Big Data, Machine Learning, and the Social Sciences

Hanna Wallach
Microsoft Research & University of Massachusetts Amherst
hanna@dirichlet.net
Computational Social Science

- Social Sciences
- Math & Statistics
- Computer Science

CSS
GOALS
MODELS
DATA
GOALS
MODELS
DATA
Research Goals

“Computer scientists may be interested in finding the needle in the haystack—such as […] the right web page to display from a search—but social scientists are more commonly interested in characterizing the haystack.”

— Gary King & Dan Hopkins
Prediction
Explanation
GOALS

MODELS

DATA
Predictive Modeling

\[ E[Y] = \clubsuit \times \kappa \times \infty \]

few constraints
Explanatory Modeling

\[ E[Y] = \beta \cdot X \]

interpretation
Role of Humans

replace?
guide?
GOALS
MODELS
DATA
Type of Data

- UCI Machine Learning Repository
- Wikipedia
- Enron
- Twitter
- Facebook
- Map of North Carolina
DATA
PRIVACY
FAIRNESS
ACCOUNTABILITY
Amazon just showed us that 'unbiased' algorithms can be inadvertently racist

Amazon Prime and the racist algorithms

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica
May 23, 2016

ON A SPRING AFTERNOON IN 2014, Brisha Borden was running late to pick up her god-sister from school when she spotted an unlocked kid’s blue Huffy bicycle and a silver Razor scooter. Borden
Ethical Challenges
The Path Forward