



Fairness, Accountability,
Transparency in
Machine Learning
FATML

A red curved line graphic is positioned on the left side of the title text, starting from the top and curving down towards the bottom right.

Solon Barocas and Moritz Hardt

NIPS 2014



CENTER FOR
INFORMATION
TECHNOLOGY
POLICY

CITP.PRINCETON.EDU

NOV 28 :
TUTORIALS
DENVER CO

NOV 29 - DEC 1:
REGULAR SESSIONS
DENVER CO

DEC 2-3:
WORKSHOPS
VAIL CO

NIPS'94

NEURAL INFORMATION PROCESSING SYSTEMS: NATURAL & SYNTHETIC

MAJOR TOPICS

ALGORITHMS AND ARCHITECTURES
VISUAL AND IMAGE PROCESSING
COGNITIVE SCIENCE & AI
APPLICATIONS
THEORY
NEUROSCIENCE
HARDWARE IMPLEMENTATION
SPEECH AND SIGNAL PROCESSING
CONTROL, NAVIGATION, & PLANNING



TUTORIALS

Andrew Barto, *Univ. of Mass., Amherst*
"Reinforcement Learning"

Leo Breiman, *UC Berkeley*
"Statistics and Nets: Understanding Nonlinear Models from Their Linear Relatives"

Dan Hammerstrom, *OGI & Adaptive Solutions*
"A Survey of Pattern Recognition Hardware"

Michael Kearns, *AT&T Bell Laboratories*
"Recent Advances in Learning Theory"

Tuovo Kohonen, *Helsinki Univ. of Technology*
"Advances in the Theory and Applications of the Self-Organizing Map"

Marc Raichle, *Washington Univ. Med. School*
"Images of the Mind: A Tutorial"

INVITED PRESENTATIONS

NIPS'94 IS THE 8TH ANNUAL MEETING OF THE PREMIER INTERDISCIPLINARY CONFERENCE COVERING ALL ASPECTS OF NEURAL PROCESSING AND COMPUTATION. NIPS BRINGS TOGETHER NEUROSCIENTISTS, COMPUTER SCIENTISTS, COGNITIVE SCIENTISTS, ENGINEERS, PHYSICISTS, MATHEMATICIANS, AND STATISTICIANS WITH INTERESTS IN NATURAL AND ARTIFICIAL NEURAL SYSTEMS. THE SINGLE-TRACK CONFERENCE INCLUDES INVITED TALKS, AND ORAL AND POSTER PRESENTATIONS OF STRINGENTLY REFERRED PAPERS. A FULL DAY OF STATE-OF-THE-ART TUTORIAL PRESENTATIONS PRECEDES THREE DAYS OF REGULAR SESSIONS. TWO DAYS OF FOCUSED WORKSHOPS FOLLOW AT A WORLD-CLASS SKI RESORT.

20 years later

Machine learning used for consequential decision-making in such areas as commerce, employment, healthcare, education, and policing



Gilian Tett (Financial Times):

"After all, as the former CPD [Chicago Police Department] computer experts point out, the **algorithms in themselves are neutral**. 'This program had absolutely nothing to do with race... but multi-variable equations,' argues Goldstein. Meanwhile, the potential benefits of predictive policing are profound."

Data as social mirror

Instant Checkmate

Ads by Google

[Latanya Sweeney, Arrested?](#)

1) Enter Name and State. 2) Access Full Background Checks Instantly.

www.instantcheckmate.com/

[Latanya Sweeney](#)

Public Records Found For: **Latanya Sweeney**. View Now.

www.publicrecords.com/

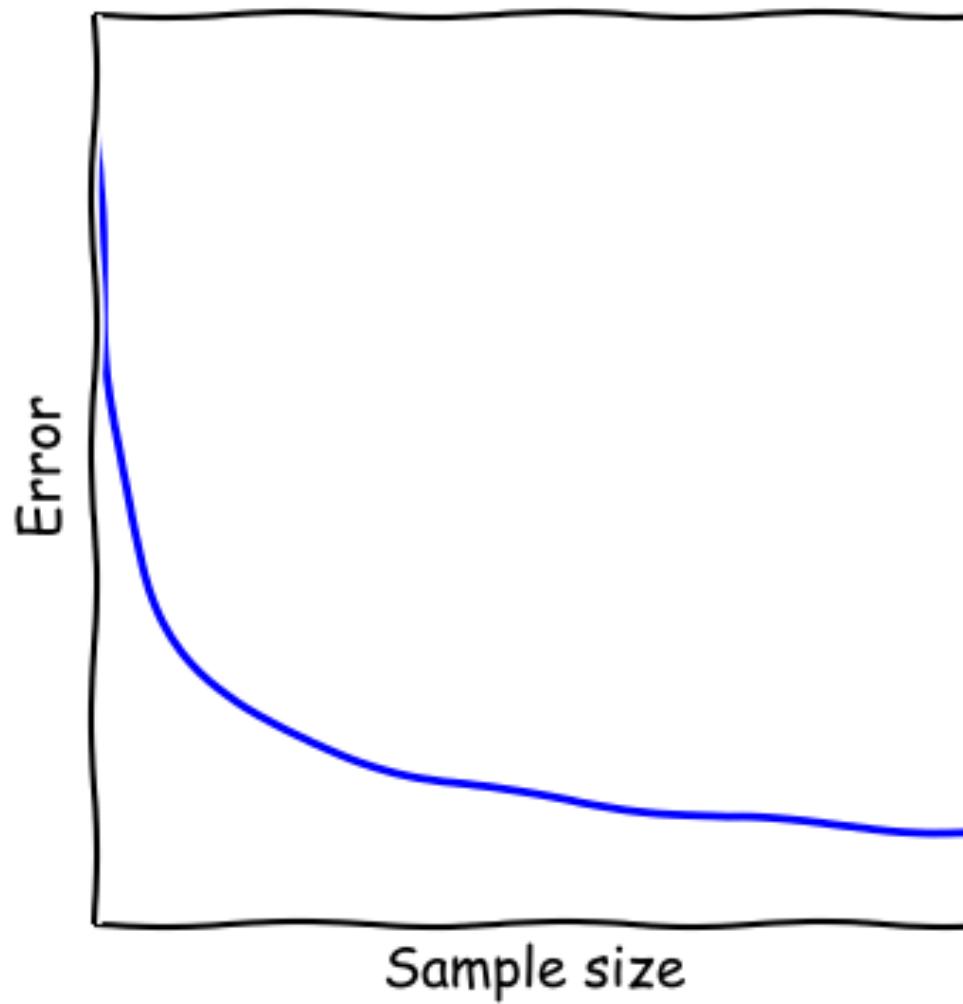
[La Tanya](#)

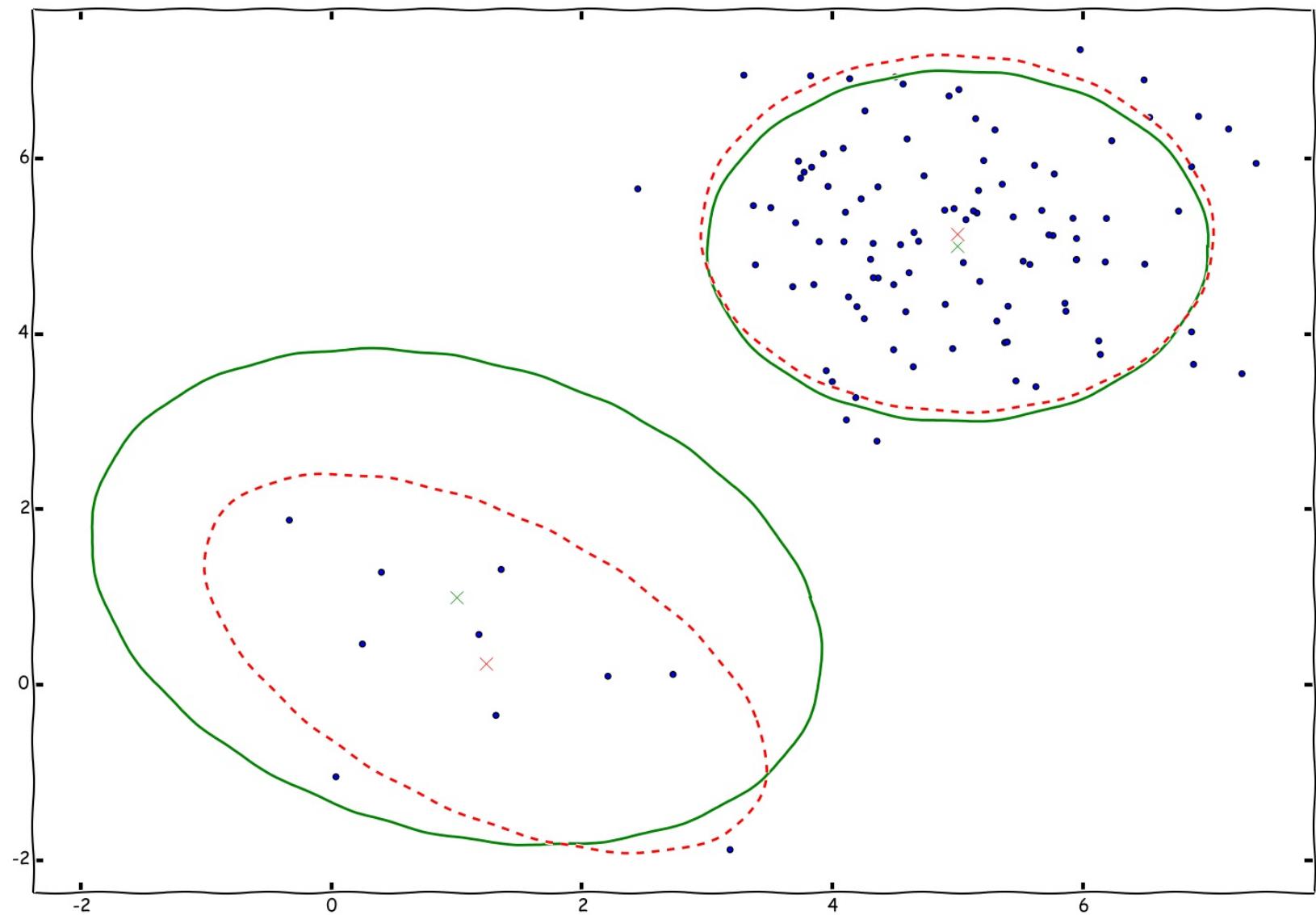
Search for La Tanya Look Up Fast Results now!

www.ask.com/La+Tanya

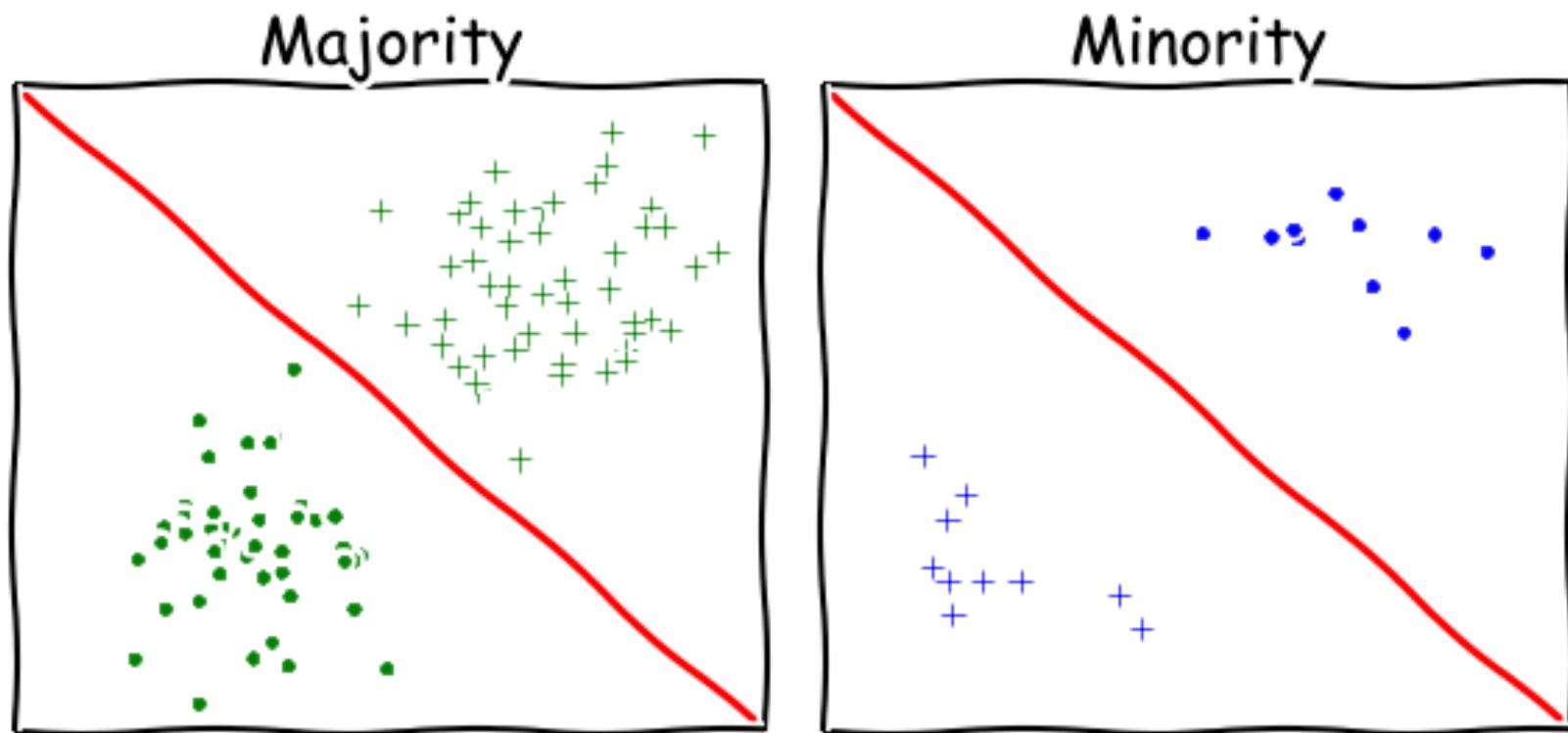
Sweeney, L., 2013. Discrimination in Online Ad Delivery.
Communications of the ACM, 56(5), pp.44–54.

Sample size disparity

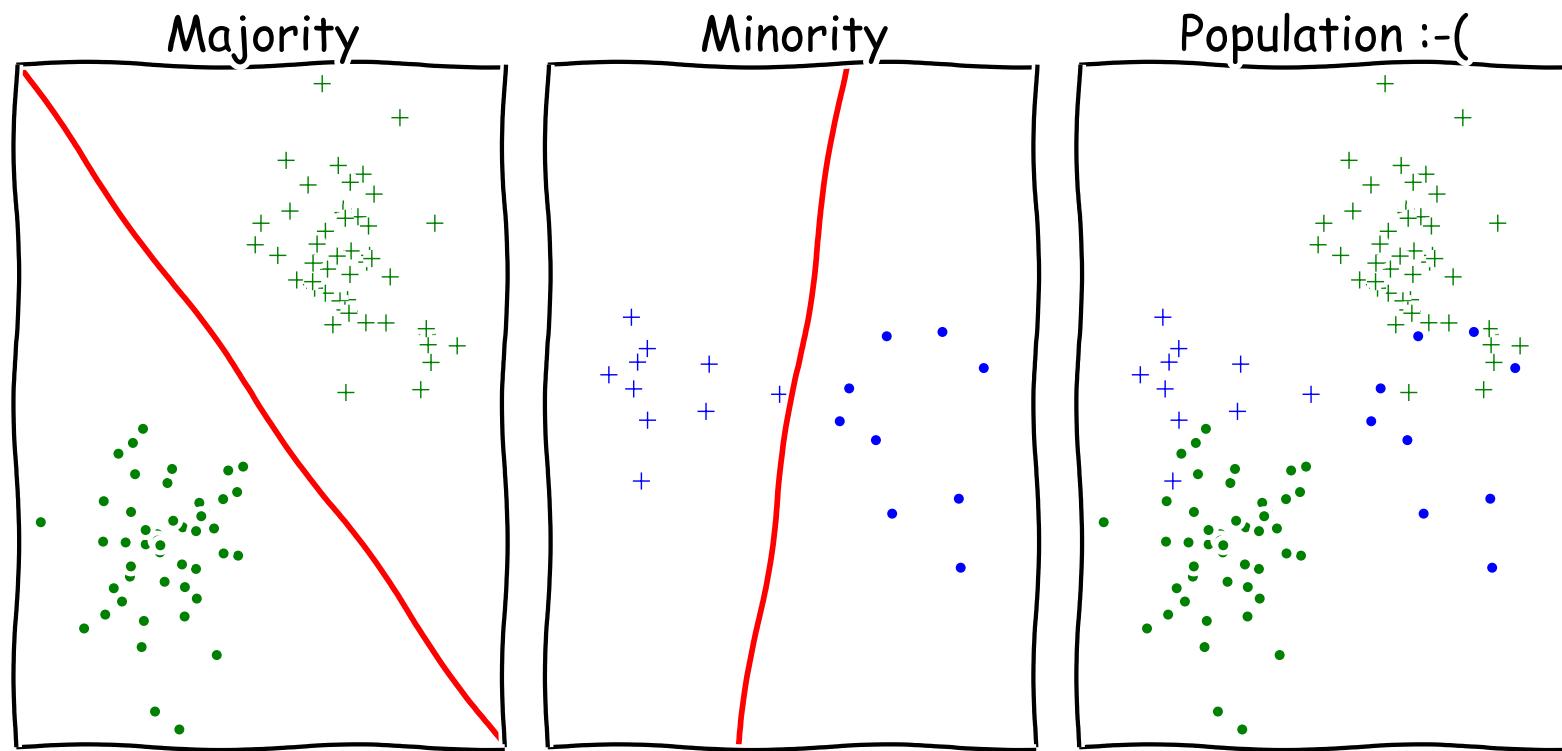




Cultural diversity



Increased complexity



Noise and the meaning of 5%

How do we distinguish between intrinsic model error and discrimination?

Redundant Encodings

How can we determine when an attribute is sufficiently relevant to justify its consideration, despite the fact that it is highly correlated with proscribed features?

- Accuracy/fairness trade-off

Interpretability

If we don't know how be "fair," can we at least explain what lead to a classification outcome
– Accuracy/interpretability trade-off

Accountability

How decisions are made, why they are made that way, and what are their effects

- Procedural fairness
- Substantive fairness



BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES

Executive Office of the President

MAY 2014



REPORT TO THE PRESIDENT BIG DATA AND PRIVACY: A TECHNOLOGICAL PERSPECTIVE

Executive Office of the President
President's Council of Advisors on
Science and Technology

May 2014





The Leadership
Conference

*The nation's premier
civil & human rights coalition*

Civil Rights Principles for the Era of Big Data

Technological progress should bring greater safety, economic opportunity, and convenience to everyone. And the collection of new types of data is essential for documenting persistent inequality and discrimination. At the same time, as new technologies allow companies and government to gain greater insight into our lives, it is vitally important that these technologies be designed and used in ways that respect the values of equal opportunity and equal justice. We aim to:

1. **Stop High-Tech Profiling.** New surveillance tools and data gathering techniques that can assemble detailed information about any person or group create a heightened risk of profiling and discrimination. Clear limitations and robust audit mechanisms are necessary to make sure that if these tools are used it is in a responsible and equitable way.
2. **Ensure Fairness in Automated Decisions.** Computerized decisionmaking in areas such as employment, health, education, and lending must be judged by its impact on real people, must operate fairly for all communities, and in particular must protect the interests of those that are disadvantaged or that have historically been the subject of discrimination. Systems that are blind to the preexisting disparities faced by such communities can easily reach decisions that reinforce existing inequities. Independent review and other remedies may be necessary to assure that a system works fairly.
3. **Preserve Constitutional Principles.** Search warrants and other independent oversight of law enforcement are particularly important for communities of color and for religious and ethnic minorities, who often face disproportionate scrutiny. Government databases must not be allowed to undermine core legal protections, including those of privacy and freedom of association.
4. **Enhance Individual Control of Personal Information.** Personal information that is known to a corporation — such as the moment-to-moment record of a person's movements or communications — can easily be used by companies and the government against vulnerable populations, including women, the formerly incarcerated, immigrants, religious minorities, the LGBT community, and young people. Individuals should have meaningful, flexible control over how a corporation gathers data from them, and how it uses and shares that data. Non-public information should not be disclosed to the government without judicial process.
5. **Protect People from Inaccurate Data.** Government and corporate databases must allow everyone — including the urban and rural poor, people with disabilities, seniors, and people who lack access to the Internet — to appropriately ensure the accuracy of personal information that is used to make important decisions about them. This requires disclosure of the underlying data, and the right to correct it when inaccurate.

Signatories:

[American Civil Liberties Union](#)

[Asian Americans Advancing Justice — AAJC](#)

[Center for Media Justice](#)

[ColorOfChange](#)

[Common Cause](#)

[Free Press](#)

[The Leadership Conference on Civil and Human Rights](#)

[NAACP](#)

[National Council of La Raza](#)

[National Hispanic Media Coalition](#)

[National Urban League](#)

[NOW Foundation](#)

[New America Foundation's Open Technology Institute](#)

[Public Knowledge](#)

Objectives

- Establish a research agenda
 - Open questions (what's missing?)
- Getting our hands dirty
 - Data sets, real world projects, impact
- Active exchange between ML and policy
- Anti-BS Board
 - Technical experts responding to BS