







Lecture 25: Fairness in ML

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 · Women's Shoes · Sneakers & Athletic Shoes · Popular Men's 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 would you like...



[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!

Dirksenstock · 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 · Women's Shoes · Sneakers & Athletic Shoes · Popular Men's Shoe Styles

Shop on Google

Sponsored ⓘ



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

ML DREAM

What news would you prefer to read...



ML DREAM

Find the best job for you...

Linked in

What is LinkedIn? Join Today Sign In

Data Scientist

United States

Find jobs

Get alerts for this search

We'll email you new jobs as they become available

Email address

Create job alert

Location

☐ New York, New York (621)

☐ San Francisco, California (575)

☒ Seattle, Washington (318)

☐ Chicago, Illinois (291)

☐ Atlanta, Georgia (181)

Company

☐ Maverick Trading (339)

☒ Amazon (265)

☐ Deloitte (235)

☐ CyberCoders (165)


☐ Jobspring Partners (145)

+ Add

Date Posted

8,715 Data Scientist jobs in United States

sort by: Relevance



Data Scientist


LeadGenius

San Francisco Bay Area

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.

Apply with Profile

13d




Data Scientist

Feedzai

Atlanta, Georgia

...client's data feeds Work with the the client to explore their data and better understand it Work...

8d




Data Scientist

Jetlore

Sunnyvale, California

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!

17d



Data Scientist

Covestro

Greater Pittsburgh Area

Covestro is in search of a Data Scientist... and data analysis to help influence changes...

5d

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

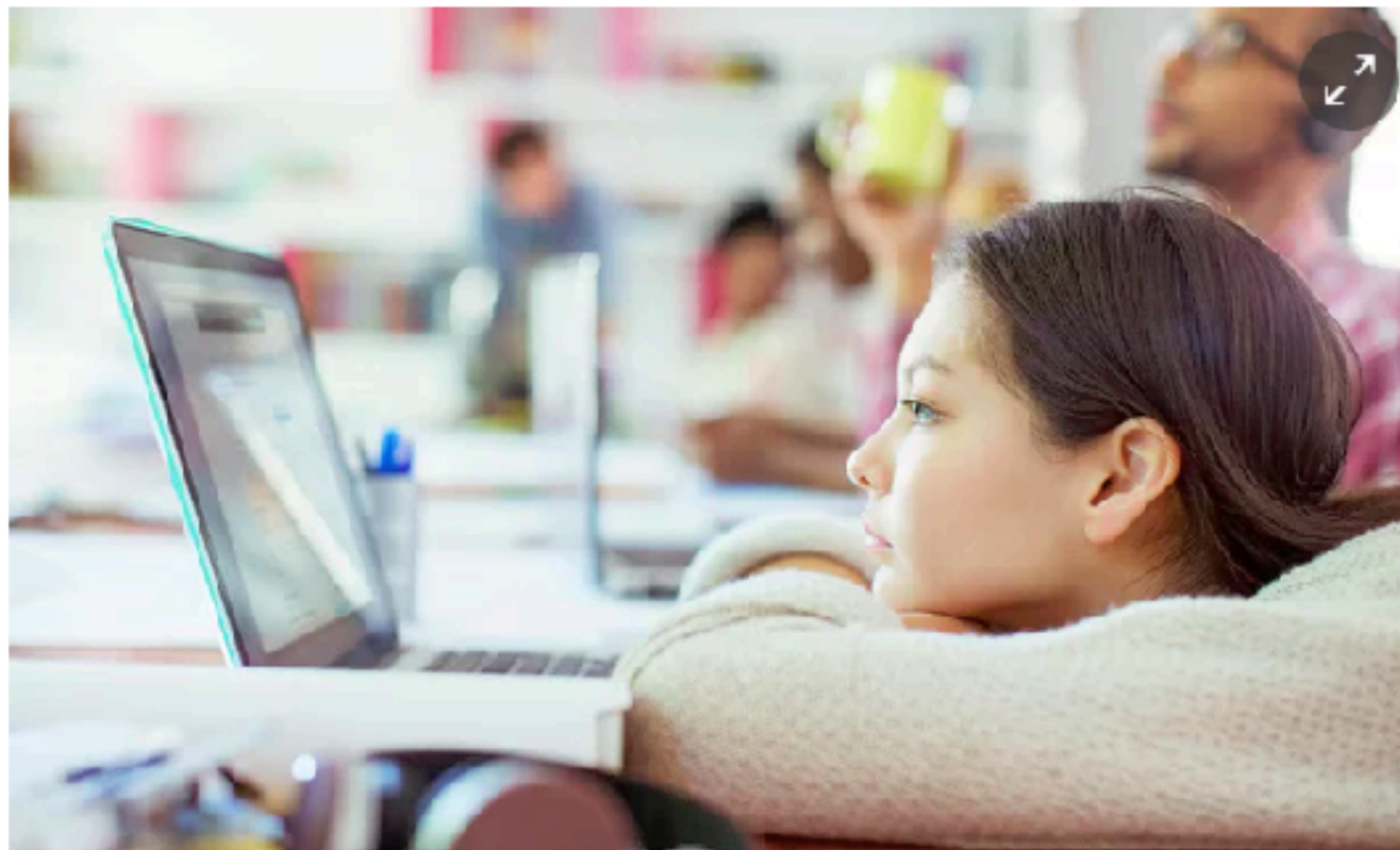
Samuel Gibbs

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 1,852 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.)

Can we make ML Fair?

- These are machine learning algorithms that learn to predict automatically
- They are not designed to be unfair
- Why is this happening?
- How do we fix them?

WHY NOW?

WHY NOW?



Loads of data collected everywhere!

WHY NOW?



Machine Learning

galvanize

 Startup.ML

coursera



METIS

datascience@berkeley

 The Data Incubator

WHY NOW?



Machine Learning

galvanize

 Startup.ML

coursera



METIS

datascience@berkeley

 The Data Incubator

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

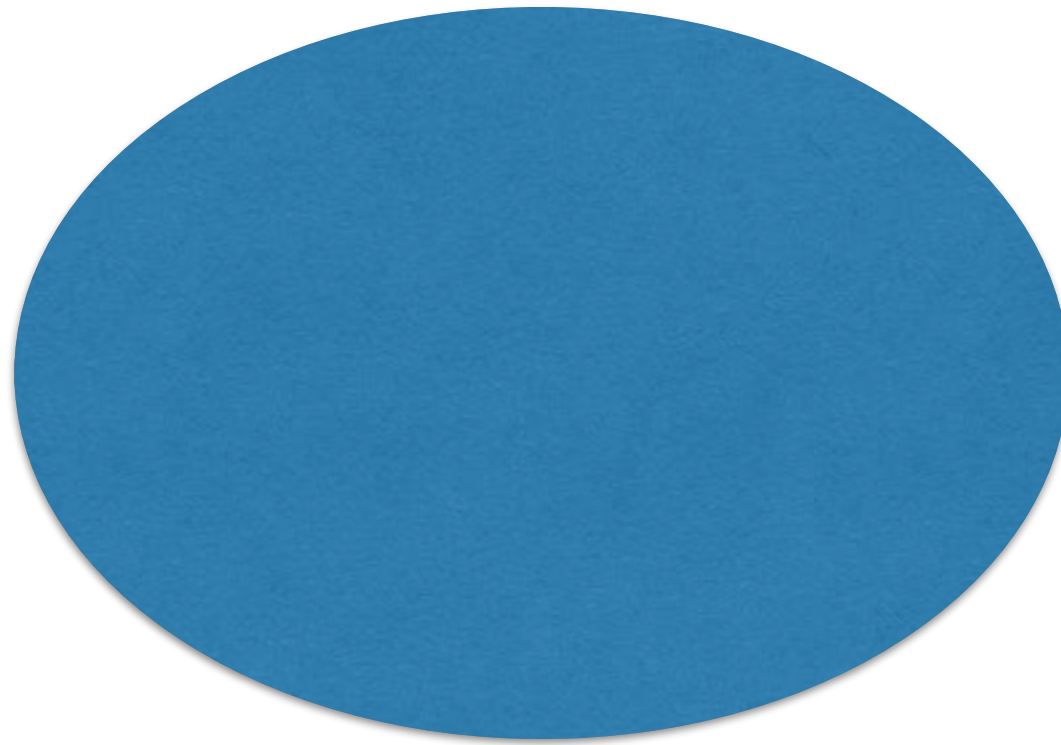
Most ML models aim for accuracy for the majority at the expense of mistakes on the smaller protected class

FAIRNESS THROUGH AWARENESS

Demographic Parity

FAIRNESS THROUGH AWARENESS

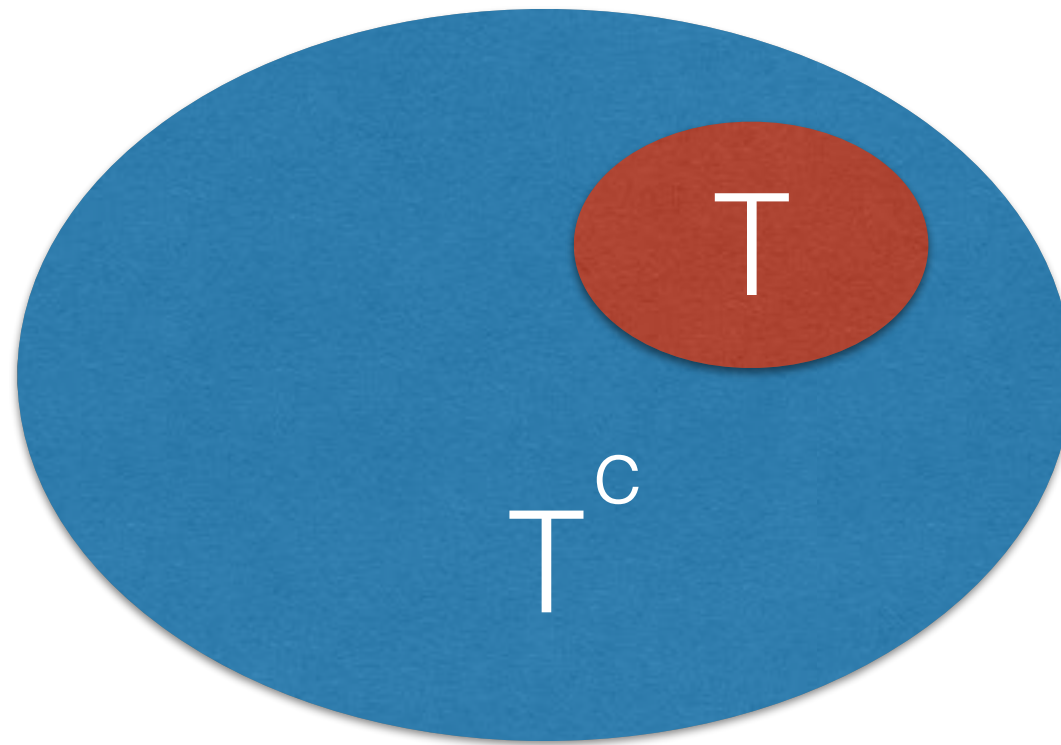
Demographic Parity



Population

FAIRNESS THROUGH AWARENESS

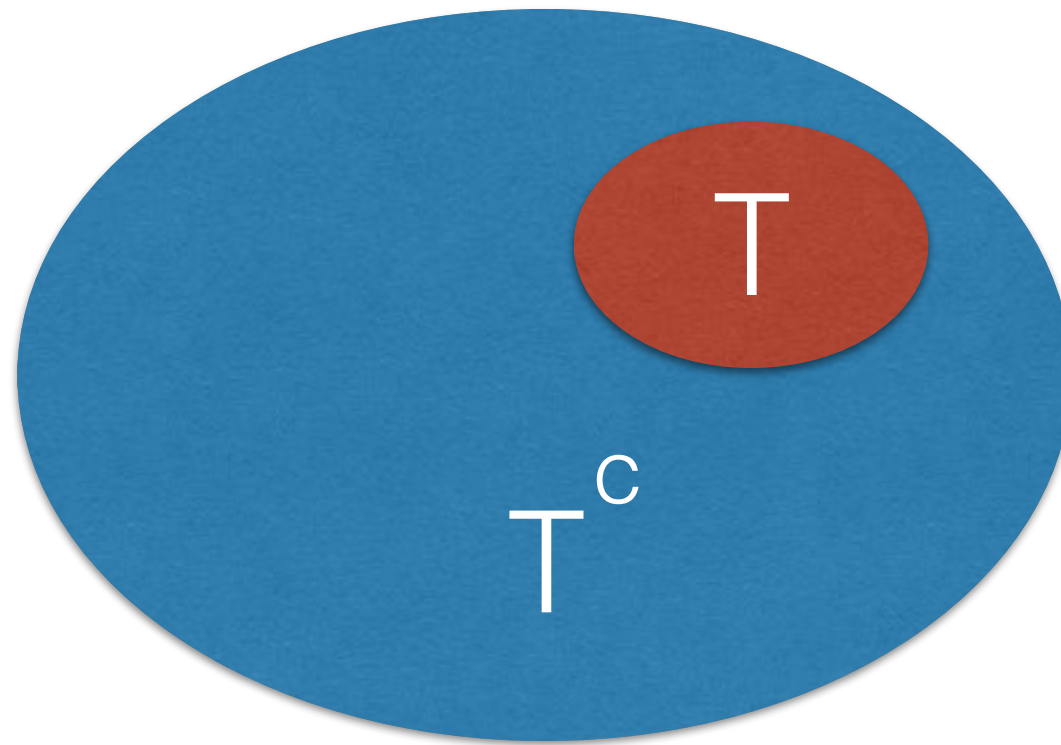
Demographic Parity



Population

FAIRNESS THROUGH AWARENESS

Demographic Parity



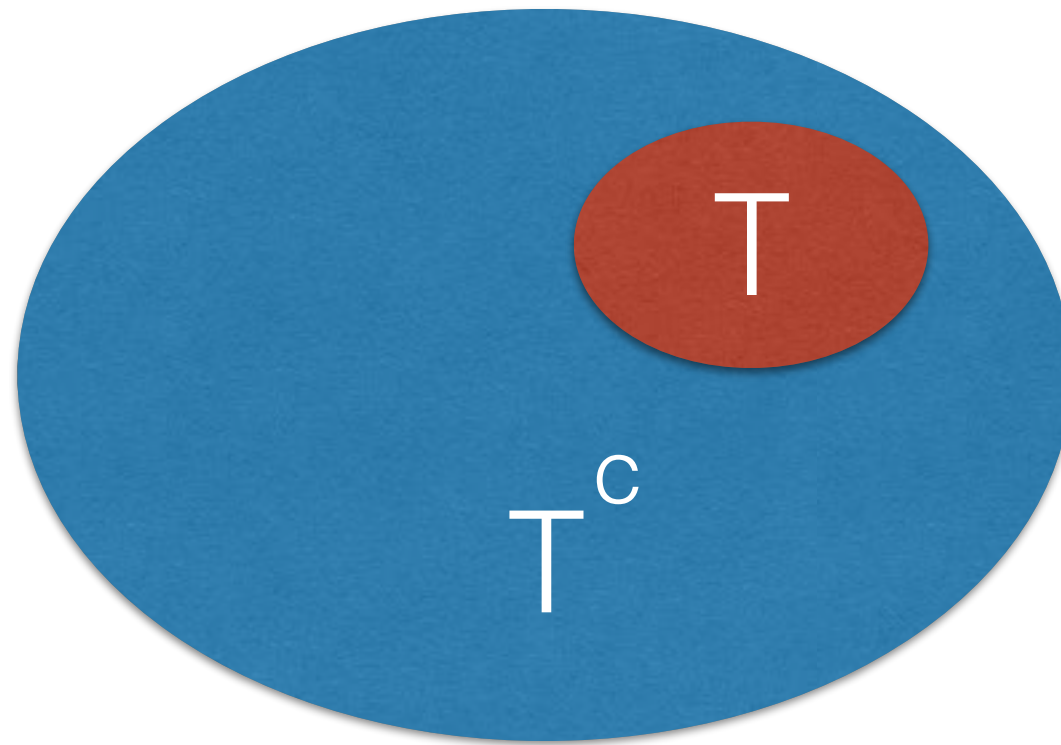
Population

T : Protected subset

T^c : Rest of the population

FAIRNESS THROUGH AWARENESS

Demographic Parity



Population

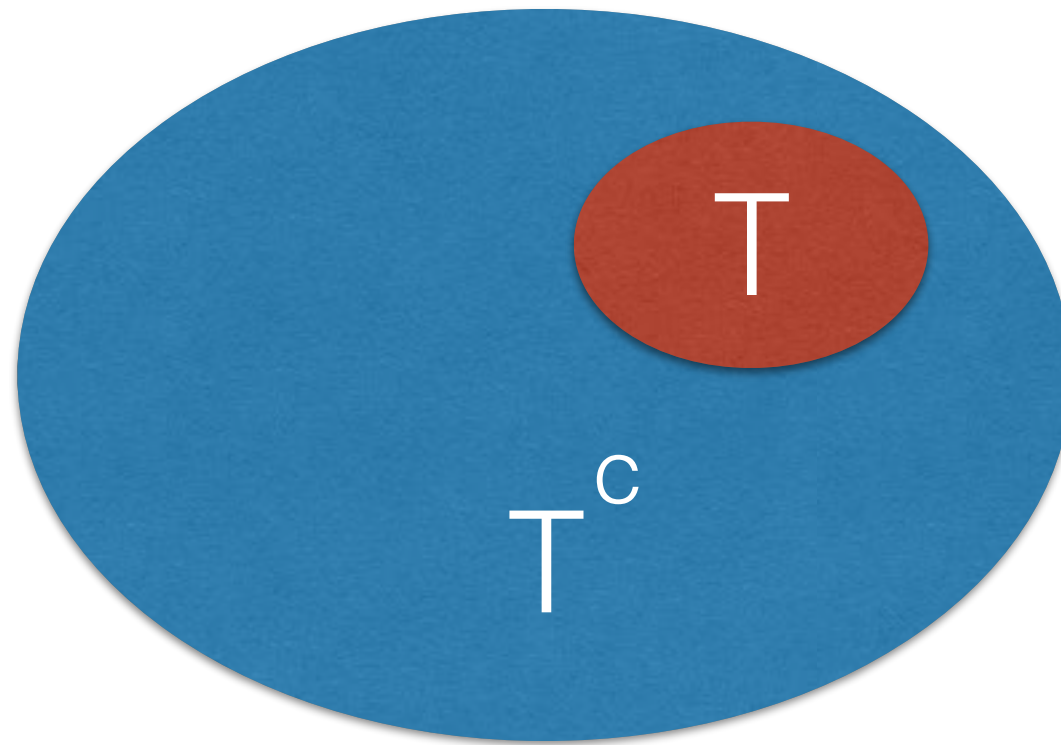
T : Protected subset

T^c : Rest of the population

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

FAIRNESS THROUGH AWARENESS

Demographic Parity



Population

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

FAIRNESS THROUGH AWARENESS

FAIRNESS THROUGH AWARENESS

- Is this good enough?

FAIRNESS THROUGH AWARENESS

- 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**

FAIRNESS THROUGH AWARENESS

- 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 way to satisfy the demographic parity:

FAIRNESS THROUGH AWARENESS

- 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 way to satisfy the demographic parity:
 - Make offer to higher income people in the unprotected class

FAIRNESS THROUGH AWARENESS

- 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 way to satisfy the demographic parity:
 - Make offer to higher income people in the unprotected class
 - Make offer to lower income people in protected class (in same proportion)

FAIRNESS THROUGH AWARENESS

- 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 way to satisfy the demographic parity:
 - 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!

FAIRNESS THROUGH AWARENESS

FAIRNESS THROUGH AWARENESS

- Setup:

FAIRNESS THROUGH AWARENESS

- Setup:
 - Variable T : indicated protected class or not

FAIRNESS THROUGH AWARENESS

- Setup:
 - Variable T : indicated protected class or not
 - Output variable O : Indicates our prediction or outcome

FAIRNESS THROUGH AWARENESS

- Setup:
 - Variable T : indicated protected class or not
 - Output variable O : Indicates our prediction or outcome
 - Variable Y : indicates true/target/desired outcome (eg. Individual capable/qualified, individual can afford etc.)

FAIRNESS THROUGH AWARENESS

- Setup:
 - Variable T: indicated protected class or not
 - Output variable O: Indicates our prediction or outcome
 - Variable Y: indicates true/target/desired outcome (eg. Individual capable/qualified, individual can afford etc.)

Demographic Parity

$$P(O=1|T=1) = P(O=1|T=0)$$

FAIRNESS THROUGH AWARENESS

- Setup:
 - Variable T: indicated protected class or not
 - Output variable O: Indicates our prediction or outcome
 - Variable Y: indicates true/target/desired outcome (eg. Individual capable/qualified, individual can afford etc.)

Demographic Parity

$$P(O=1|T=1) = P(O=1|T=0)$$

Problem: when $T=0$, O can correlate with Y and if $T=1$, O can be random

FAIRNESS THROUGH AWARENESS

Equalized Odds

FAIRNESS THROUGH AWARENESS

Equalized Odds

For all o, y in $\{0, 1\}$

$$P(O=o|Y=y, T=1) = P(O=o|Y=y, T=0)$$

FAIRNESS THROUGH AWARENESS

Equalized Odds

For all o, y in $\{0, 1\}$

$$P(O=o|Y=y, T=1) = P(O=o|Y=y, T=0)$$

- O is independent of T given Y

FAIRNESS THROUGH AWARENESS

Equalized Odds

For all o, y in $\{0, 1\}$

$$P(O=o|Y=y, T=1) = P(O=o|Y=y, T=0)$$

- O is independent of T given Y
- Proposed in Hardt, Price, Srebro 2016 and Zafar, Valera, Rodriguez, Gummadi (2016)

FAIRNESS THROUGH AWARENESS

Equalized Odds

For all o, y in $\{0, 1\}$

$$P(O=o|Y=y, T=1) = P(O=o|Y=y, T=0)$$

- O is independent of T given Y
- Proposed in Hardt, Price, Srebro 2016 and Zafar, Valera, Rodriguez, Gummadi (2016)
- Incentive to reduce error uniformly in all groups

FAIRNESS THROUGH AWARENESS

Equalized Odds

For all o, y in $\{0, 1\}$

$$P(O=o|Y=y, T=1) = P(O=o|Y=y, T=0)$$

- O is independent of T given Y
- Proposed in Hardt, Price, Srebro 2016 and Zafar, Valera, Rodriguez, Gummadi (2016)
- Incentive to reduce error uniformly in all groups

Problem: Say in T , 2/100 people qualify and outside 50/100 qualify
Company can make 26 offers: 25 to qualifying people in T' and 1 in T

FAIRNESS THROUGH AWARENESS

Sufficiency or Predictive Rate Parity

FAIRNESS THROUGH AWARENESS

Sufficiency or Predictive Rate Parity

For all o, y in $\{0, 1\}$

$$P(Y=y|O=o, T=1) = P(Y=y|O=o, T=0)$$

Sufficiency or Predictive Rate Parity

For all o, y in $\{0, 1\}$

$$P(Y=y|O=o, T=1) = P(Y=y|O=o, T=0)$$

- Y is independent of T given O

FAIRNESS THROUGH AWARENESS

Sufficiency or Predictive Rate Parity

For all o, y in $\{0, 1\}$

$$P(Y=y|O=o, T=1) = P(Y=y|O=o, T=0)$$

- Y is independent of T given O
- Equal chance of success($Y=1$) given acceptance

Sufficiency or Predictive Rate Parity

For all o, y in $\{0, 1\}$

$$P(Y=y|O=o, T=1) = P(Y=y|O=o, T=0)$$

- Y is independent of T given O
- Equal chance of success($Y=1$) given acceptance
- Proposed in Zafar et al (2017)

FAIRNESS THROUGH AWARENESS

Sufficiency or Predictive Rate Parity

For all o, y in $\{0, 1\}$

$$P(Y=y|O=o, T=1) = P(Y=y|O=o, T=0)$$

- Y is independent of T given O
- Equal chance of success($Y=1$) given acceptance
- Proposed in Zafar et al (2017)

Problem: Same as equal odds

IMPOSSIBILITY RESULT

IMPOSSIBILITY RESULT

- Turns out that other than degenerate cases, any two of the three criterion are mutually exclusive

IMPOSSIBILITY RESULT

- Turns out that other than degenerate cases, any two of the three criterion are mutually exclusive
- Chouldechova, 2016 and Kleinberg et al. 2016

IMPOSSIBILITY RESULT

- Turns out that other than degenerate cases, any two of the three criterion are mutually exclusive
- Chouldechova, 2016 and Kleinberg et al. 2016
- Demographic parity Vs Sufficiency: if T is dependent of Y .

IMPOSSIBILITY RESULT

- Turns out that other than degenerate cases, any two of the three criterion are mutually exclusive
- Chouldechova, 2016 and Kleinberg et al. 2016
- Demographic parity Vs Sufficiency: if T is dependent of Y .
- Demographic parity Vs Equal Odds: T is dependent of Y and O is dependent of Y .

IMPOSSIBILITY RESULT

- Turns out that other than degenerate cases, any two of the three criterion are mutually exclusive
- Chouldechova, 2016 and Kleinberg et al. 2016
- Demographic parity Vs Sufficiency: if T is dependent of Y .
- Demographic parity Vs Equal Odds: T is dependent of Y and O is dependent of Y .
- Equal Odds Vs Sufficiency: If T is dependent of Y

ACHIEVING FAIRNESS

- Preprocessing: While doing feature extraction, extract features that ensure independence of feature to T (Eg. Equal odds)
- While training: Find model that minimizes training error subject to fairness constraints
- Post-processing: Learn model as before on training data, as post processing use extra training data to learn a bias parameter to correct for fairness

On to the next social issue...

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

Fairness, Accountability, and Transparency in Machine Learning

DATA
TRANSPARENCY
LAB

DTL Conferences ▾

Become a sponsor

Grants Program ▾

Technical Program ▾

About ▾

Up to Date

Workshop ▾

ES

Contact

Newsletter



DTL CONFERENCES

DTL2017

Barcelona **December 11-13th**