Privacy & Fairness in Data Science

CS848 Fall 2019
Instructor

Xi He:

• Research interest: privacy and fairness for big-data management and analysis
• CS848, Fall 2019:
  – Tue: 3:00pm - 5:50pm (DC2568)
Tell me ...

... why do you want to do this course?
Personalization ...

Based on your internet history, you might be dumb enough to enjoy extreme sports.

Click here to buy a ticket to base jump from the international space station.

I think the internet is trying to kill me.

We call it "machine learning."
Online Advertising

TOP 10: GLOBAL ADVERTISING REVENUE (IN BILLIONS)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Alphabet</td>
<td>$43.7</td>
<td>$51.1</td>
<td>$59.6</td>
<td>$67.4</td>
<td>$79.4</td>
</tr>
<tr>
<td>COMCAST</td>
<td>$11.5</td>
<td>$10.7</td>
<td>$11.8</td>
<td>$17.1</td>
<td>$26.9</td>
</tr>
<tr>
<td>CBS</td>
<td>$8.5</td>
<td>$8.8</td>
<td>$11.5</td>
<td>$11.5</td>
<td>$12.9</td>
</tr>
<tr>
<td>Disney</td>
<td>$7.8</td>
<td>$8.0</td>
<td>$8.2</td>
<td>$10.3</td>
<td>$10.4</td>
</tr>
<tr>
<td>21ST CENTURY FOX</td>
<td>$7.6</td>
<td>$7.6</td>
<td>$8.1</td>
<td>$8.6</td>
<td>$8.6</td>
</tr>
<tr>
<td>iHeartMedia</td>
<td>$6.0</td>
<td>$7.0</td>
<td>$7.8</td>
<td>$6.1</td>
<td>$7.8</td>
</tr>
<tr>
<td>Viacom</td>
<td>$4.8</td>
<td>$6.1</td>
<td>$7.7</td>
<td>$6.1</td>
<td>$7.7</td>
</tr>
<tr>
<td>Bertelsmann</td>
<td>$4.7</td>
<td>$5.1</td>
<td>$6.1</td>
<td>$5.8</td>
<td>$6.3</td>
</tr>
<tr>
<td>Time Warner</td>
<td>$4.3</td>
<td>$4.9</td>
<td>$5.0</td>
<td>$5.0</td>
<td>$6.1</td>
</tr>
<tr>
<td>Facebook</td>
<td>$4.3</td>
<td>$4.6</td>
<td>$4.6</td>
<td>$4.7</td>
<td>$6.1</td>
</tr>
</tbody>
</table>

SOURCE: Bloomberg, Zenith Media
Online Advertising

TOP 10: GLOBAL ADVERTISING REVENUE (IN BILLIONS)

Ad-Supported Internet Brings Over $1 Trillion To The U.S. Economy, Representing 6 Percent Of Country’s Total GDP, According To IAB Study Led By Harvard Business School Professor

03.15.17

SOURCE: Bloomberg, Zenith Media
Health

**Red**: official numbers from Center for Disease Control and Prevention; weekly
**Black**: based on Google search logs; daily (potentially instantaneously)

**Detecting influenza epidemics using search engine query data**

http://www.nature.com/nature/journal/v457/n7232/full/nature07634.html
IMPRECISION MEDICINE

For every person they do help (blue), the ten highest-grossing drugs in the United States fail to improve the conditions of between 3 and 24 people (red).

1. ABILIFY (aripiprazole)
   Schizophrenia

2. NEXIUM (esomeprazole)
   Heartburn

3. HUMIRA (adalimumab)
   Arthritis

4. CRESTOR (rosuvastatin)
   High cholesterol

https://www.nature.com/news/personalized-medicine-time-for-one-person-trials-1.17411
Precision Medicine

One-size-fit-all Medicine

From

Stratified Medicine

To

Precision Medicine

1. Patients are grouped by:
   - Disease Subtypes
   - Risk Profiles
   - Demographics
   - Socio-economic
   - Clinical Features
   - Biomarker
   - Molecular sub-populations

2. Individual patient level:
   - Genomics and Omics
   - Lifestyle
   - Preferences
   - Health History
   - Medical Records
   - Compliance
   - Exogenous Factors

Companion Diagnostic (CDx) Biomarker

Therapy (Rx + Dx = CDx)

Precision medicine ensures delivery of the right intervention to the right patient at the right time.

Source: forbes.com
Predictive Policing

How can predictive policing drive proactive crime prevention?

Manchester Police Department

- Protects and serves the 110,000 citizens of Manchester, New Hampshire
- Needed a smarter way to decide where its 237 officers should patrol
- Worked with Ironside to harness IBM® SPSS® Modeler software to help predict where crimes were likely to occur

- 12% reduction in robberies
- 21% reduction in burglaries
- 32% reduction in thefts from vehicles
Predictive Policing
The dark side of the force…
39% of the experts agree...

Thanks to many changes, including the building of “the Internet of Things,” human and machine analysis of **Big Data will cause more problems than it solves** by 2020. The existence of huge data sets for analysis will **engender false confidence in our predictive powers** and will lead many to make **significant and hurtful mistakes**. Moreover, analysis of Big Data will be **misused by powerful people and institutions with selfish agendas** who manipulate findings to make the case for what they want. And the advent of Big Data has a harmful impact because it **serves the majority (at times inaccurately) while diminishing the minority** and ignoring important outliers. Overall, the rise of Big Data is a big negative for society in nearly all respects.

— 2012 Pew Research Center Report
http://pewinternet.org/Reports/2012/Future-of-Big-Data/Overview.aspx
Harm due to personalized data analytics ... 

• Privacy

• Fairness
Where is the data coming from?
Where is the data coming from?

- Census surveys
- IRS Records
- Medical records
- Insurance records
- Search logs
- Browse logs
- Shopping histories

- Photos
- Videos
- Smart phone Sensors
- Mobility trajectories

Very sensitive information ...
How is this data collected?

When you visit a website...

... tiny tracking files watch what you do online...

... and develop a profile of your behavior.

Some sell your data on an exchange...

... which can combine it with other sources of personal data...

... to be sold to advertisers looking for consumers like you.

You might like this book!

You might like this car!

Advertisers buy ad space from websites at auctions.

Often, a tracking company sells this information directly to advertisers.

BACK TO YOU
The websites you visit show you ads or other content based on the description of you in the dossiers they've built and analyzed.

http://graphicsweb.wsj.com/documents/divSlider/media/ecosystem100730.png
Isn’t my data anonymous?
Device Fingerprinting

A typical computer broadcasts hundreds of details about itself when a Web browser connects to the Internet. Companies tracking people online can use those details to ‘fingerprint’ browsers and follow their users.

**Timestamp** One fingerprinting technique compares the time on a person’s computer to the time on a Web server down to the millisecond.

**User ID** Once a device has been fingerprinted, it is assigned a ‘token,’ or ID number, that can be used to track a user’s online activities.

**Device Token**: 28AB-ECDD-7A8C-3D7A-2563-AE87-C551-5D4D

**Fonts** Not all machines have the same typefaces installed. The order the fonts were installed can also distinguish one computer from another.

**Screen Size** Things like the size of the screen and its color settings can help websites display content correctly, but also can be used to identify machines.

**Browser Plugins** The mix of QuickTime, Flash and other ‘plugins’ (small pieces of optional software within a browser) can vary widely.

**User Agent** This is tech-speak for the type of Web-browsing software used. It can include specific details about the computer’s operating system, too.
Your browser fingerprint *appears to be unique* among the 2,050,572 tested in the past 45 days.

Currently, we estimate that your browser has a fingerprint that conveys at least 20.97 bits of identifying information.
Let’s get rid of unique identifiers …
The Massachusetts Governor Privacy Breach [Sweeney IJUFKS 2002]

- Name
- SSN
- Visit Date
- Diagnosis
- Procedure
- Medication
- Total Charge
- Zip
- Birth date
- Sex

Medical Data
The Massachusetts Governor Privacy Breach [Sweeney IJUFKS 2002]

Medical Data
- Name
- SSN
- Visit Date
- Diagnosis
- Procedure
- Medication
- Total Charge

Voter List
- Name
- Address
- Date Registered
- Party affiliation
- Date last voted

• Name
• SSN
• Visit Date
• Diagnosis
• Procedure
• Medication
• Total Charge

• Name
• Address
• Date Registered
• Party affiliation
• Date last voted

• Zip
• Birth date
• Sex
The Massachusetts Governor Privacy Breach [Sweeney IJUFKS 2002]

- Name
- SSN
- Visit Date
- Diagnosis
- Procedure
- Medication
- Total Charge

- Name
- Address
- Date Registered
- Party affiliation
- Date last voted

- Governor of MA uniquely identified using ZipCode, Birth Date, and Sex.

Name linked to Diagnosis

Medical Data  Voter List
The Massachusetts Governor Privacy Breach [Sweeney IJUFKS 2002]

- Name
- SSN
- Visit Date
- Diagnosis
- Procedure
- Medication
- Total Charge
- Zip
- Birth date
- Sex
- Name
- Address
- Date Registered
- Party affiliation
- Date last voted

Medical Data Voter List

87% of US population uniquely identified using ZipCode, Birth Date, and Sex.

Quasi Identifier
AOL data publishing fiasco

IN SOLIDARITY WITH THE MANY AOL USERS WHOSE OFTEN EMBARRASSING WEB SEARCHES WERE RELEASED TO THE PUBLIC, I OFFER A SAMPLE OF MY OWN SEARCH HISTORY:

```
velociraptors
site:imdb.com "jurassic park"
raptors
dromaeosaurids
utahraptor
"home depot" deadbolts
security home improvement
surviving a raptor attack
robert bakker paleontologist
robert bakker "possible raptor sympathizer"
site:en.wikipedia.org surviving a raptor attack
learning from mistakes in jurassic park
big-game rifles
tire irons
treating raptor wounds
do raptors fear fire
how to make a molotov cocktail
do raptors fear death
can raptors pick locks
how to tell if my neighbors are raptors
```
AOL data publishing fiasco ...

| Xi222    | Uefa cup                                |
| Xi222    | Uefa champions league                   |
| Xi222    | Champions league final                  |
| Xi222    | Champions league final 2013             |
| Abel156  | exchangeability                         |
| Abel156  | Proof of deFinitti’s theorem            |
| Jane12345| Zombie games                            |
| Jane12345| Warcraft                                |
| Jane12345| Beatles anthology                       |
| Jane12345| Ubuntu breeze                           |
| Bob222   | Python in thought                       |
| Bob222   | Enthought Canopy                        |
User IDs replaced with random numbers

<table>
<thead>
<tr>
<th>User ID</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>865712345</td>
<td>Uefa cup</td>
</tr>
<tr>
<td>865712345</td>
<td>Uefa champions league</td>
</tr>
<tr>
<td>865712345</td>
<td>Champions league final</td>
</tr>
<tr>
<td>865712345</td>
<td>Champions league final 2013</td>
</tr>
<tr>
<td>236712909</td>
<td>exchangeability</td>
</tr>
<tr>
<td>236712909</td>
<td>Proof of deFinitti’s theorem</td>
</tr>
<tr>
<td>112765410</td>
<td>Zombie games</td>
</tr>
<tr>
<td>112765410</td>
<td>Warcraft</td>
</tr>
<tr>
<td>112765410</td>
<td>Beatles anthology</td>
</tr>
<tr>
<td>112765410</td>
<td>Ubuntu breeze</td>
</tr>
<tr>
<td>865712345</td>
<td>Python in thought</td>
</tr>
<tr>
<td>865712345</td>
<td>Enthought Canopy</td>
</tr>
</tbody>
</table>
Privacy Breach

A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr.
Published: August 9, 2006

[NYTimes 2006]
Machine learning models can reveal sensitive information

Facebook Profile

Online Data

Number of Impressions

+ Who are interested in Men

25

+ Who are interested in Women

0

Facebook’s learning algorithm uses private information to predict match to ad

[Korolova JPC 2011]
Genome wide association studies

Results of a GWAS study

High density SNP profile of Bob

Did Bob participate in the study

[Homer et al PLOS Genetics 08]
Harm due to personalized data analytics ...

- Privacy

- Fairness
The red side of learning

- **Redlining**: the practice of denying, or charging more for, services such as banking, insurance, access to health care, or even supermarkets, or denying jobs to residents in particular, often racially determined, areas.
Predictive Policing

- Predictive policing systems use machine learning algorithms to predict crime.

- But … the algorithms learn … patterns not about crime, per se, but about how police record crime.

- This can amplify existing biases
The online world is shaped by forces beyond our control, determining the stories we read on Facebook, the people we meet on OkCupid and the search results we see on Google. Big data is used to make decisions about health care, employment, housing, education and policing.

But can computer programs be discriminatory?

There is a widespread belief that software and algorithms that rely on data are objective. But software is not free of human influence. Algorithms are written and maintained by people, and machine learning algorithms adjust what they do based on people’s behavior. As a result, say researchers in computer science, ethics and law, algorithms can reinforce human prejudices.

Google’s online advertising system, for instance, showed an ad for high-income jobs to men much more often than it showed the ad to women, a new study by Carnegie Mellon University researchers found.

Research from Harvard University found that ads for arrest records were significantly more likely to show up on searches for distinctively black names or a historically black fraternity. The Federal Trade Commission said advertisers are able to target people who live in low-income neighborhoods with high-interest loans.
BRACE YOURSELF

DEEP LEARNING IS COMING
Deep Learning

Incredibly powerful tool for ...

• Extracting regularities from data according to a given data

• Amplifying bias!
Word embeddings

Can convert words to vectors of numbers - at the hearth of most NLP applications with deep learning

http://slides.com/simonescardapane/the-dark-side-of-deep-learning
Embeddings are *highly* sexists!


http://slides.com/simonescardapane/the-dark-side-of-deep-learning
Deep Learning

Incredibly powerful tool for …

• Extracting regularities from data according to a given data

• Amplifying privacy concerns!
Given access to a black-box classifier, can we infer whether a specific example was part of the training dataset? We can with **shadow training**:


<table>
<thead>
<tr>
<th>Dataset</th>
<th>Training Accuracy</th>
<th>Testing Accuracy</th>
<th>Attack Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult</td>
<td>0.848</td>
<td>0.842</td>
<td>0.503</td>
</tr>
<tr>
<td>MNIST</td>
<td>0.984</td>
<td>0.928</td>
<td>0.517</td>
</tr>
<tr>
<td>Location</td>
<td>1.000</td>
<td>0.673</td>
<td>0.678</td>
</tr>
<tr>
<td>Purchase (2)</td>
<td>0.999</td>
<td>0.984</td>
<td>0.505</td>
</tr>
<tr>
<td>Purchase (10)</td>
<td>0.999</td>
<td>0.866</td>
<td>0.550</td>
</tr>
<tr>
<td>Purchase (20)</td>
<td>1.000</td>
<td>0.781</td>
<td>0.590</td>
</tr>
<tr>
<td>Purchase (50)</td>
<td>1.000</td>
<td>0.693</td>
<td>0.860</td>
</tr>
<tr>
<td>Purchase (100)</td>
<td>0.999</td>
<td>0.659</td>
<td>0.935</td>
</tr>
<tr>
<td>TX hospital stays</td>
<td>0.668</td>
<td>0.517</td>
<td>0.657</td>
</tr>
</tbody>
</table>

**TABLE II:** Accuracy of the Google-trained models and the corresponding attack precision.
This course:

Learn to combat the dark side
You will ...

• mathematically formulate privacy.
• mathematically formulate fairness.
Differential Privacy

For every pair of inputs that differ in one row:

\[ D_1 \quad D_2 \]

Adversary should not be able to distinguish between any \( D_1 \) and \( D_2 \) based on any \( O \)

\[ \log \left( \frac{\Pr[A(D_1) = O]}{\Pr[A(D_2) = O]} \right) < \varepsilon \quad (\varepsilon > 0) \]
You will …

• mathematically formulate privacy.
• mathematically formulate fairness.

• design algorithms to ensure privacy
• design algorithms to ensure fairness
Differential Privacy in practice

OnTheMap [ICDE 2008]

[CCS 2014]

[Apple WWDC 2016]
You will ...

- mathematically formulate privacy.
- mathematically formulate fairness.
- design algorithms to ensure privacy
- design algorithms to ensure fairness
- do research into the interplay between privacy and fairness.
Course Format

• Module 1: Intro to Privacy
  - In-class Exercise
  - In-class Mini-project
  - Lectures

• Module 2: Intro to Fairness
  - Reading papers
  - Mini-critiques
  - Research Project

• Module 3: Paper Reading by Topics
  - privacy v.s. fairness
  - private machine learning
  - deployments of DP
  - sources of bias
  - fairness mechanisms
\[ \forall i \in [n], d \in S, \ln \frac{\Pr[T_i \in T | d_i = d]}{\Pr[T_i \in T | d_i = \text{NULL}]} \leq \ln \left( \frac{e^e \cdot \frac{1+e^e}{1}}{1+e^e} \right) = \varepsilon \]

\[ \frac{\alpha_{\text{client}}(d) = t}{\alpha_{\text{client}}(\text{null}) = t} \leq \ln \left( \frac{e^e \cdot \frac{1+e^e}{1}}{1+e^e} \right) = \varepsilon \]

\[ \alpha = \frac{3k + 2c_\varepsilon \sqrt{\ln(6mk/\beta)}}{\sqrt{n}} = O \left( \frac{\sqrt{\log(p/\beta)}}{\varepsilon \sqrt{n}} \right) \]

\[ \left\{ \left( \frac{v[j] \cdot b[j] + 1}{2} \right), \forall j \in [m] \right\} \]
What we expect you to know ... 

• Strong background in
  – Probability
  – Proof techniques

• Some knowledge of
  – Programming with Python
  – Machine learning
  – Statistics
  – Algorithms
Misc. course info

• **Website**: [https://cs.uwaterloo.ca/~xihe/cs848](https://cs.uwaterloo.ca/~xihe/cs848)
  – Schedule (with links to lecture slides, readings, projects, etc.)

• **Grading**
  – In class mini-projects: 10% x 2
  – Mini-critiques: 10%
  – Class participation and presentation: 20%
    • Attending class!
  – Project: 50%

• **LEARN** for submission and grades:
  – [https://learn.uwaterloo.ca/d2l/home/492027](https://learn.uwaterloo.ca/d2l/home/492027)
Academic Integrity

• See course website
• Mini-project reports and paper critiques are individual work and submission.
• Group discussion okay (and encouraged), but
  – Acknowledge help you receive from others
  – Make sure you “own” your solution
• All suspected cases of violation will be aggressively pursued
Reference

• Course materials are adapted from: https://sites.duke.edu/cs590f18privacyfairness/