

Machine Learning for Data Science (CS4786)

Lecture 26

Fairness, Transparency and other Moral Issues in Machine Learning







Course Webpage :

<http://www.cs.cornell.edu/Courses/cs4786/2016fa/>

Announcements

- Survey 2, just over 80%
- Make sure you fill out the course eval
- (If the participation on this is above 90% I will still drop worst assignment)

ML DREAM



[All](#) [Shopping](#) [Maps](#) [Images](#) [News](#) [More](#) [Settings](#) [Tools](#)

About 1,810,000,000 results (1.10 seconds)

Shoes at Zappos.com
Ad www.zappos.com/Shoes
4.7 ★★★★★ rating for zappos.com
Fast, Free Shipping & Free 365 Day Returns on Huge Selection of Shoes!
Birkenstock · Nike · Converse · New Balance · Born · Frye
Types: Sneakers, Slippers, Heels, Boots, Flats, Running Shoes


Womens Shoes at Macy's - Save 40-60% on Cyber Monday - macys.com
Ad www.macys.com/Womens_Shoes/Holiday_Deals
4.2 ★★★★★ rating for macys.com
Cyber Monday Ends 11/30, Deals Going Fast + Extra 20% Off!
Styles: Boots, Wedges, Loafers, Flats, Slippers, Sneakers, Pumps, Espadrilles, Booties
Ratings: Returns 9/10 - Product quality 9/10 - Shipping 8.5/10 - Service 8.5/10 - Selection 8.5/10

Converse® Official Site - Converse.com
Ad www.converse.com/Shoes
Full Converse Collection. Shop Our New Designs for Men, Women & Kids.
Free Shipping for Members · Free 60 Day Returns
Types: Men's Sneakers, Women's Sneakers, Kid's Sneakers, Custom Sneakers

Brand Name Shoes for Less - Up to 70% Off - zulily.com
Ad www.zulily.com/
Exclusive Deals On All Shoes, Socks, & More at Zulily. Shop & Save Today!

Men and Womens shoes, Shipped Free | Zappos.com
www.zappos.com/shoes
Boots Sneakers & Athletic Heels Flats Sandals View All... Sneakers & Athletic Boots Oxfords Loafers Sandals View All... Sneakers and Athletic Boots Slippers Flats Sandals view all...
Sandals · Womens Shoes · Sneakers & Athletic Shoes · Popular Mens Shoe Styles

Shop on Google Sponsored ⓘ



Adidas NEO Baseline Women's Dicast-Leather Sneakers, Size: 8, White
\$34.99 - Kohl's

ML DREAM

What product you would like...

The image shows a Google search interface for the query "shoes". The search bar at the top contains the word "shoes" and shows a search count of "About 1,810,000,000 results (1.10 seconds)". Below the search bar are navigation tabs for "All", "Shopping", "Maps", "Images", "News", "More", "Settings", and "Tools".

The search results are divided into two main sections. The left section, highlighted with a red border, contains four organic search results:

- Shoes at Zappos.com**: An advertisement for Zappos.com with a 4.7-star rating. It highlights "Fast, Free Shipping & Free 365 Day Returns on Huge Selection of Shoes!" and lists brands like Birkenstock, Nike, Converse, New Balance, Born, and Frye. Types include Sneakers, Slippers, Heels, Boots, Flats, and Running Shoes.
- Womens Shoes at Macy's - Save 40-60% on Cyber Monday - macys.com**: An advertisement for Macy's with a 4.2-star rating. It states "Cyber Monday Ends 11/30, Deals Going Fast + Extra 20% Off!" and lists styles like Boots, Wedges, Loafers, Flats, Slippers, Sneakers, Pumps, Espadrilles, and Booties. Ratings for Returns (9/10), Product quality (9/10), Shipping (8.5/10), Service (8.5/10), and Selection (8.5/10) are provided.
- Converse® Official Site - Converse.com**: An advertisement for Converse.com with a 4.2-star rating. It promotes a "Full Converse Collection" and "Free Shipping for Members · Free 60 Day Returns". Types include Men's Sneakers, Women's Sneakers, Kid's Sneakers, and Custom Sneakers.
- Brand Name Shoes for Less - Up to 70% Off - zulily.com**: An advertisement for Zulily.com with a 4.2-star rating. It offers "Exclusive Deals On All Shoes, Socks, & More at Zulily. Shop & Save Today!".

The right section, also highlighted with a red border, features a sponsored product listing:

- Shop on Google**: A sponsored result for Adidas NEO Baseline Women's Dicast-Leather Sneakers, Size: 8, White, priced at \$34.99 from Kohl's. The listing includes a small image of the shoe and a "Sponsored" label.

At the bottom of the page, there is a link for "Men and Womens shoes, Shipped Free | Zappos.com" with a URL to www.zappos.com/shoes. Below this link are several category links: "Boots Sneakers & Athletic Heels Flats Sandals View All...", "Sneakers & Athletic Boots Oxfords Loafers Sandals View All...", "Sneakers and Athletic Boots Slippers Flats Sandals view all...", and "Sandals · Womens Shoes · Sneakers & Athletic Shoes · Popular Mens Shoe Styles".

ML DREAM

What news would you prefer to read...



ML DREAM

Find the best job for you...

The screenshot shows the LinkedIn job search interface. At the top, the LinkedIn logo is on the left, and navigation links for 'What is LinkedIn?', 'Join Today', and 'Sign In' are on the right. Below this is a search bar with 'Data Scientist' entered, a location filter for 'United States', and a 'Find jobs' button. On the left side, there are three filter panels: 'Get alerts for this search' with an email input and 'Create job alert' button; 'Location' with checkboxes for New York, San Francisco, Seattle, Chicago, and Atlanta; and 'Company' with checkboxes for Maverick Trading, Amazon, Deloitte, CyberCoders, and Jobspring Partners, plus an '+ Add' button. The main content area shows 8,715 jobs in the United States, sorted by Relevance. Four job listings are visible:

- Data Scientist** at **LeadGenius**, San Francisco Bay Area, posted 13d. Description: "We are looking for a seasoned Data Scientist/Machine Learning engineer to build the next generation mission critical data platform. Solid engineering and coding skills." Includes an "Apply with Profile" button.
- Data Scientist** at **Feedzai**, Atlanta, Georgia, posted 8d. Description: "...client's data feeds Work with the the client to explore their data and better understand it Work..."
- Data Scientist** at **Jetlore**, Sunnyvale, California, posted 17d. Description: "We are looking for an exceptional data scientist who is excited to work on challenging problems Involving massive amount of data. Ping-pong skills is a plus!"
- Data Scientist** at **Covestro**, Greater Pittsburgh Area, posted 5d. Description: "Covestro is in search of a Data Scientist... and data analysis to help influence changes..."

ML DREAM

- For every user predict: Ads, products, news, ...
- Have tons of data to learn this task well
- Have right models that can learn from all this data

ML DREAM

- For every user predict: Ads, products, news, ...
- Have tons of data to learn this task well
- Have right models that can learn from all this data

With Big Data comes Bigger Responsibilities ...

IS ML FAIR, IMPARTIAL?

Google

Women less likely to be shown ads for high-paid jobs on Google, study shows

Automated testing and analysis of company's advertising system reveals male job seekers are shown far more adverts for high-paying executive jobs

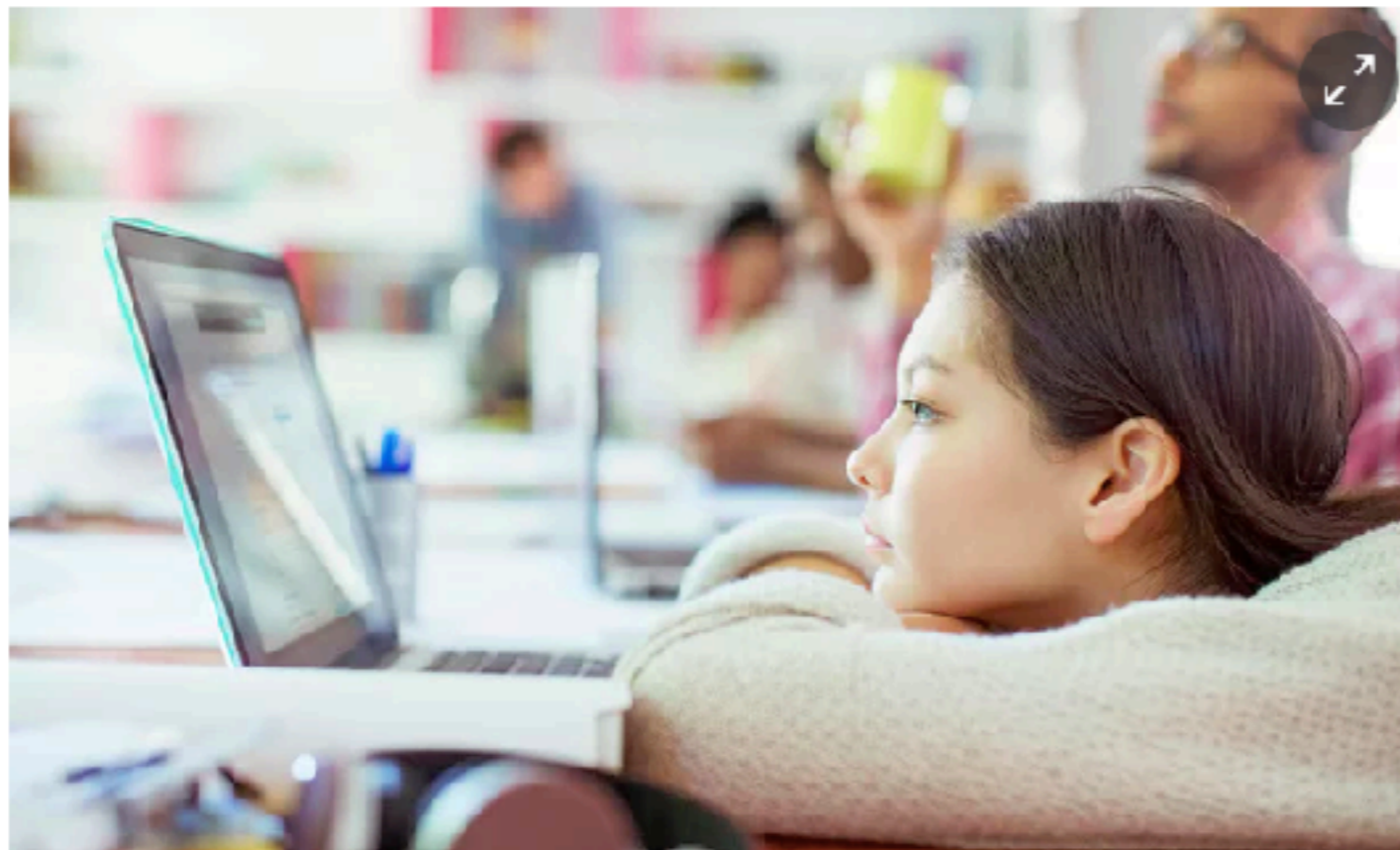
Samuel Gibbs

Wednesday 8 July 2015
06:29 EDT



This article is 1 year old

1120 140



One experiment showed that Google displayed adverts for a career coaching service for executive jobs 1,857 times to the male group and only 318 times to the female group. Photograph: Alamy

Female job seekers are much less likely to be shown adverts on [Google](#) for highly paid jobs than men, researchers have found.

Advertisement



In the spirit of Giving Tuesday, we're donating our ads to charity today

IS ML FAIR, IMPARTIAL?

Prediction Fails Differently for Black Defendants

| | WHITE | AFRICAN AMERICAN |
|---|-------|------------------|
| Labeled Higher Risk, But Didn't Re-Offend | 23.5% | 44.9% |
| Labeled Lower Risk, Yet Did Re-Offend | 47.7% | 28.0% |

Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes. (Source: ProPublica analysis of data from Broward County, Fla.)

WHY NOW?

WHY NOW?



Loads of data collected everywhere!

WHY NOW?



Machine Learning

galvanize

 Startup.ML

coursera



datascience@berkeley

 The Data Incubator

METIS

WHY NOW?



Machine Learning

galvanize

 Startup.ML

coursera



datascience@berkeley

 The Data Incubator

METIS

Raise in number of Data Scientists!

WHY IS ML UNFAIR?

the algorithms in themselves are neutral. “This program had absolutely nothing to do with race... but multi-variable equations,”

WHY IS ML UNFAIR?

- Data collection, labeling etc. can have unintentional biases
 - We learn from past data, historic biases
- Data in itself nor algorithms explicitly know of social inequities

FAIRNESS THROUGH BLINDNESS?

- Ignore all protected attributes.
Eg. Don't look at race, gender etc.

FAIRNESS THROUGH BLINDNESS?

- Ignore all protected attributes.
Eg. Don't look at race, gender etc.
- **Problem: You don't need to look to be able to predict**

FAIRNESS THROUGH BLINDNESS?

- Ignore all protected attributes.
Eg. Don't look at race, gender etc.
- **Problem: You don't need to look to be able to predict**

Eg. User visits “www.artofmanliness.com”
...highly likely to be male

EG. REAL VS FAKE NAMES

- Biases are often not intentional ...

EG. REAL VS FAKE NAMES



- Biases are often not intentional ...
- Most training examples standard white American names: James, John, Robert, Jennifer, Michael, ...
- Ethnic names often unique, much fewer training examples

EG. REAL VS FAKE NAMES

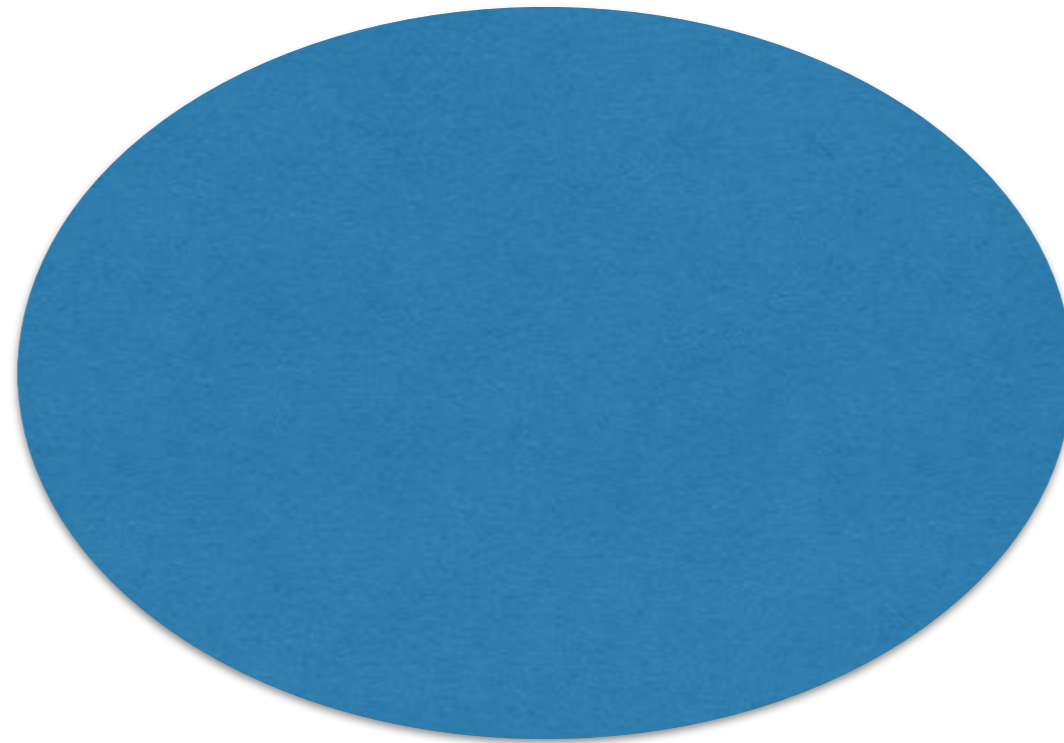


- Biases are often not intentional ...
- Most training examples standard white American names: James, John, Robert, Jennifer, Michael, ...
- Ethnic names often unique, much fewer training examples

Most ML models aim for accuracy for the majority at the expense of mistakes on the smaller protected class

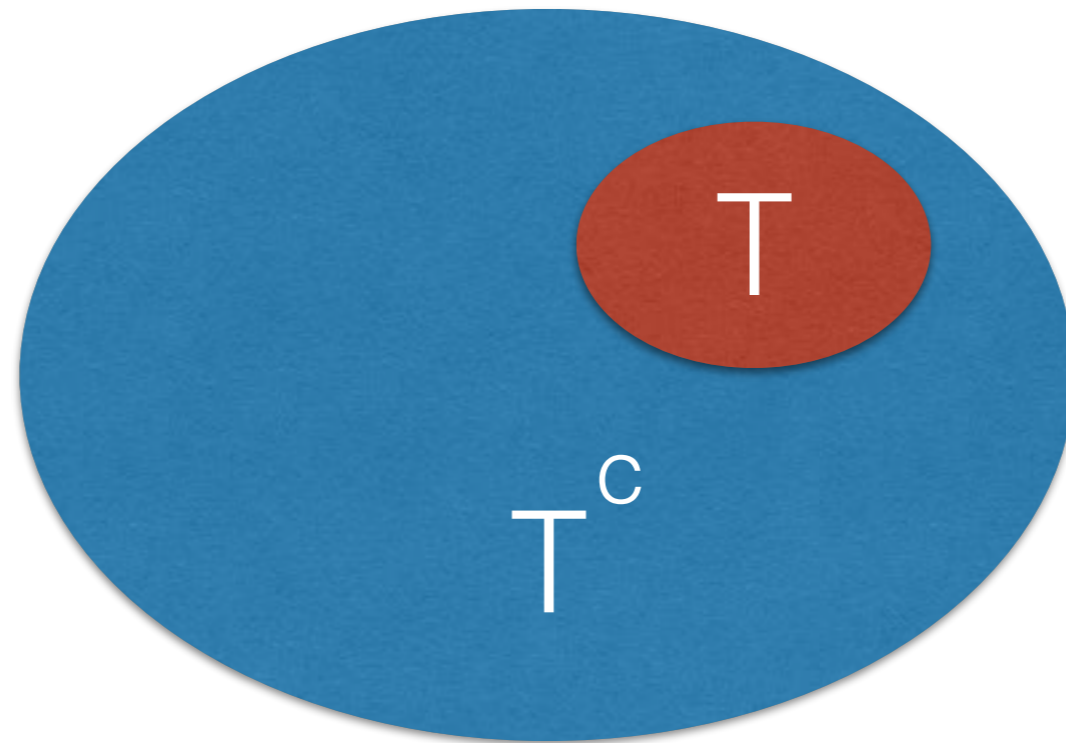
FAIRNESS THROUGH AWARENESS

FAIRNESS THROUGH AWARENESS



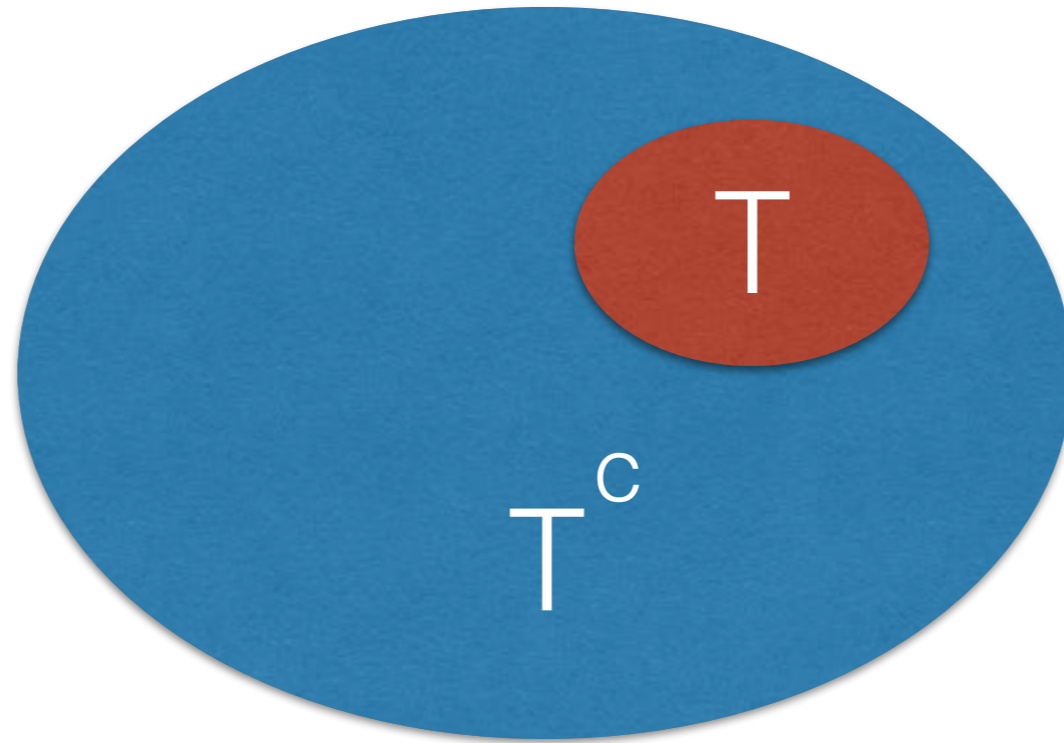
Population

FAIRNESS THROUGH AWARENESS



Population

FAIRNESS THROUGH AWARENESS

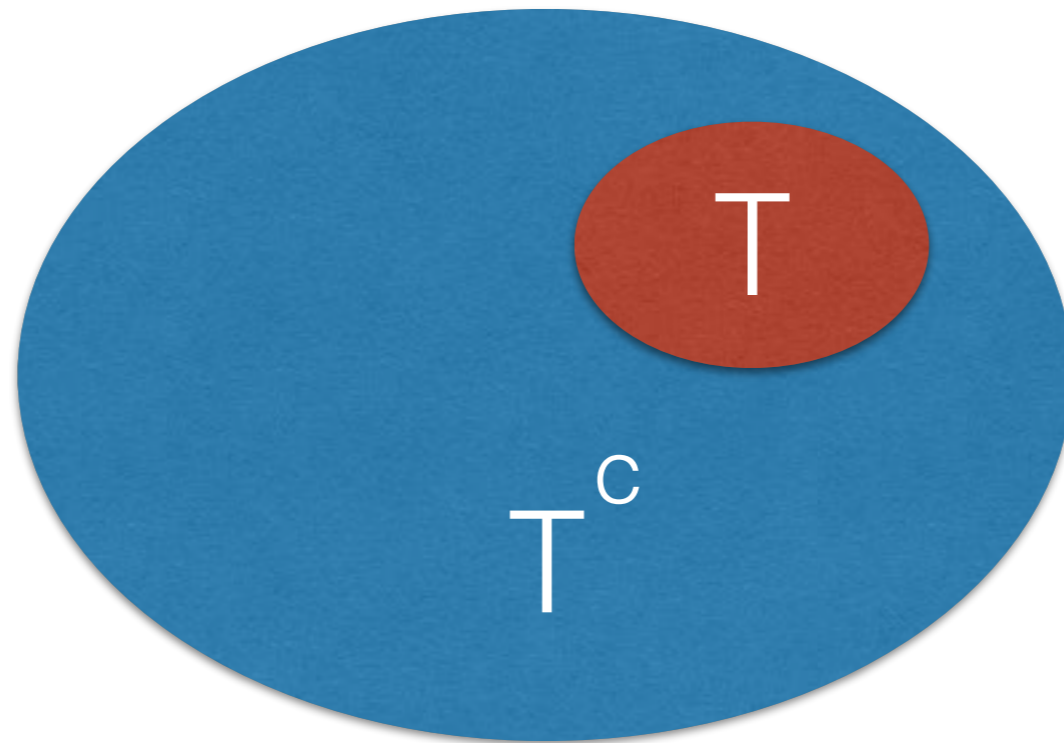


Population

T : Protected subset

T^c : Rest of the population

FAIRNESS THROUGH AWARENESS



Population

T : Protected subset

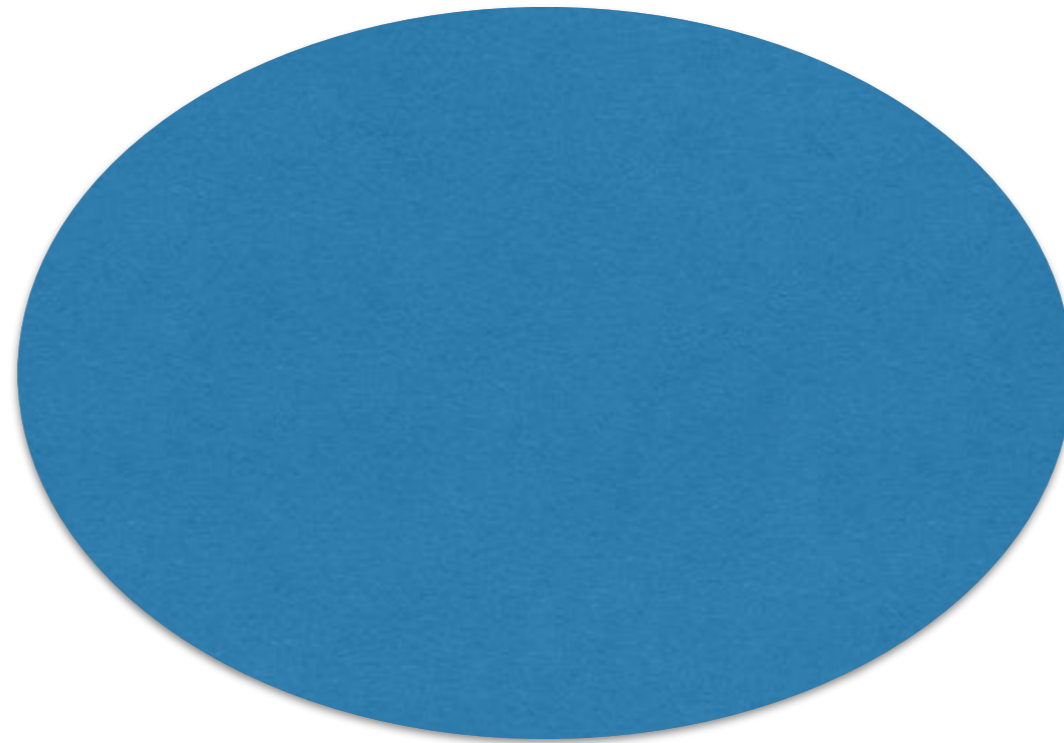
T^c : Rest of the population

$$P(\text{Outcome}|T) \approx P(\text{Outcome}|T^c)$$

FAIRNESS THROUGH AWARENESS

Eg. Fraction of people shown high paying jobs in T and in T^c is equal

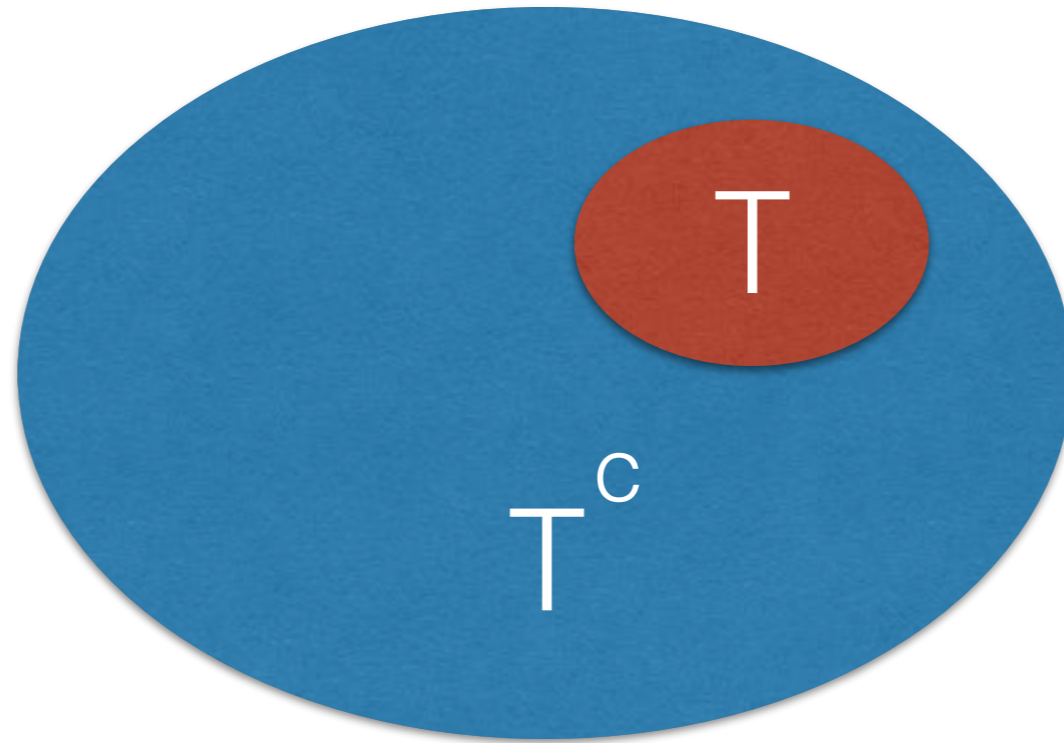
FAIRNESS THROUGH AWARENESS



Population

Eg. Fraction of people shown high paying jobs in T and in T^c is equal

FAIRNESS THROUGH AWARENESS



Population

Eg. Fraction of people shown high paying jobs in T and in T^c is equal

EG. FAIR K-MEANS CLUSTERING (very naive)

EG. FAIR K-MEANS CLUSTERING (very naive)

$$\text{Objective} = \sum_{j=1}^K \sum_{t \in C_j} \|\mathbf{x}_t - \mathbf{r}_j\|_2^2$$

$$\text{where } \mathbf{r}_j = \frac{1}{|C_j|} \sum_{\mathbf{x}_t \in C_j} \mathbf{x}_t$$

EG. FAIR K-MEANS CLUSTERING (very naive)

$$\text{Objective} = \sum_{j=1}^K \sum_{t \in C_j} \|\mathbf{x}_t - \mathbf{r}_j\|_2^2$$

$$\text{where } \mathbf{r}_j = \frac{1}{|C_j|} \sum_{\mathbf{x}_t \in C_j} \mathbf{x}_t$$

$$\text{Fairness constraints: } \forall j \in [K], \quad \sum_{t: c_t=j} \mathbf{1}_{x_t \in T} = \sum_{t: c_t=j} \mathbf{1}_{x_t \notin T}$$

EG. FAIR K-MEANS CLUSTERING (very naive)

$$\text{Objective} = \sum_{j=1}^K \sum_{t \in C_j} \|\mathbf{x}_t - \mathbf{r}_j\|_2^2$$

$$\text{where } \mathbf{r}_j = \frac{1}{|C_j|} \sum_{\mathbf{x}_t \in C_j} \mathbf{x}_t$$

$$\text{Fairness constraints: } \forall j \in [K], \quad \sum_{t: c_t=j} \mathbf{1}_{x_t \in T} = \sum_{t: c_t=j} \mathbf{1}_{x_t \notin T}$$

Number of protected in cluster j = Number of unprotected in cluster j

FAIR CLASSIFICATION

A view from a mile above:

FAIR CLASSIFICATION

A view from a mile above:

Minimize Classification objective
(or whatever other surrogate loss you use usually)

FAIR CLASSIFICATION

A view from a mile above:

Minimize Classification objective
(or whatever other surrogate loss you use usually)

Added Constraint: subject to proportion of labels in each class being same for protected and unprotected population

FAIR CLASSIFICATION

FAIR CLASSIFICATION

- Is this good enough?

FAIR CLASSIFICATION

- Is this good enough?
- Say there is this algorithm to select people to invite to apply for this **exclusive, credit card with high annual fee**

FAIR CLASSIFICATION

- Is this good enough?
- Say there is this algorithm to select people to invite to apply for this **exclusive, credit card with high annual fee**
- One simple way to satisfy the fairness constraint:

FAIR CLASSIFICATION

- Is this good enough?
- Say there is this algorithm to select people to invite to apply for this **exclusive, credit card with high annual fee**
- One simple way to satisfy the fairness constraint:
 - Make offer to higher income people in the unprotected class

FAIR CLASSIFICATION

- Is this good enough?
- Say there is this algorithm to select people to invite to apply for this **exclusive, credit card with high annual fee**
- One simple way to satisfy the fairness constraint:
 - Make offer to higher income people in the unprotected class
 - Make offer to lower income people in protected class (in same proportion)

FAIR CLASSIFICATION

- Is this good enough?
- Say there is this algorithm to select people to invite to apply for this **exclusive, credit card with high annual fee**
- One simple way to satisfy the fairness constraint:
 - Make offer to higher income people in the unprotected class
 - Make offer to lower income people in protected class (in same proportion)

NOT REALLY FAIR!

FAIR CLASSIFICATION

A view from a mile above:

Minimize Classification objective
(or whatever other surrogate loss you use usually)

FAIR CLASSIFICATION

A view from a mile above:

Minimize Classification objective
(or whatever other surrogate loss you use usually)

Added Constraint: subject to proportion of labels for
similar instances
in each class being same for protected
and unprotected population

Its not always about fairness
There are other issues too!

ML DREAM

What news would user prefer to read...



ML DREAM

ML DREAM

- Just because we can predict, should we?

ML DREAM

- Just because we can predict, should we?
- Say we have a fair, unbiased algorithm for prediction

ML DREAM

- Just because we can predict, should we?
- Say we have a fair, unbiased algorithm for prediction
- Can there be other issues?

EXTREMIZING EFFECT OF ML

EXTREMIZING EFFECT OF ML



EXTREMIZING EFFECT OF ML



User 1

User 2

EXTREMIZING EFFECT OF ML



User 1

Apples are extremely rich in important antioxidants, flavanoids, and dietary fiber. The phytonutrients and antioxidants in **apples** may help reduce the risk of developing cancer, hypertension, diabetes, and heart



User 2

For fewer calories per fruit, **oranges** have higher levels of Vitamin C, folate, potassium, and protein.

EXTREMIZING EFFECT OF ML



User 1

Apples are extremely rich in important antioxidants, flavanoids, and dietary fiber. The phytonutrients and antioxidants in **apples** may help reduce the risk of developing cancer, hypertension, diabetes, and heart



User 2

For fewer calories per fruit, **oranges** have higher levels of Vitamin C, folate, potassium, and protein.

The Health Benefits of Oranges

- 
- Packed with fiber to promote healthy digestion
 - Full of folate to help the body form red blood cells
 - A good source of immune-boosting vitamin C
 - Contains potassium to ensure a healthy heart
 - Keeps vision clear and eyes healthy with its content of vitamin A
 - A great source of calcium for healthy and strong bones

EXTREMIZING EFFECT OF ML



User 1

Apples are extremely rich in important antioxidants, flavanoids, and dietary fiber. The phytonutrients and antioxidants in **apples** may help reduce the risk of developing cancer, hypertension, diabetes, and heart

TOP 10 Health Benefits of Apples

- 1. Cancer Prevention
- 2. Antioxidant Activity
- 3. Antihyperglycemic
- 4. Anti-diabetes
- 5. Cardiovascular Protection
- 6. Cholesterol Reduction
- 7. Anti-asthma
- 8. Weight Reduction



User 2

For fewer calories per fruit, **oranges** have higher levels of Vitamin C, folate, potassium, and protein.

The Health Benefits of Oranges

- Packed with fiber to promote healthy digestion
- Full of folate to help the body form red blood cells
- A good source of immune-boosting vitamin C
- Contains potassium to ensure a healthy heart
- Keeps vision clear and eyes healthy with its content of vitamin A
- A great source of calcium for healthy and strong bones



EXTREMIZING EFFECT OF ML



Vs



User 1

Apples are extremely rich in important antioxidants, flavanoids, and dietary fiber. The phytonutrients and antioxidants in **apples** may help reduce the risk of developing cancer, hypertension, diabetes, and heart



User 2

For fewer calories per fruit, **oranges** have higher levels of Vitamin C, folate, potassium, and protein.

The Health Benefits of Oranges

- Packed with fiber to promote healthy digestion
- Full of folate to help the body form red blood cells
- A good source of immune-boosting vitamin C
- Contains potassium to ensure a healthy heart
- Keeps vision clear and eyes healthy with its content of vitamin A
- A great source of calcium for healthy and strong bones



EXTREMIZING EFFECT OF ML

EXTREMIZING EFFECT OF ML

- Possible fixes:

EXTREMIZING EFFECT OF ML

- Possible fixes:
 - Good contents is the best fix

EXTREMIZING EFFECT OF ML

- Possible fixes:
 - Good contents is the best fix
 - Mix up articles: provide both sides

EXTREMIZING EFFECT OF ML

- Possible fixes:
 - Good contents is the best fix
 - Mix up articles: provide both sides
 - what if there are multiple sides and worse yet you can't identify the sides

EXTREMIZING EFFECT OF ML

- Possible fixes:
 - Good contents is the best fix
 - Mix up articles: provide both sides
 - what if there are multiple sides and worse yet you can't identify the sides
 - Mix up user profiles from time to time (or have a canonical user and every user is mix of individual and the canonical user)

EXTREMIZING EFFECT OF ML

- Possible fixes:
 - Good contents is the best fix
 - Mix up articles: provide both sides
 - what if there are multiple sides and worse yet you can't identify the sides
 - Mix up user profiles from time to time (or have a canonical user and every user is mix of individual and the canonical user)
- This is a completely open topic ...

EXTREMIZING EFFECT OF ML

- Possible fixes:
 - Good contents is the best fix
 - Mix up articles: provide both sides
 - what if there are multiple sides and worse yet you can't identify the sides
 - Mix up user profiles from time to time (or have a canonical user and every user is mix of individual and the canonical user)
- This is a completely open topic ...
 - But a very important one ...

TRANSPARENCY IN ML

- Another issue: ML methods are complex and we don't understand semantic meaning

TRANSPARENCY IN ML

- Another issue: ML methods are complex and we don't understand semantic meaning
- We need transparency of method for accountability

TRANSPARENCY IN ML

- Another issue: ML methods are complex and we don't understand semantic meaning
- We need transparency of method for accountability
- Transparency via interpretability.
 - Provide explanation for each decision
 - What makes an instance a negative instance according to the algorithm

November 18th 2016 / New York University, NYC

Fairness, Accountability, and Transparency in Machine Learning

Co-located with the [Data Transparency Lab Conference](#) and
the [Workshop on Data and Algorithmic Transparency](#)

The workshop is now over but a recording of the event will be
available later.

Sponsors



mozilla



DATA
TRANSPARENCY
LAB



NYU Technology Law
and Policy Clinic

Telefonica



Time for General Questions