18734: Foundations of Privacy

Influence in Classification

Anupam Datta

CMU

Fall 2015

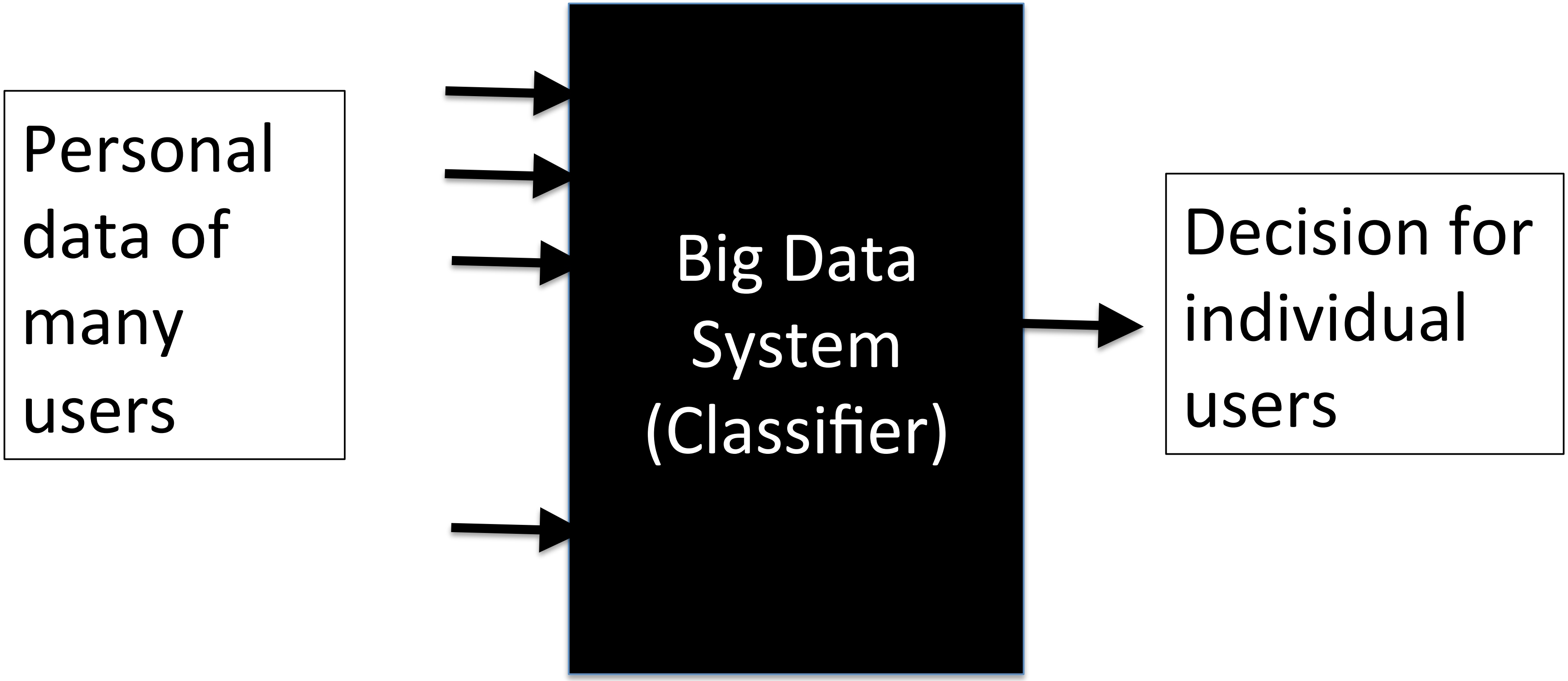
Big Data Analysis and Transparency

- Big data is big business.
- It is “good”: able to identify trends, produce accurate results.

It is not transparent!

- It is hard to tell what factors determine classification outcomes.

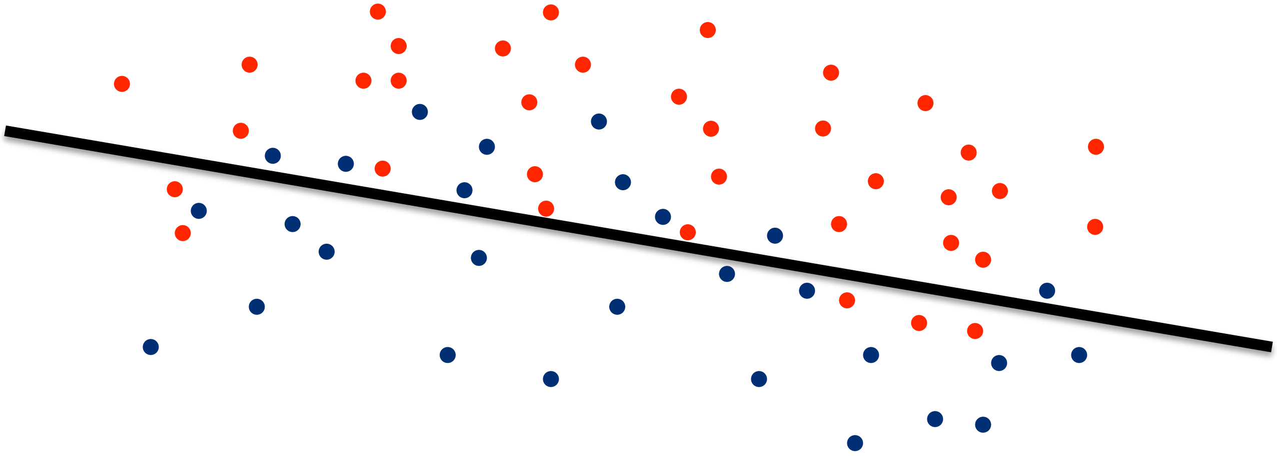
System Model


Personal data of many users
Big Data System (Classifier)
Decision for individual users

<u>Goal:</u> Measure influence of features on classifier's decision

Measuring Feature Importance

How important was the -th feature in determining the classifier's output?



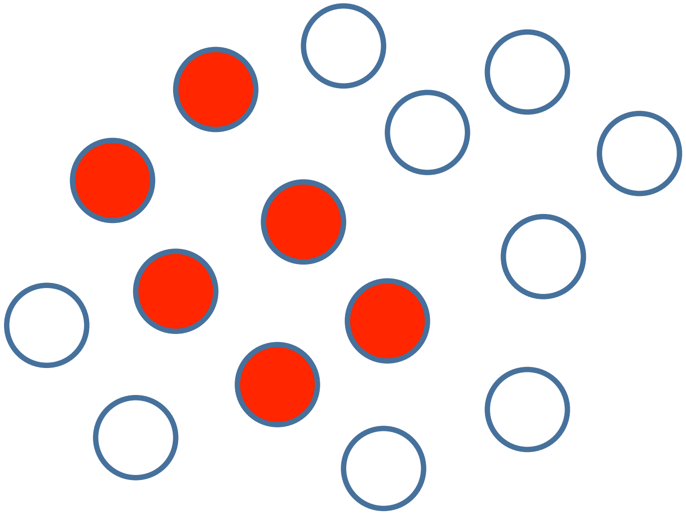
Causal Influence Measure: Idea

[Datta, Datta, Procaccia, Zick 2015]

Counts number of times a change in state causes a change in classifier's decision.

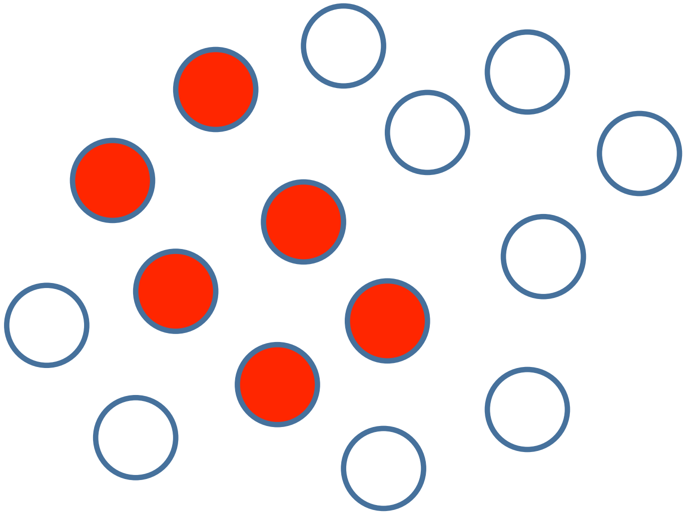
Notation

- A set of **features** $N=\{1,\ldots,n\}$
- For each $i\in N$, A_i : set of possible **states**.
- $A=\prod_{i\in N} A_i$: all possible profiles.
- $v:A\rightarrow\{0,1\}$, labels data.
- Dataset: $\langle B,v\rangle$, where $B\subseteq A$
  (we don't see all profiles)



Notation

- An influence measure: a function φ that, given a dataset $\langle B,v\rangle$, outputs a value $\varphi\downarrow i$ for every feature $i\in N$.
  “how important is gender for this classification?”



Causal Influence Measure

[Datta, Datta, Procaccia, Zick 2015]

; here:

and is a constant independent of (but may depend on).

Relation to Linear Classifiers

Theorem: suppose that is a linear classifier, defined by and . Then if and only if .

High weight translates to high influence!

Implementation

- To test our measure's behavior, we measure influence on a generated dataset.
- We employ the AdFisher framework [Datta et al. 2014] to create fake Google user profiles and observe the ads that they are presented.

Experimental Setup

- 12 x 100 simulated users, different setting of
  - Gender: male or female
  - Age: 18-24, 35-44, 55-64
  - Language: {English, Spanish}
- Go to bbc.com/news, collect the ads displayed.

Influence of Features

- *$v(a)$* for profile *a*: a vector measuring counts of for each unique ad served

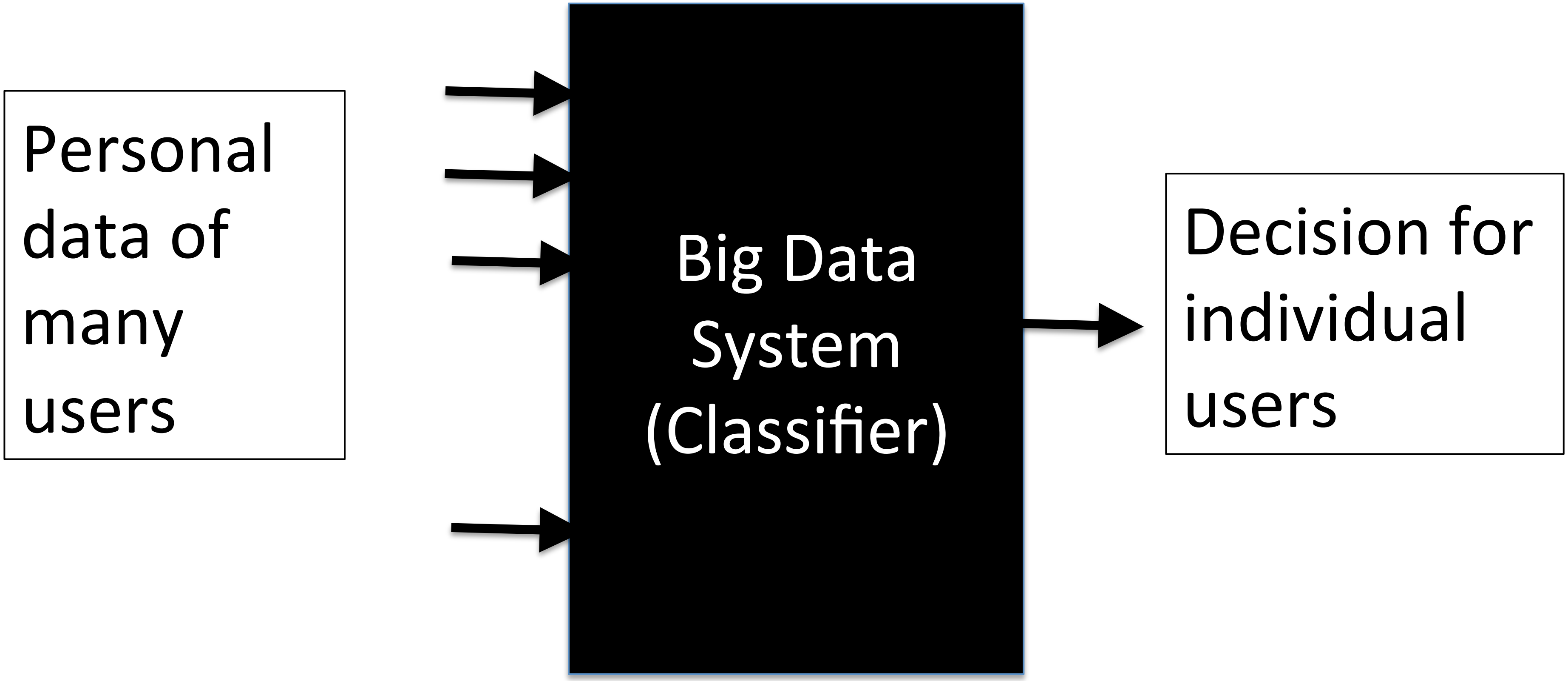
  Example: $v_{i}(M,18-24,\mathrm{En})=\#$of times i^{th} unique ad was displayed to this profile

- Influence measures for:
  - Gender: 0.124; Age: 0.120; Language: 0.141
- Language most influential but not by much

Future Work

- Leverage knowledge of priors on data
- Account for correlated features

System Model



Personal data of many users

Big Data System (Classifier)

Decision for individual users

Goal: Measure influence of features on classifier's decision

Extensions

- **State Influence:** how influential is being 25-35, vs. how important is age.

- **Generalized distance measure:** replacing / $v(\mathbf{a})-v(\mathbf{a}\downarrow -i\ ,b)$/ with a pseudo-distance $d(\mathbf{a},(\mathbf{a}\downarrow -i\ ,b))$.

Top Ads for Age

Title/Ad Description	Influence
Buy Home For Taxes Owed/Or Get 18-36% Interest! Watch 8min Video That Explains All.	0.07
Jim Rickards Project 2015/Economist, Jim Rickards explains the coming economic crash.	0.0663
”My Insomnia Trick”/Naturally Fall Asleep Fast, Stay Asleep All Night – Wake Up Refreshed	0.0661
Get In Now With Graphene/Money-Making Mineral Set To Launch Can Shape The World And Your Wealth	0.0611
Sciatica Exercises?/Stop: What You MUST know Before attempting to Treat your Sciatica:	0.0606

Top Ads for Gender

Title/Ad Description	Influence
Jim Rickards Project 2015/Economist, Jim Rickards explains the coming economic crash.	0.07
Buy Home For Taxes Owed/Or Get 18-36% Interest! Watch 8min Video That Explains All.	0.0583
Tech Gadgets/Daily Deals on Modern Gadgets. Exclusive Pricing - Up To 70% Off.	0.0564
Get In Now With Graphene/Money-Making Mineral Set To Launch Can Shape The World And Your Wealth	0.0561
Elabore su Presupuesto/Nuestros Consejeros Certificados Est´an listos para ayudarlo	0.0534

Top Ads for Language

Title/Ad Description	Influence
Elabore su Presupuesto/Nuestros Consejeros Certificados Est´an listos para ayudarlo	0.1667
The Greatest Penny Stocks/Get free daily penny stock alerts. Join now. New pick out soon.	0.0755
Business Leads CRM/Business Lead Manager, Dialer, CRM. 400% Boost in Conversion Rates.	0.0683
Get In Now With Graphene/Money-Making Mineral Set To Launch Can Shape The World And Your Wealth	0.0644
Buy Home For Taxes Owed/Or Get 18-36% Interest! Watch 8min Video That Explains All.	0.06