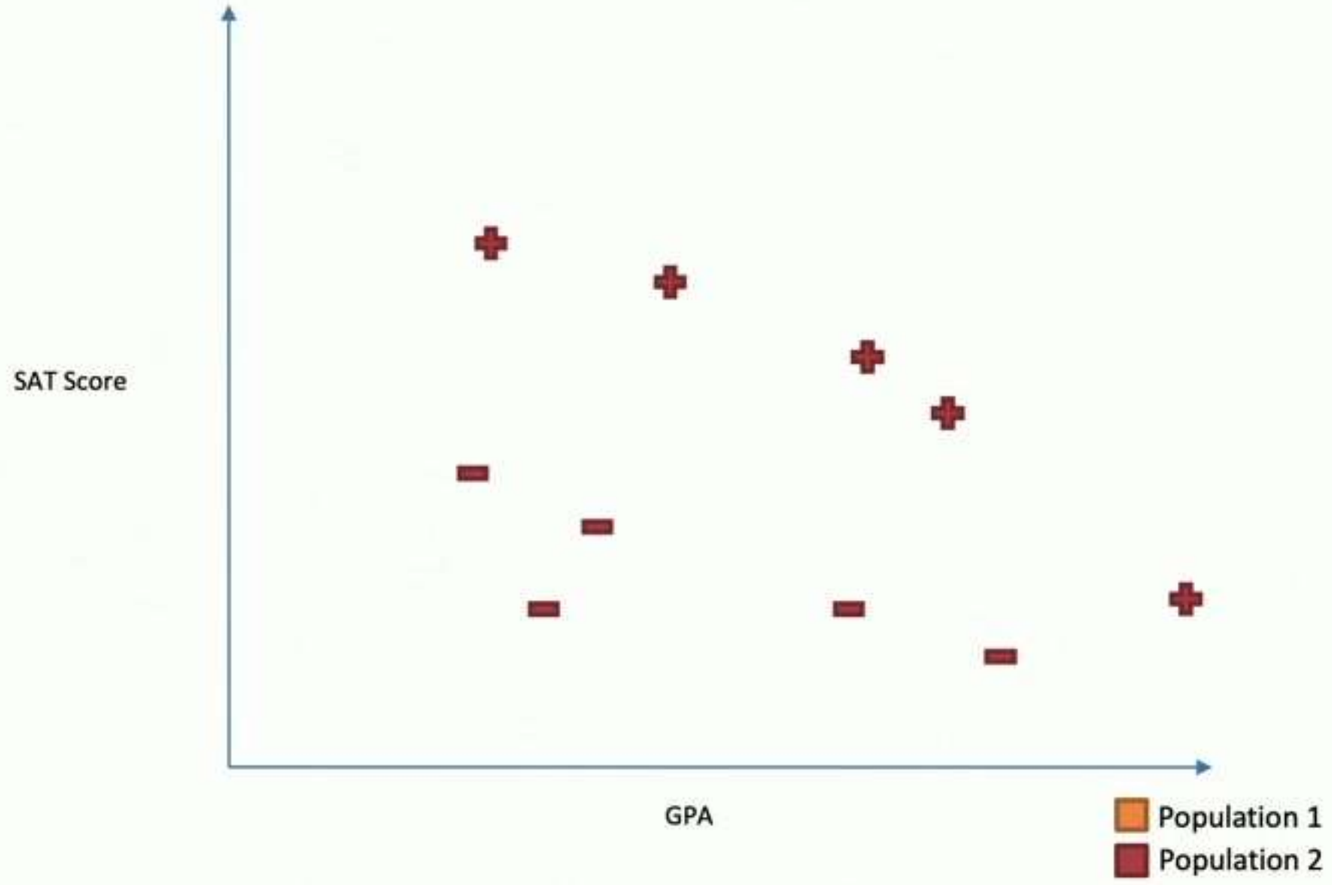
By Sridhar Natarajan and Shahien Nasiripour

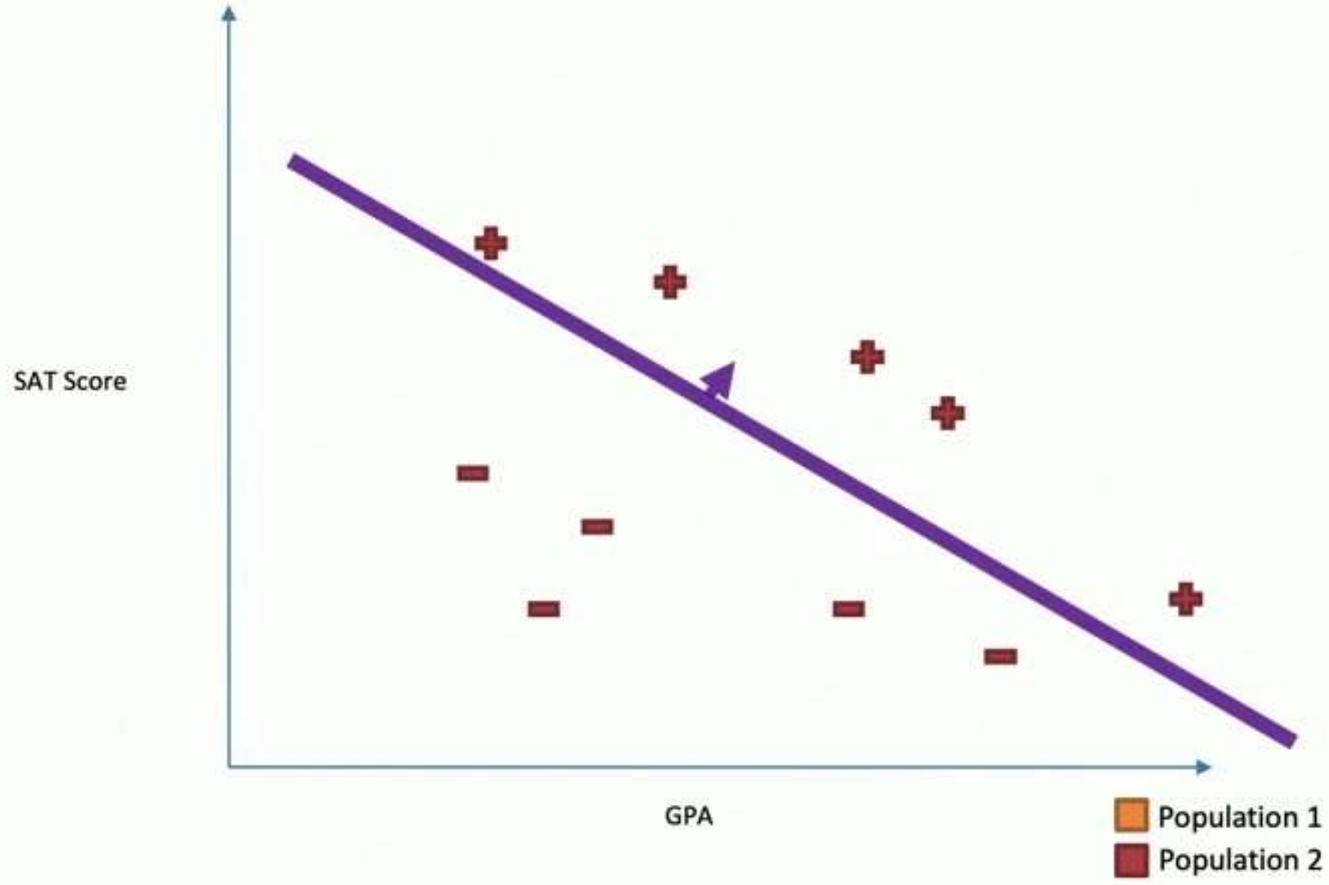
November 9, 2019, 3:52 PM EST *Updated on November 9, 2019, 8:53 PM EST*

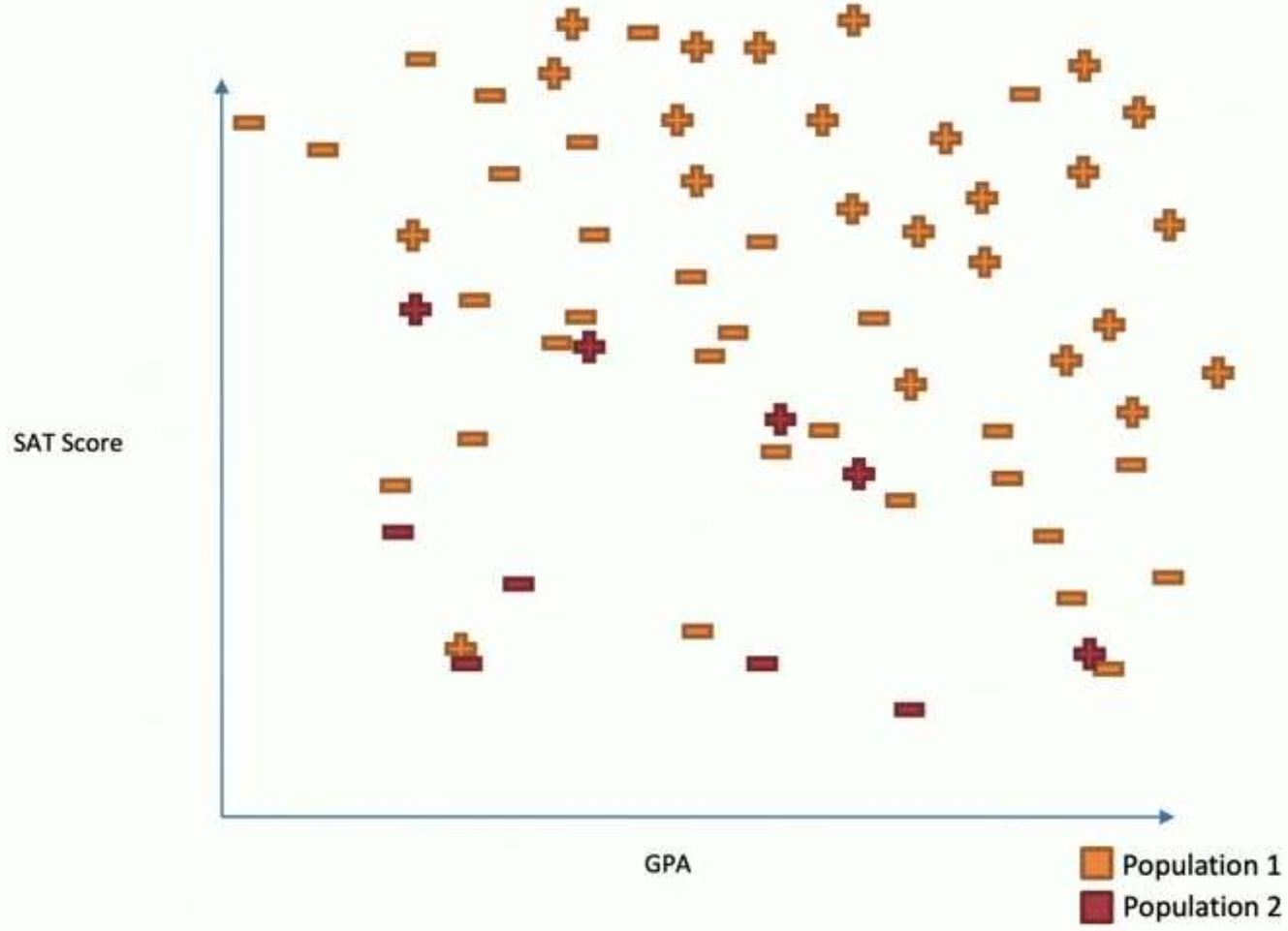
- Tech entrepreneur alleged inherent bias in algorithms for card
- The card is part of Goldman's new main street business lines

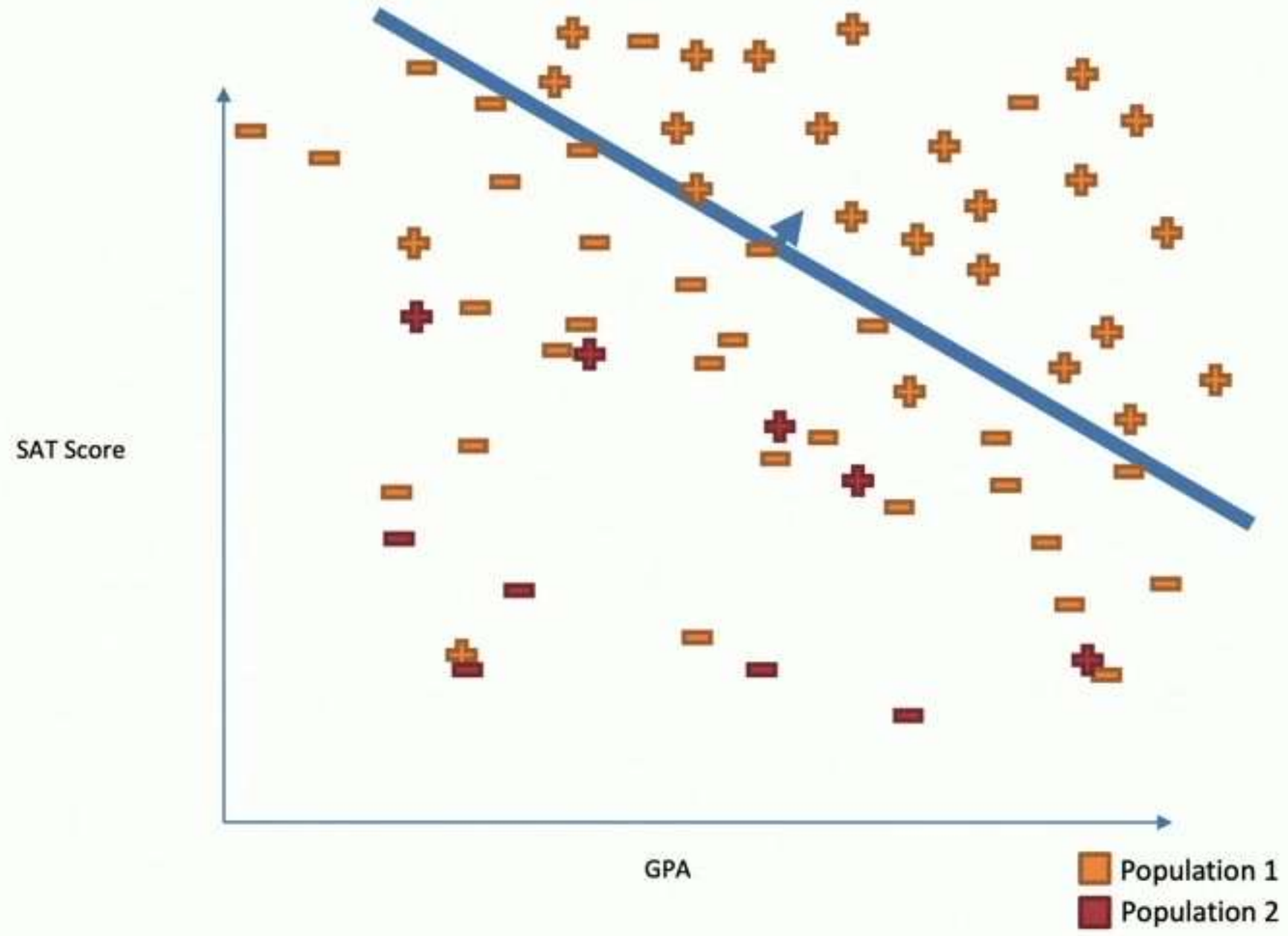

SAT Score
GPA
Population 1
Population 2

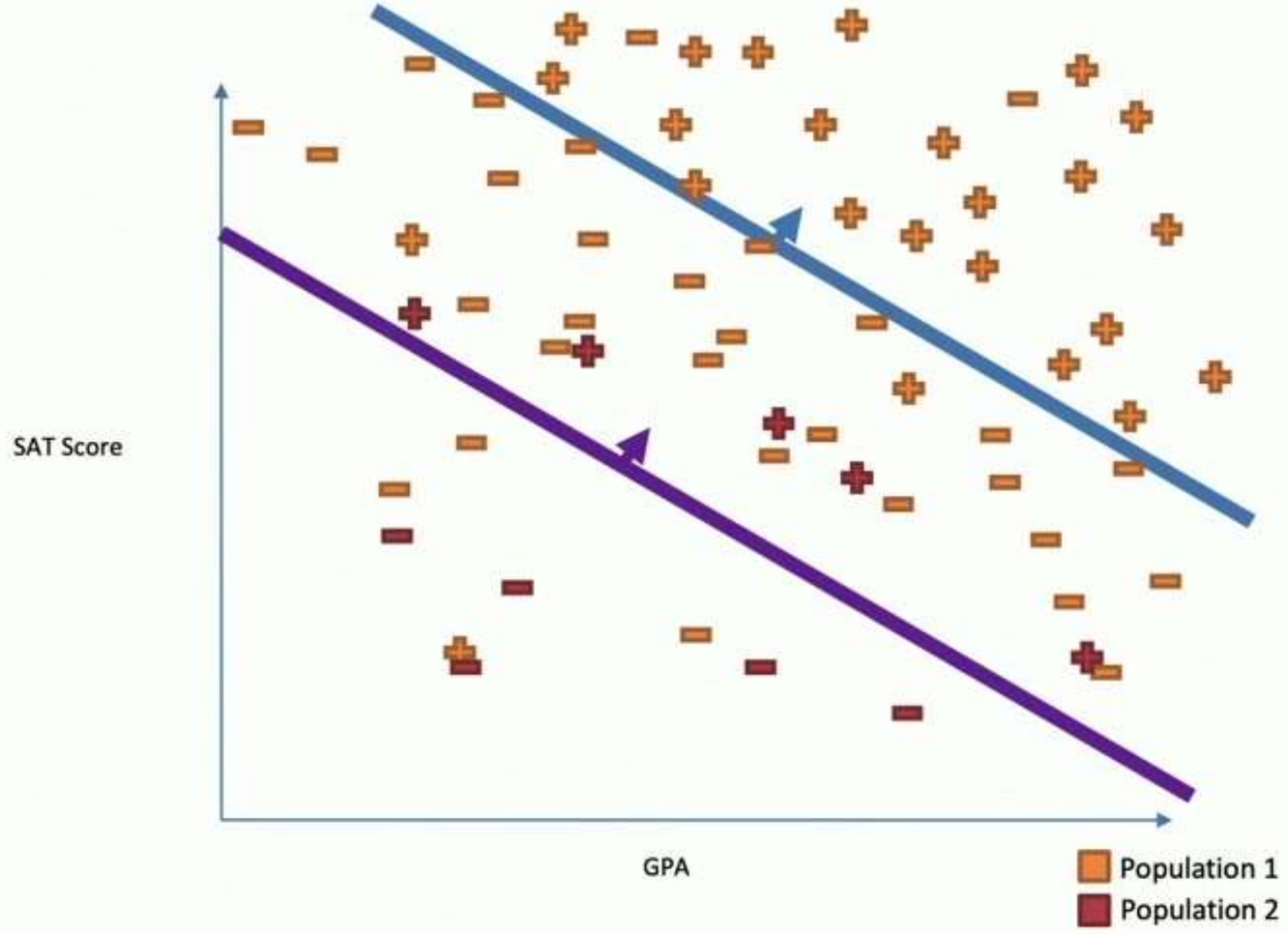

SAT Score
GPA
Population 1
Population 2

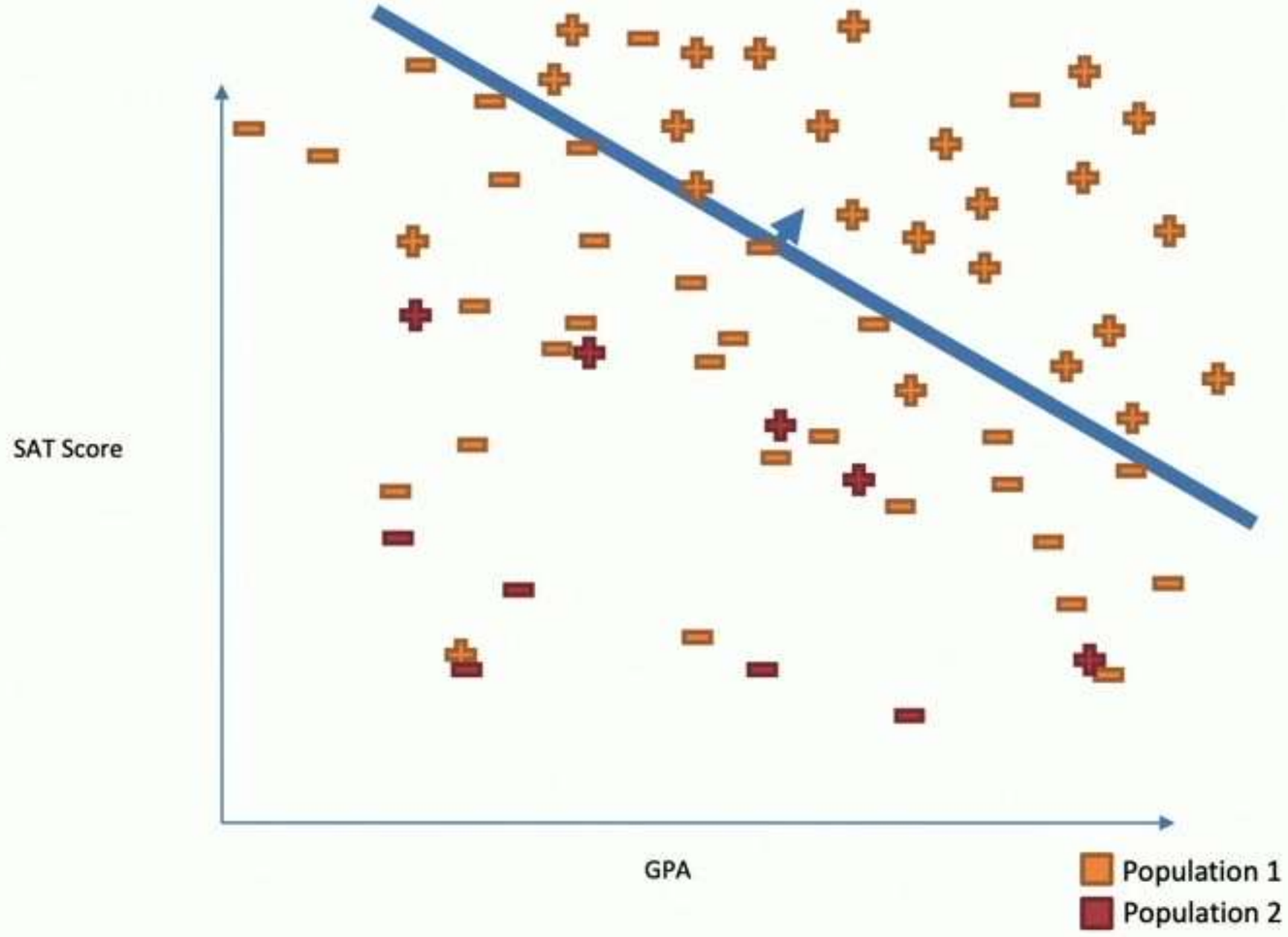

SAT Score
GPA
Population 1
Population 2

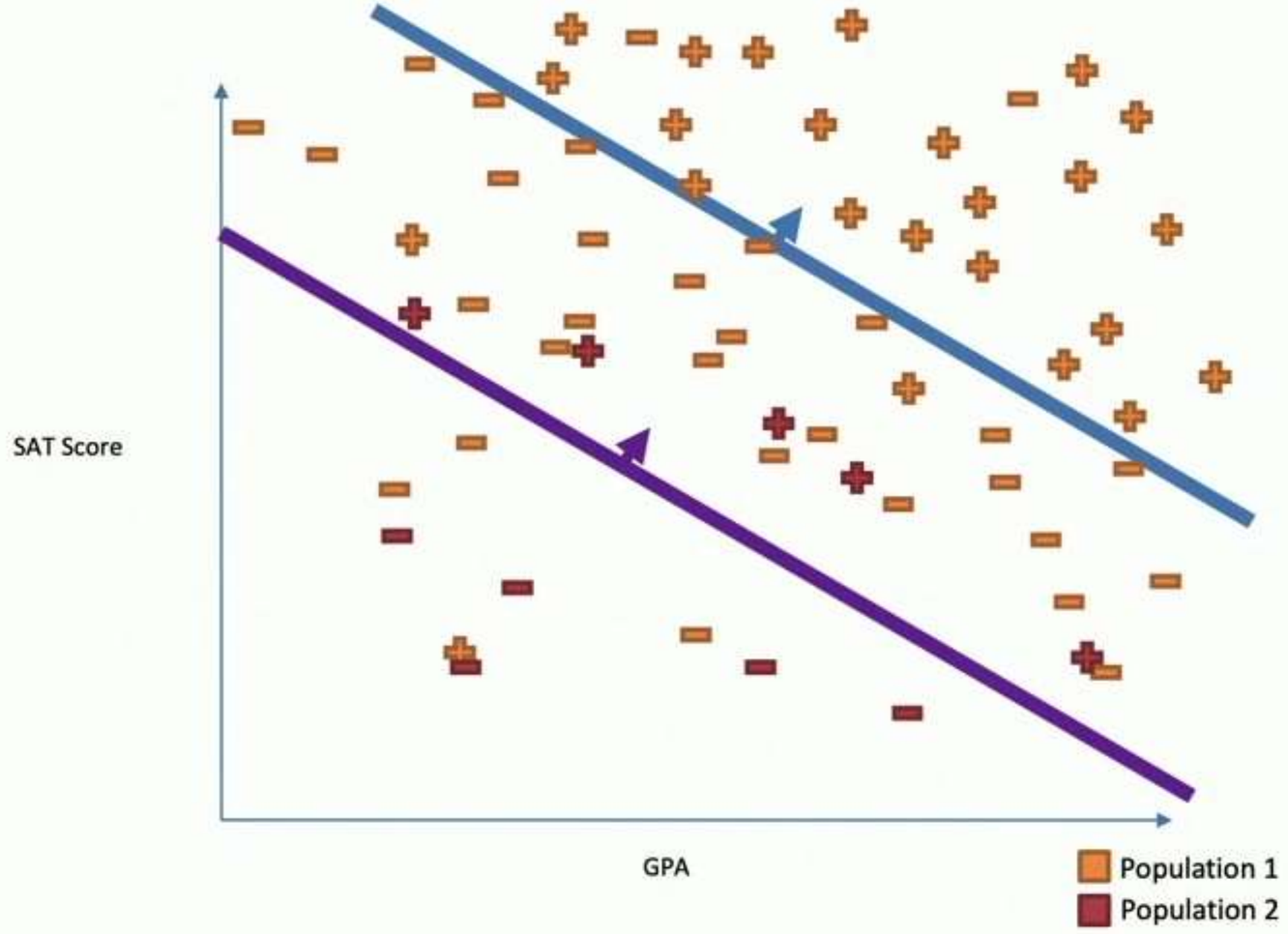

SAT Score
GPA
Population 1
Population 2


SAT Score
GPA
Population 1
Population 2

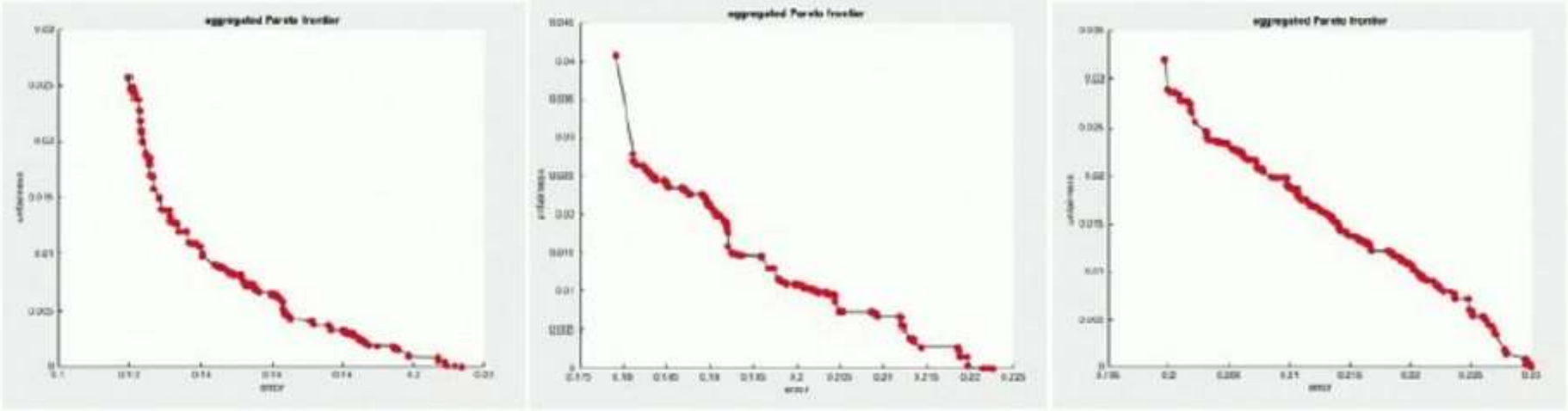

SAT Score
GPA
Population 1
Population 2


SAT Score
GPA
Population 1
Population 2

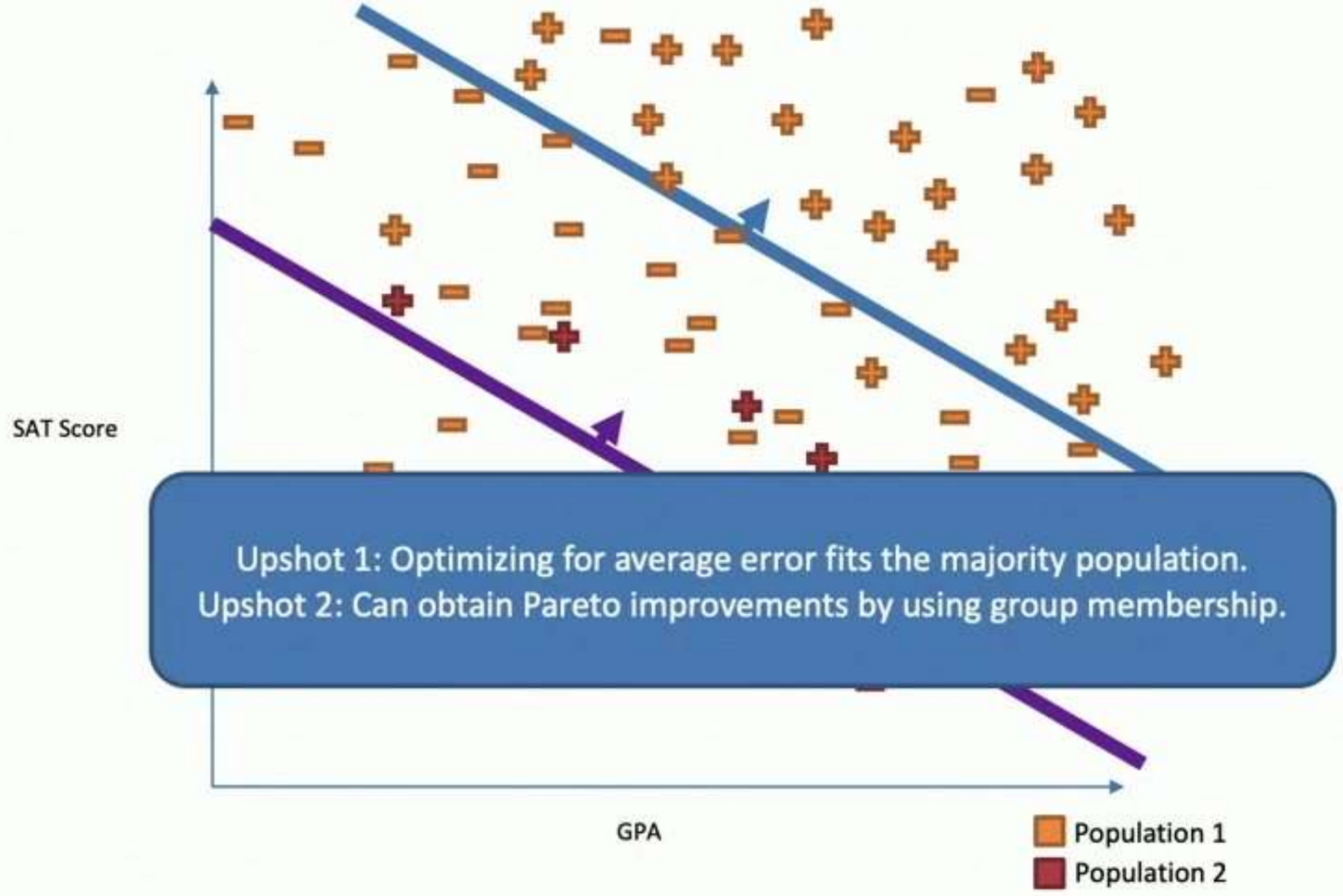

SAT Score
GPA
Population 1
Population 2


SAT Score
GPA
Population 1
Population 2

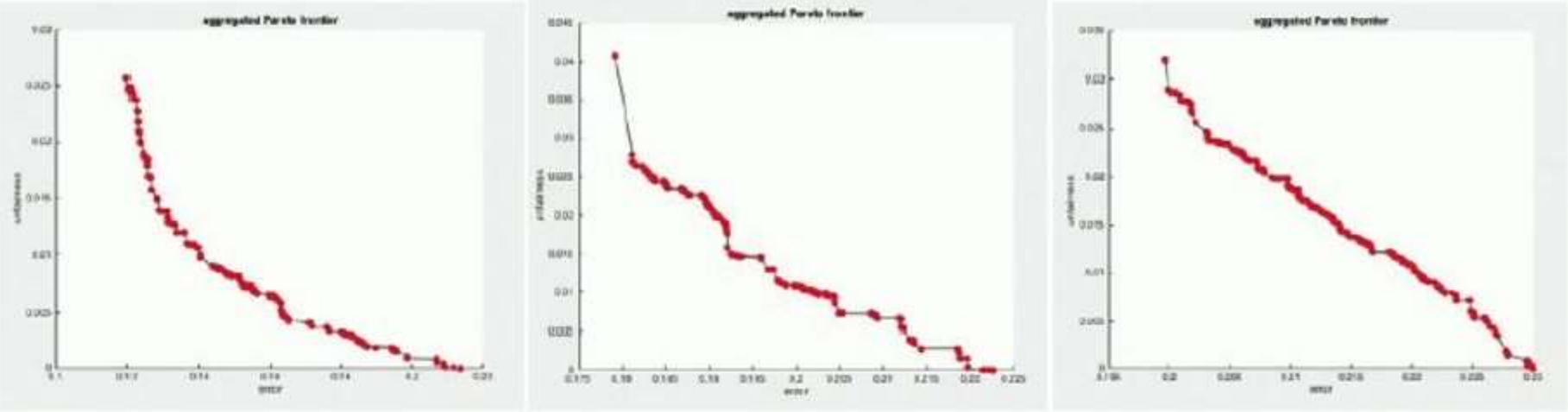
Efficient Frontiers



Examples of Pareto frontiers of error (x axis) and an unfairness measure (y axis) for three different real data sets. The curves differ in their shapes and the actual numeric values on the error and fairness axes, thus presenting different trade-offs.


SAT Score
GPA
Upshot 1: Optimizing for average error fits the majority population.
Upshot 2: Can obtain Pareto improvements by using group membership.
Population 1
Population 2

Efficient Frontiers



Examples of Pareto frontiers of error (x axis) and an unfairness measure (y axis) for three different real data sets. The curves differ in their shapes and the actual numeric values on the error and fairness axes, thus presenting different trade-offs.

Other Topics

Games People Play (with Algorithms)

Games Scientists Play (with Data)

Interpretability, Accountability, Morality…. The Singularity

Frontiers of Fairness: From Groups to Individuals

Joint works with Chris Jung, Seth Neel, Aaron Roth, Saeed Sharifi, Logan Stapleton, Steven Wu

Types of Fairness Definitions

- Group Fairness
  - E.g. equality of error or false negative rates across gender, racial groups, etc.
  - Strong theory, practical implementations (e.g. fairness regularization)
  - But no guarantees to individuals
- Individual Fairness
  - E.g. metric fairness ("fairness through awareness"), meritocratic fairness
  - Binds at the individual level
  - But strong assumptions required (e.g. realizability) have prevented practical implementations

A Framework for Fair ML

- Begin by expressing training as a (linear or convex) constrained optimization problem
  - E.g. minimize error subject to various fairness constraints
  - In interesting cases, model space of learning algo and number of constraints may be exponential/infinite
  - Want to avoid explicit enumeration
- Use LP duality to pass to Lagrangian and recast as two-player, zero-sum game
  - Learner/Primal: wants to minimize error subject to constraints so far
  - Regulator/Dual: presents learner with violated constraints
  - Nash equilibrium is solution to constrained optimization problem
- If we can:
  - Formulate best responses as instances of *cost-sensitive classification*
  - Implement at least one player as a *no-regret* algorithm w.r.t. their strategy space

  … then algorithm provably converges in polynomial time given access to a standard learning heuristic
- Directly implement on top of your favorite "unfair" learning algorithm
- Applications:
  - Preventing "fairness gerrymandering"
  - Subjective individual fairness
  - Average individual fairness

Average Individual Fairness