

# Last Lecture

# SUMMARY: STATISTICAL LEARNING

- Useful in scenarios where we can collect training data from the same pool of examples as we get training data from in an iid fashion (Eg. typical object recognition etc.)
- Measure of Performance: Excess risk bounds w.r.t. best model in a class of models
- Algorithm of choice : Empirical Risk Minimization (ERM), Regularized ERM
- Analysis:
  - Bounds on excess risk via uniform convergence.
    - For binary classification, worst case rates characterized by VC dimension
    - More generally, Rademacher complexity gives us a handle on rates
  - Bounds via algorithmic stability: depend on algorithm used.

# SUMMARY: ONLINE LEARNING

- Useful in scenarios where no iid assumptions on data hold but we know that a fixed model class is good for our problem. Needs continuous feedback
- Measure of Performance: Regret against best model in hindsight
- Algorithms: Deriving algorithm and proving they work go hand in hand. Classic algorithms like ERM don't work
- Examples:
  - Online gradient descent:  $f_{t+1} \leftarrow f_t - \eta \nabla \ell(f_t, (x_t, y_t))$
  - Exponential weights algorithm:  $q_t(f) \propto \exp(-\eta \sum_{j=1}^{t-1} \ell(f; (x_j, y_j)))$

# SUMMARY: BANDIT PROBLEMS

- Useful in practical scenarios where we can't evaluate every model on every time step but only get limited feedback on the loss of the chosen model or prediction or action on a given instance.
- Stochastic setting: Using Lower (or upper) confidence bound algorithm. Optimism pays off, either we learn to eliminate quickly or we are correct.
- Adversarial Setting: Use full information (classic) algorithms but used unbiased estimate of losses on every round.

# SUMMARY: COMPUTATIONAL LEARNING THEORY

- There are problems that can be learnt in sample efficient way but not computationally efficiently
- Proper Vs Improper learning makes a huge difference in terms of computational efficiency of learning
- Proper learning hardness can be shown via NP reductions
- Improper learning hardness results we need to use other methods like cryptographic hardness
- Hardness results let us know what to focus on Eg. in theory of deep learning

# SUMMARY: DIFFERENTIAL PRIVACY

- We need to be aware of privacy concerns while developing ML algorithms
- Differential Privacy is one such mechanism where we build randomized algorithms that are not too sensitive to any one data point
- Typical mechanism, inject noise into algorithm either at the output or within the algorithm
- Beware of reusing data, can lead to faulty conclusions
- Differential privacy can be used to alleviate this issue.

# ML DREAM

shoes

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](http://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](http://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](http://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/](http://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](http://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...

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 contains four organic search results, each enclosed in a red rectangular box:

- 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 promotes "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, Product quality, Shipping, Service, and Selection are all 8.5/10.
- Converse® Official Site - Converse.com**: An advertisement for Converse.com with a 4.2-star rating. It promotes "Full Converse Collection. Shop Our New Designs for Men, Women & Kids." and lists types like 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 promotes "Exclusive Deals On All Shoes, Socks, & More at Zulily. Shop & Save Today!"

The right section contains a sponsored product listing, also enclosed in a red rectangular box:

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

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](http://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...", "Sandals", "Womens Shoes", "Sneakers & Athletic Shoes", and "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 'What is LinkedIn?', 'Join Today', and 'Sign In' are on the right. Below this is a search bar with 'Data Scientist' and 'United States' entered, and a 'Find jobs' button. On the left side, there are three filter sections: '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. The main area displays 8,715 jobs, sorted by Relevance. Four job listings are visible, each with a company logo, job title, location, description, and an 'Apply with Profile' button.

**LinkedIn** 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** 13d  
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

**Data Scientist** 8d  
feedzai  
Atlanta, Georgia  
...client's data feeds Work with the the client to explore their data and better understand it Work...

**Data Scientist** 17d  
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!

**Data Scientist** 5d  
Covestro  
Greater Pittsburgh Area  
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

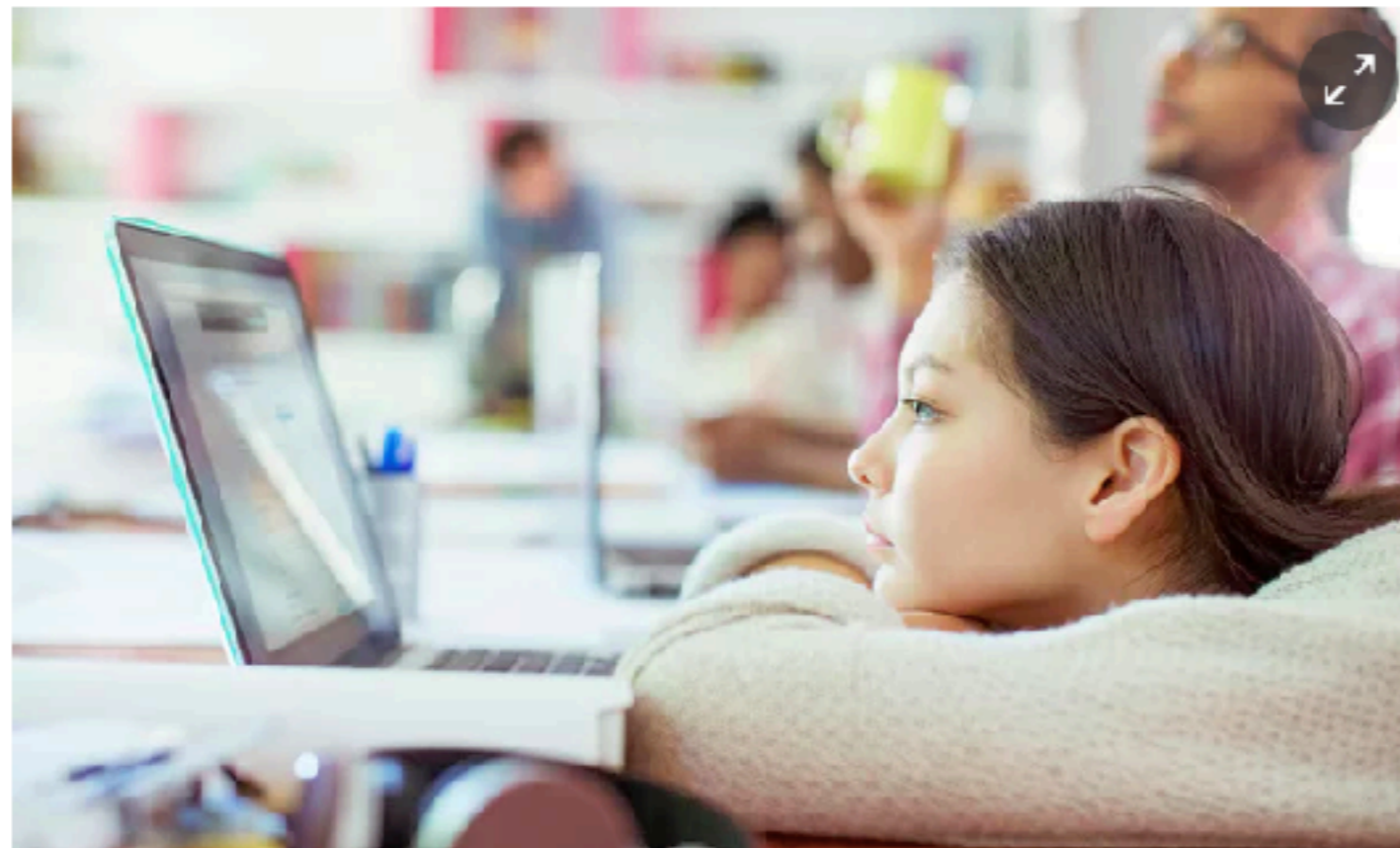
Samuel Gibbs

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.)*

# 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?

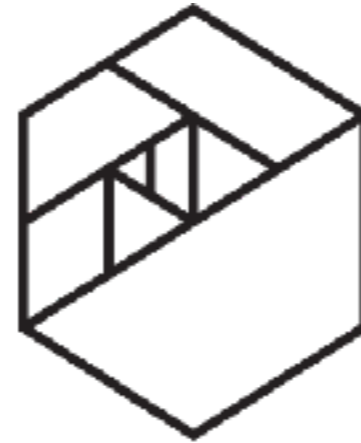


Machine Learning

galvanize

 Startup.ML

**coursera**



**METIS**

**datascience@berkeley**

 The Data Incubator

# WHY NOW?



Machine Learning

galvanize

 Startup.ML

**coursera**



**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](http://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

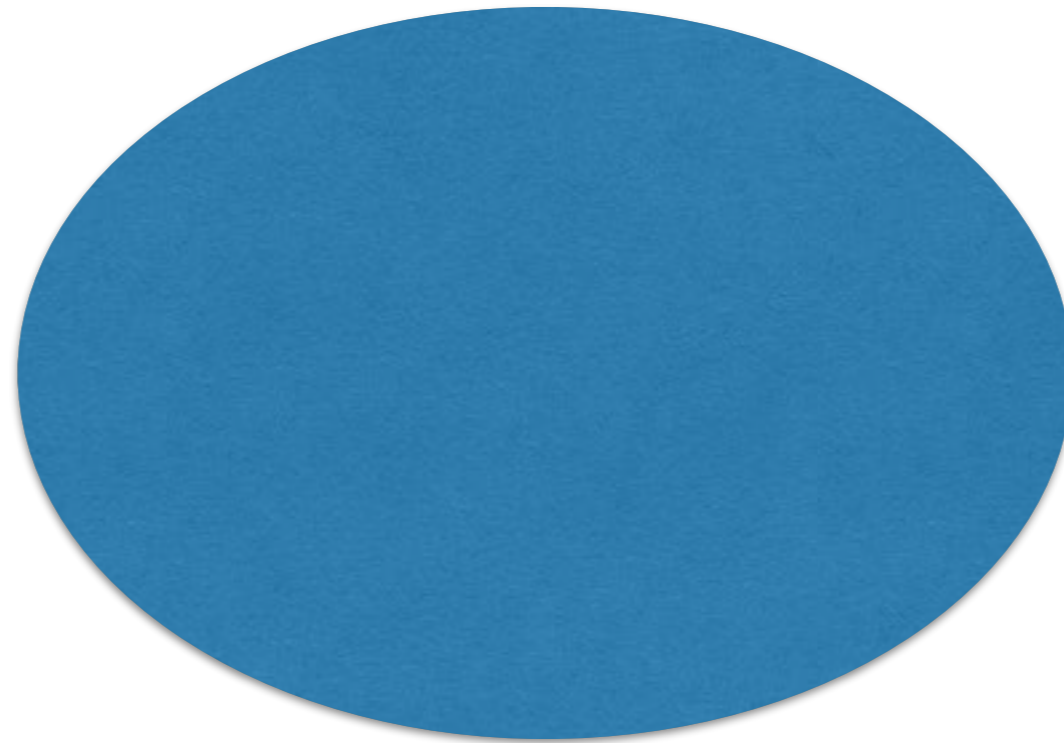
Most ML models aim for accuracy for the majority at the expense of mistakes on the smaller protected class

# FAIRNESS THROUGH AWARENESS

## **Demographic Parity**

# FAIRNESS THROUGH AWARENESS

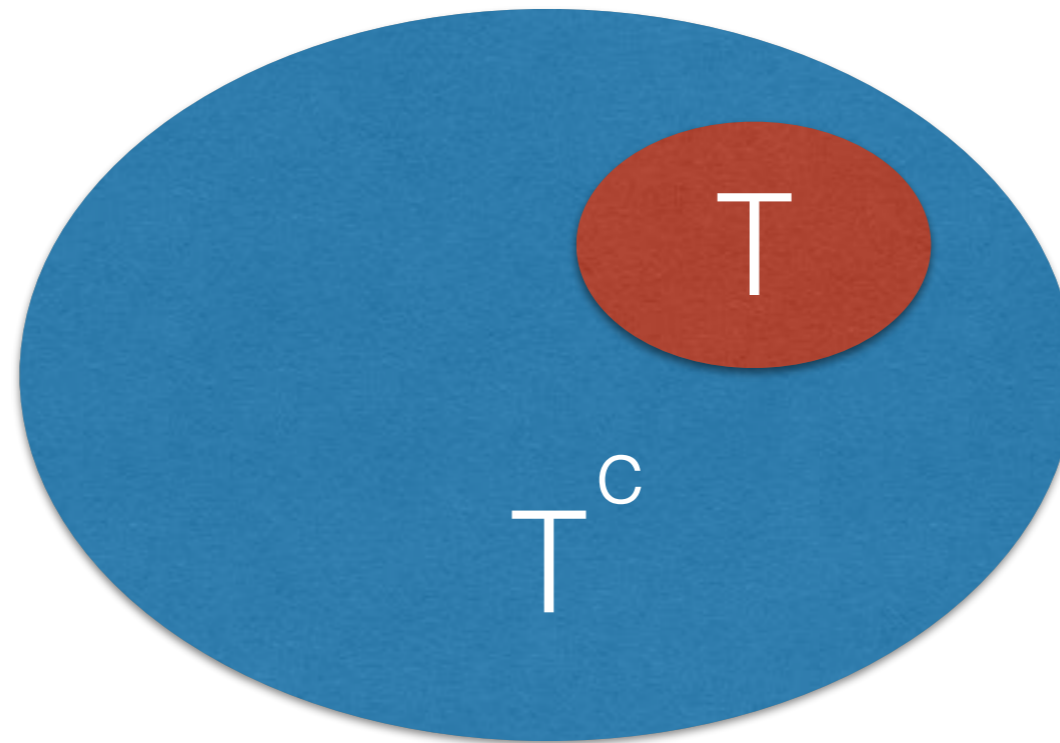
## **Demographic Parity**



Population

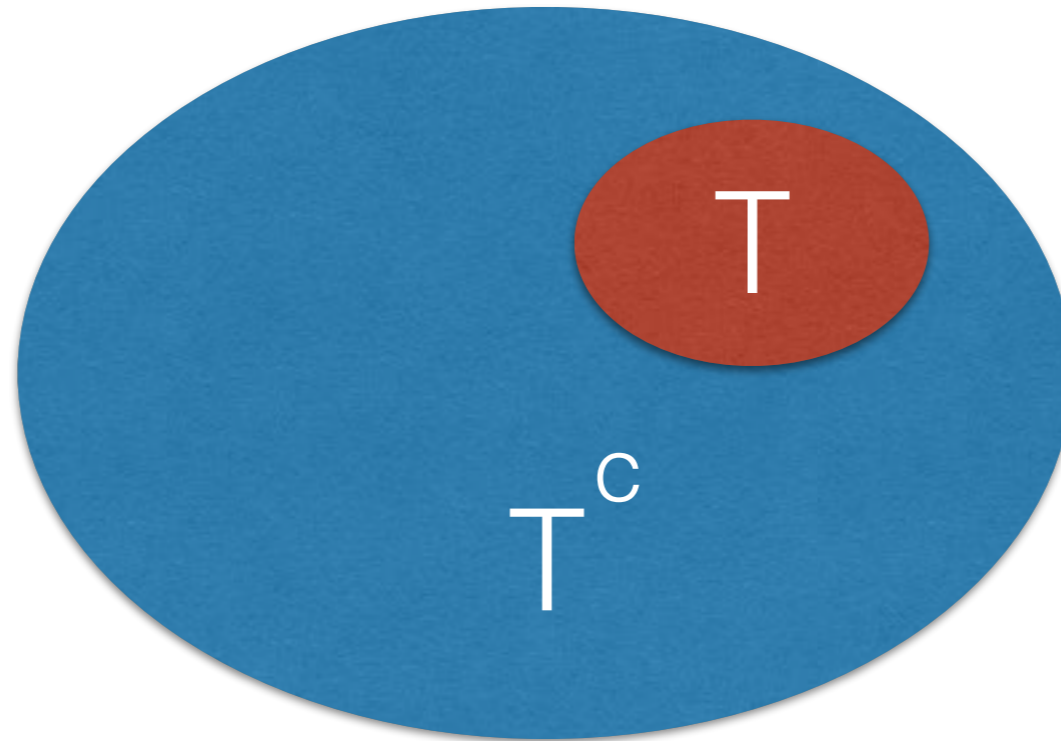
# FAIRNESS THROUGH AWARENESS

## Demographic Parity



Population

## Demographic Parity

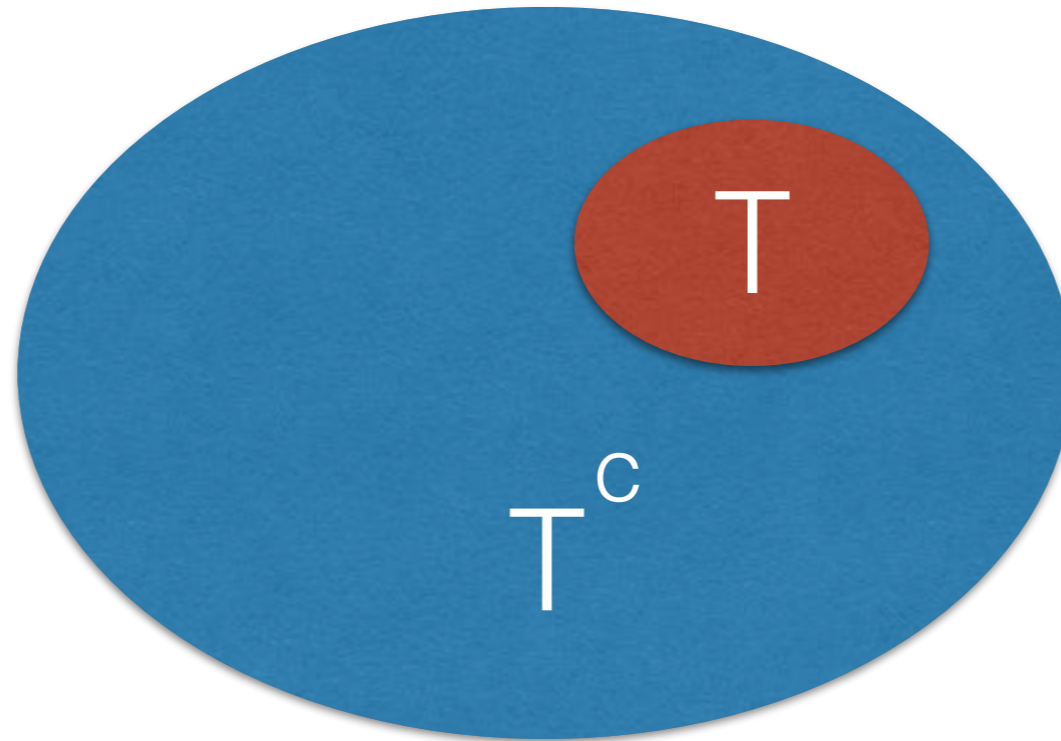


Population

$T$  : Protected subset

$T^c$  : Rest of the population

## Demographic Parity



Population

$T$  : Protected subset

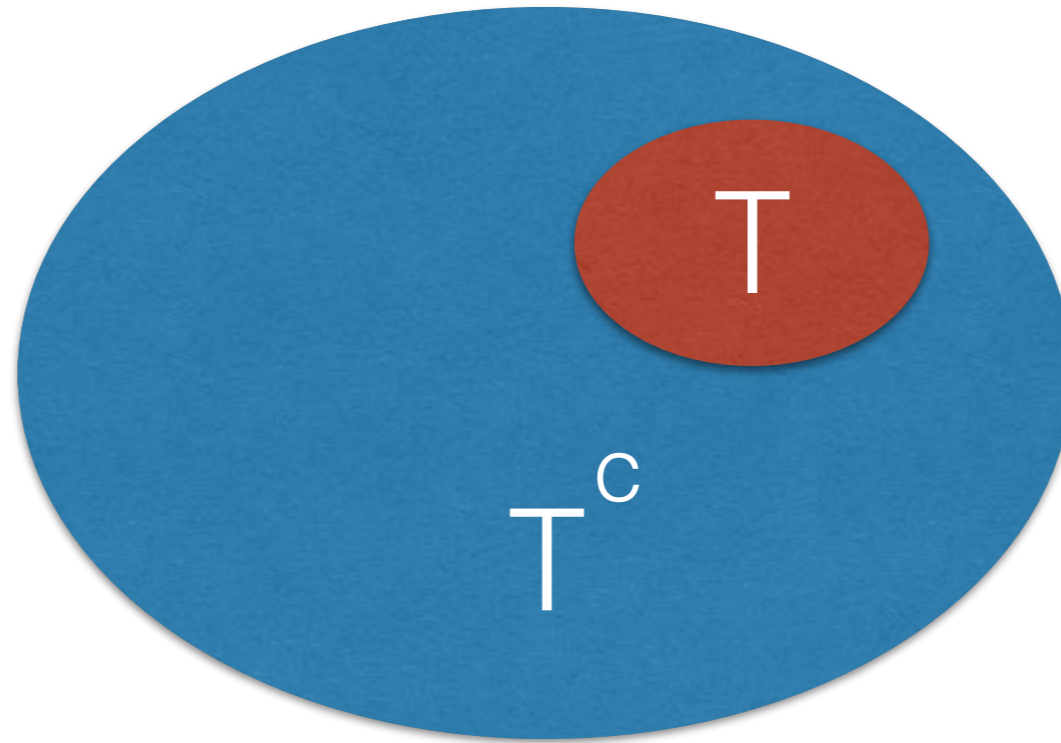
$T^c$  : Rest of the population

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



# 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

## **Equalized Odds**

# FAIRNESS THROUGH AWARENESS

## **Equalized Odds**

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

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

## 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  $Y$

## 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  $\{0, 1\}$

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



# 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  $O$

# 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)

## 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
- While there are no perfect fairness criterion, we still want to ensure at least some notion of fairness,



# 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
- While there are no perfect fairness criterion, we still want to ensure at least some notion of fairness,
- how do we do it?

# 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
- While there are no perfect fairness criterion, we still want to ensure at least some notion of fairness,
- how do we do it?

# 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

# ACHIEVING FAIRNESS

- While training: Find model that minimizes training error subject to fairness constraints

# 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

# ACHIEVING FAIRNESS

- Post-processing:
  - Learn model as before on training data,
  - As post processing use fresh training data to learn a bias parameter to correct for fairness
- Eg. Equal Odds (Binary classification)
  - Learn mapping  $f$  from training set such that from input to reals such that  $Y = 1$  if  $f(X) > 0$  and  $Y = 0$  if not
  - Now on fresh dataset, learn new threshold  $\theta$  such that for protected class,  $Y = 1$  if  $f(X) > \theta$  and  $Y = 0$  if not
  - $\theta$  is chosen so as to ensure Equal odds