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 cant 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.

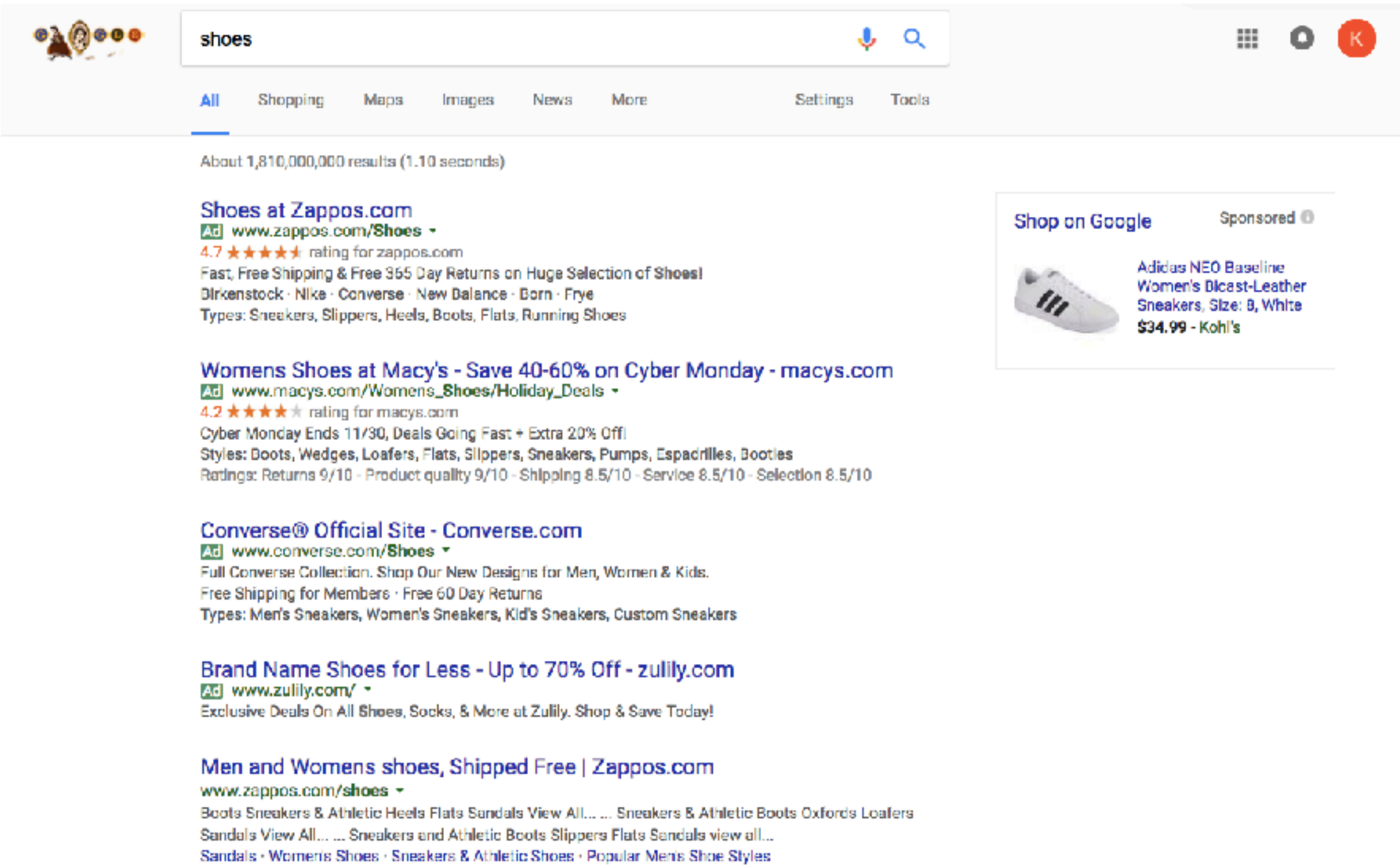
Summary: Computational Learning Theory

- 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

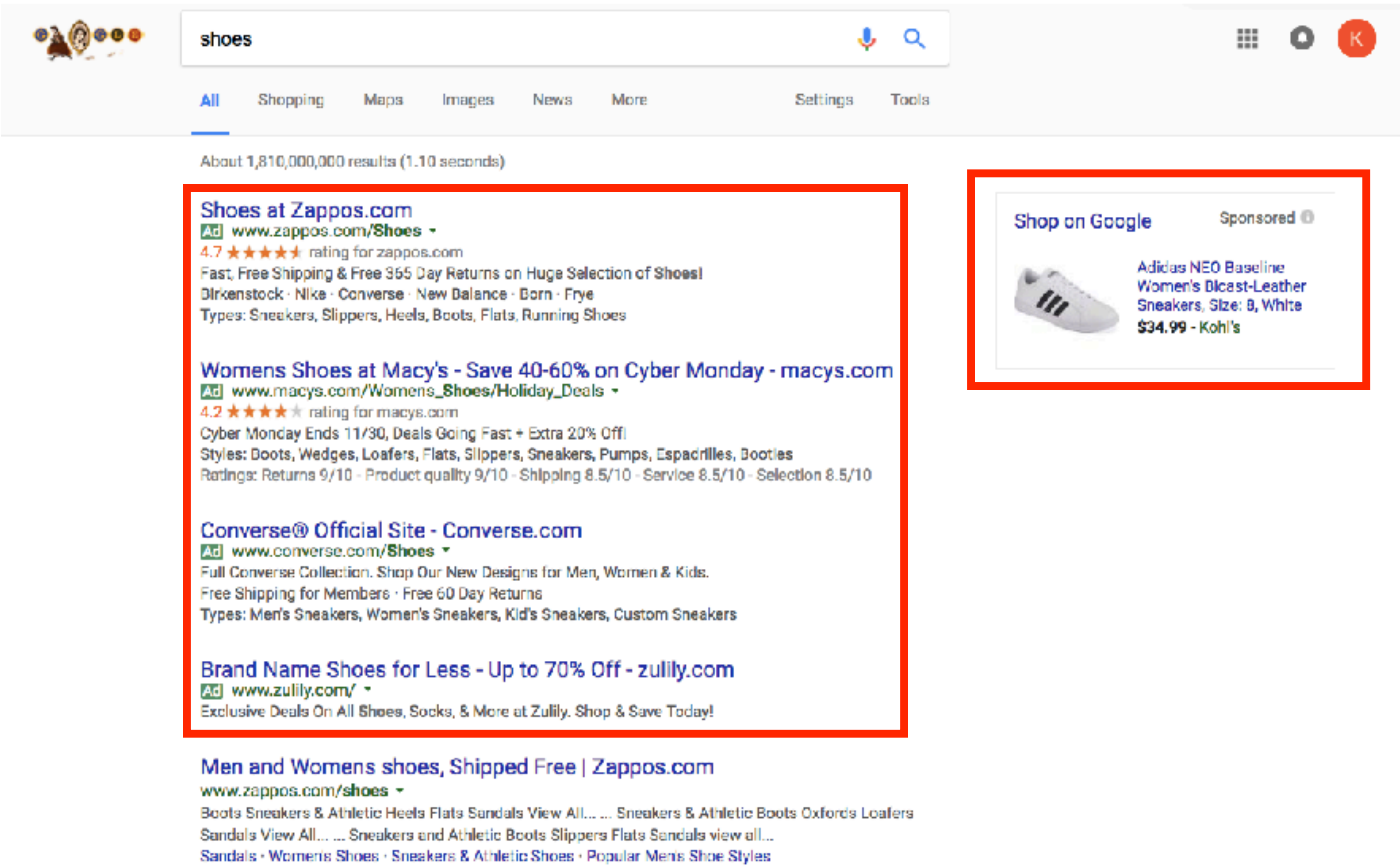
- We need to be aware of privacy concerns while developing ML algorithms
- Differential Privacy in one such mechanism where we build randomized algorithms that are no too sensitive to any one data point
- Typical mechanism, inject noise into algorithm either at the output or within the algorithm
- Beware 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
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 Blcast-Leather Sneakers, Size: 8, White
$34.99 - Kohl's

ML Dream

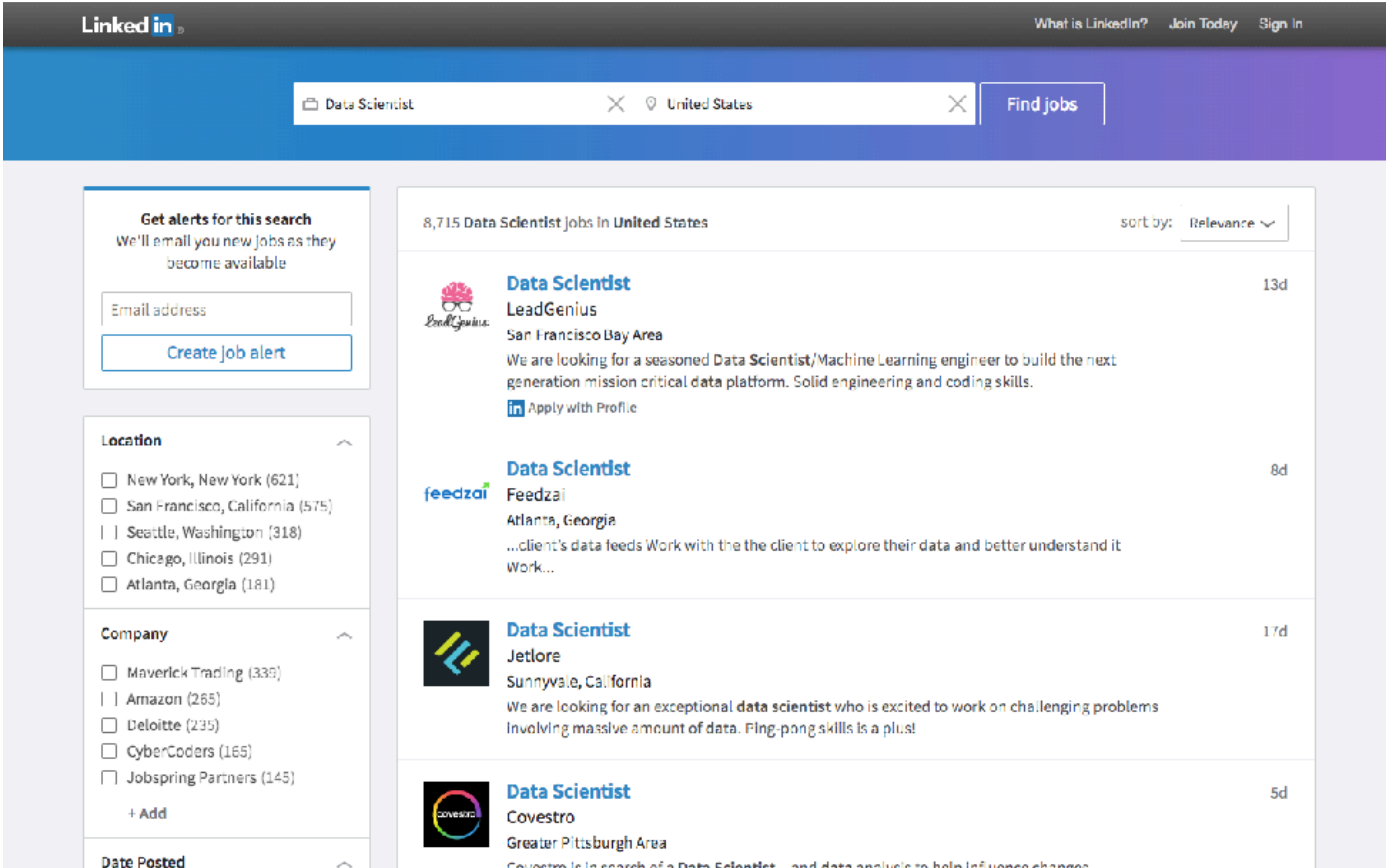
What product would you like…



What news would you prefer to read…

ML DREAM

Find the best job for you…



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

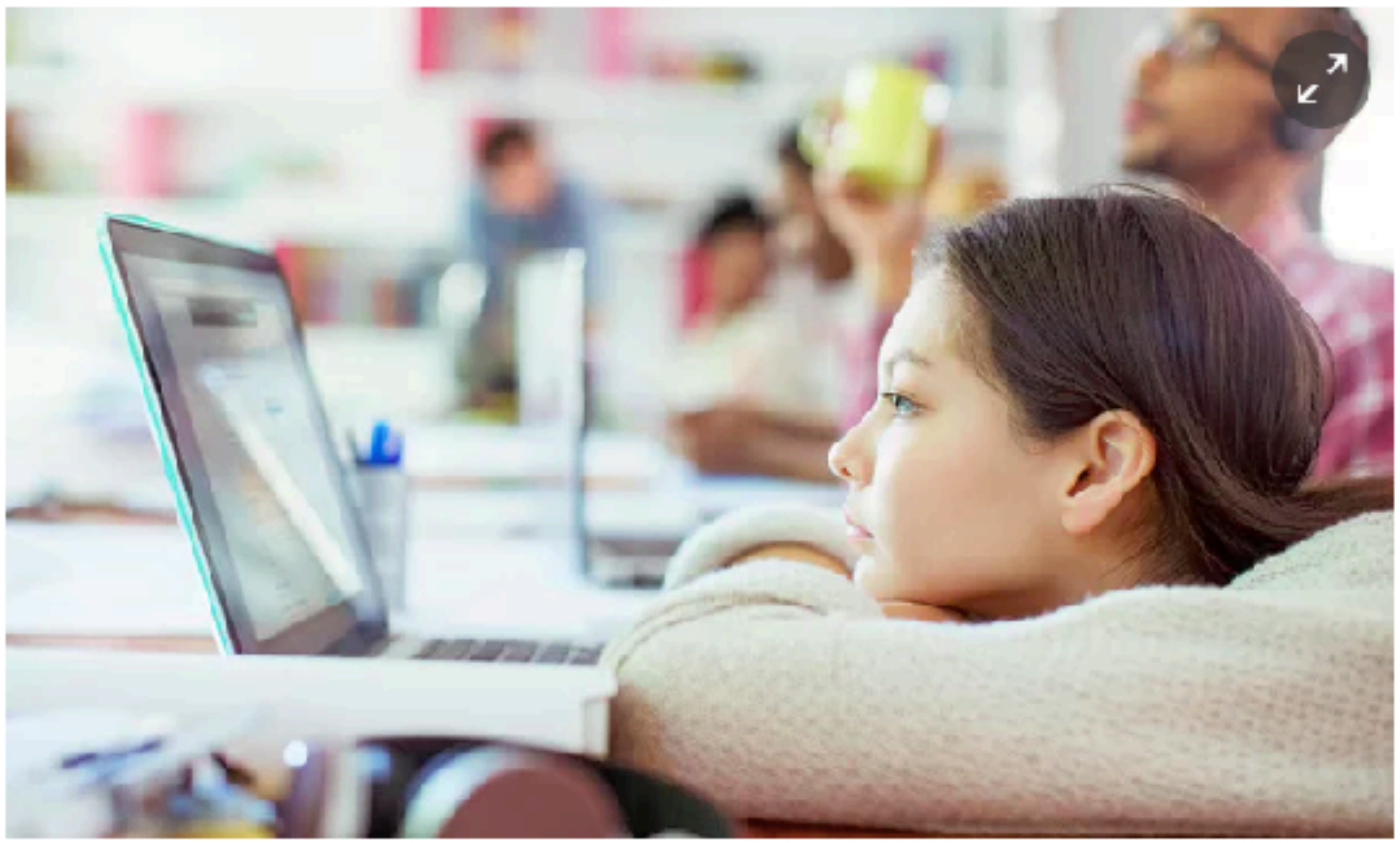
Automated testing and analysis of company's advertising system reveals male job seekers are shown far more adverts for high-paying executive jobs

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.

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?

galvanize

Startup.ML

coursera

datascience@berkeley

The Data Incubator

METIS

Why Now?

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

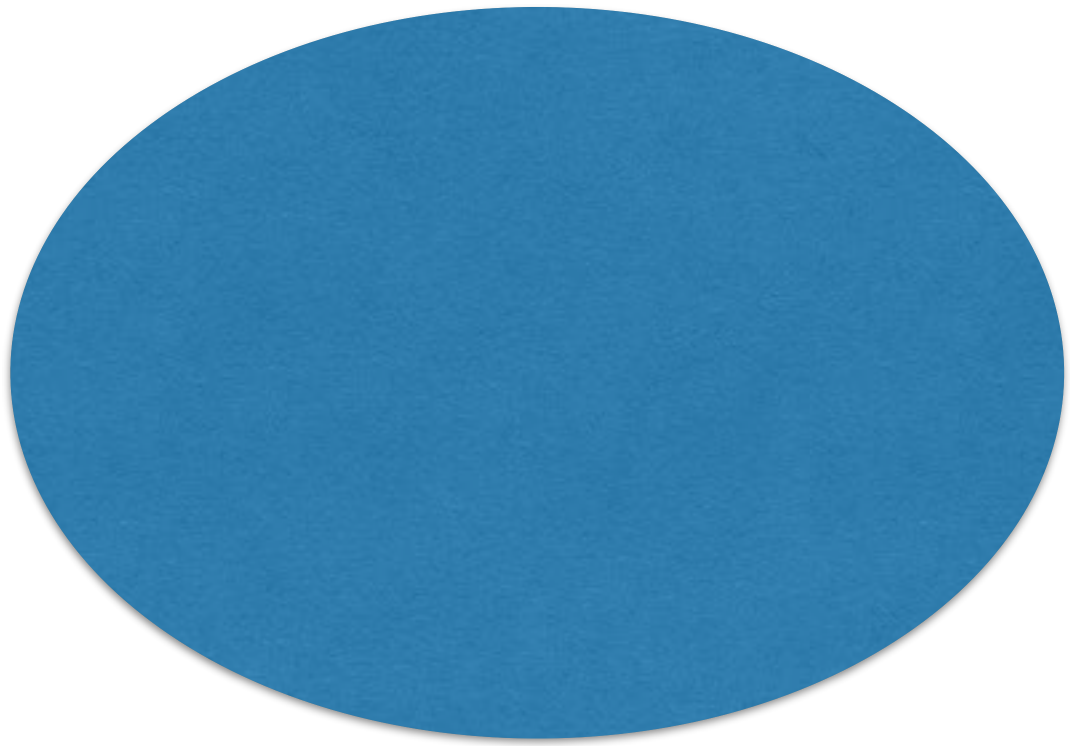
Eg. Real Vs Fake Names

- 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

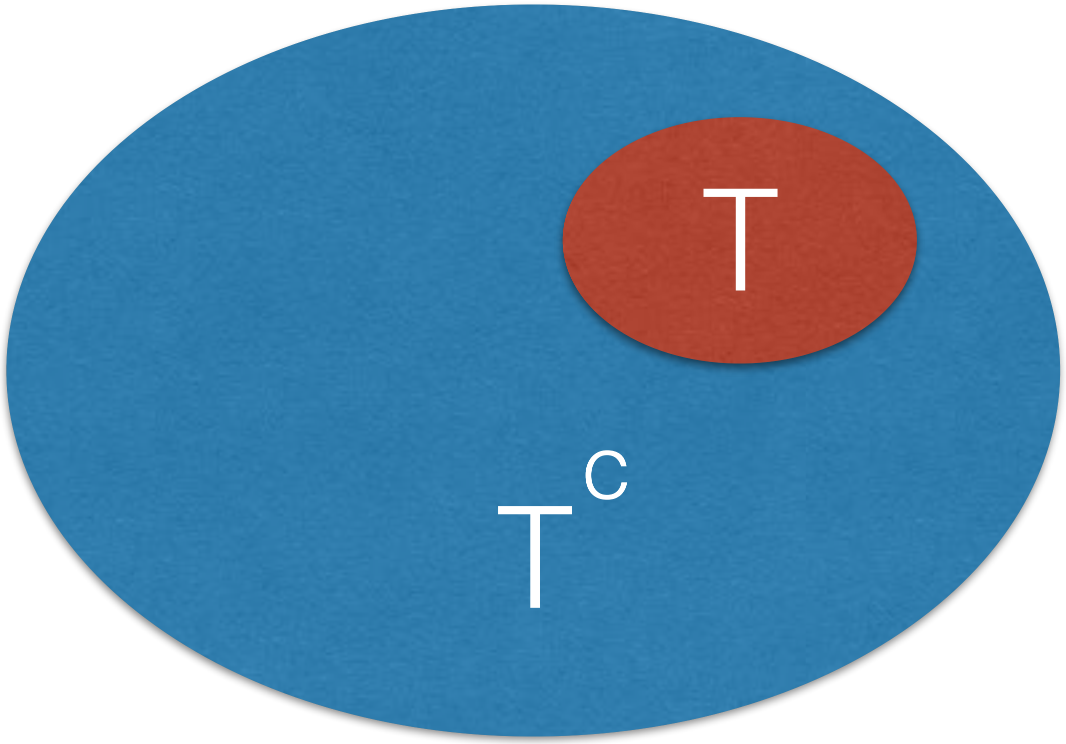
Demographic Parity

Demographic Parity



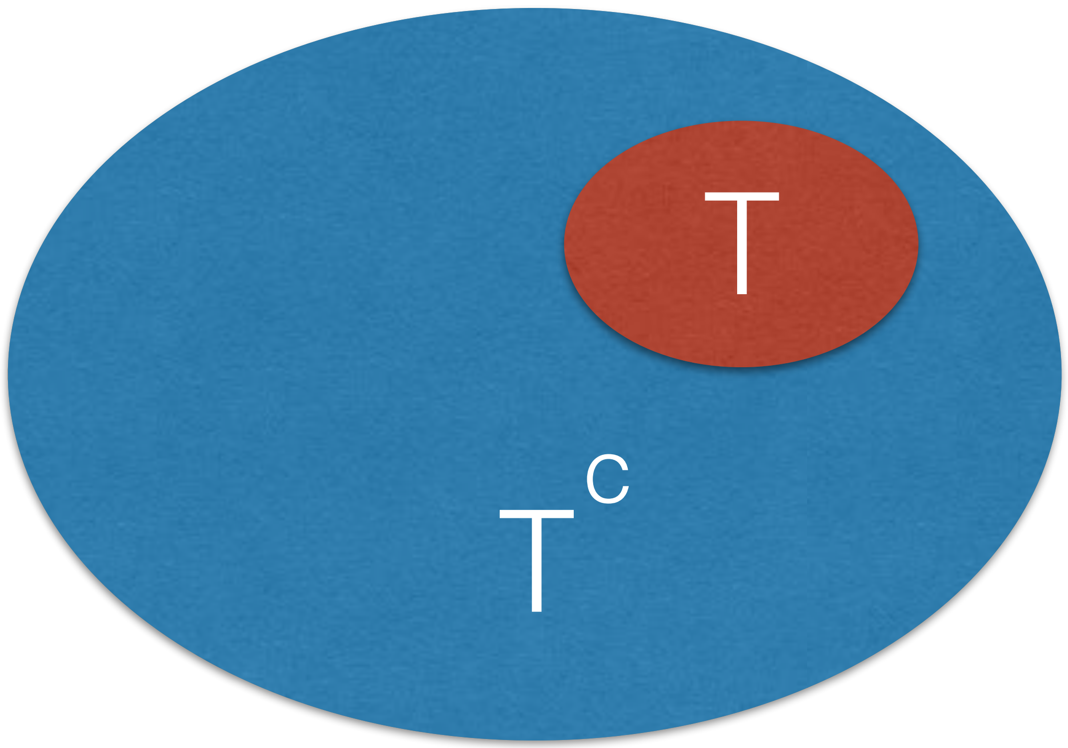
Population

Demographic Parity



Population

Demographic Parity

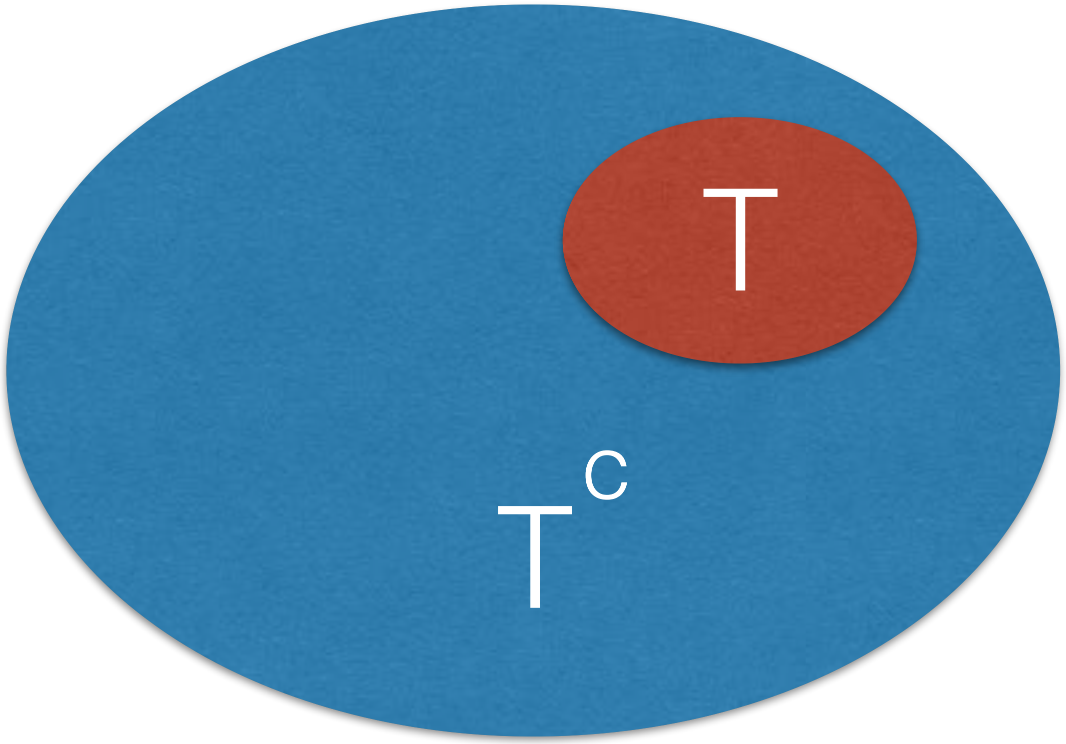


Population

T : Protected subset

T^c : Rest of the population

Demographic Parity



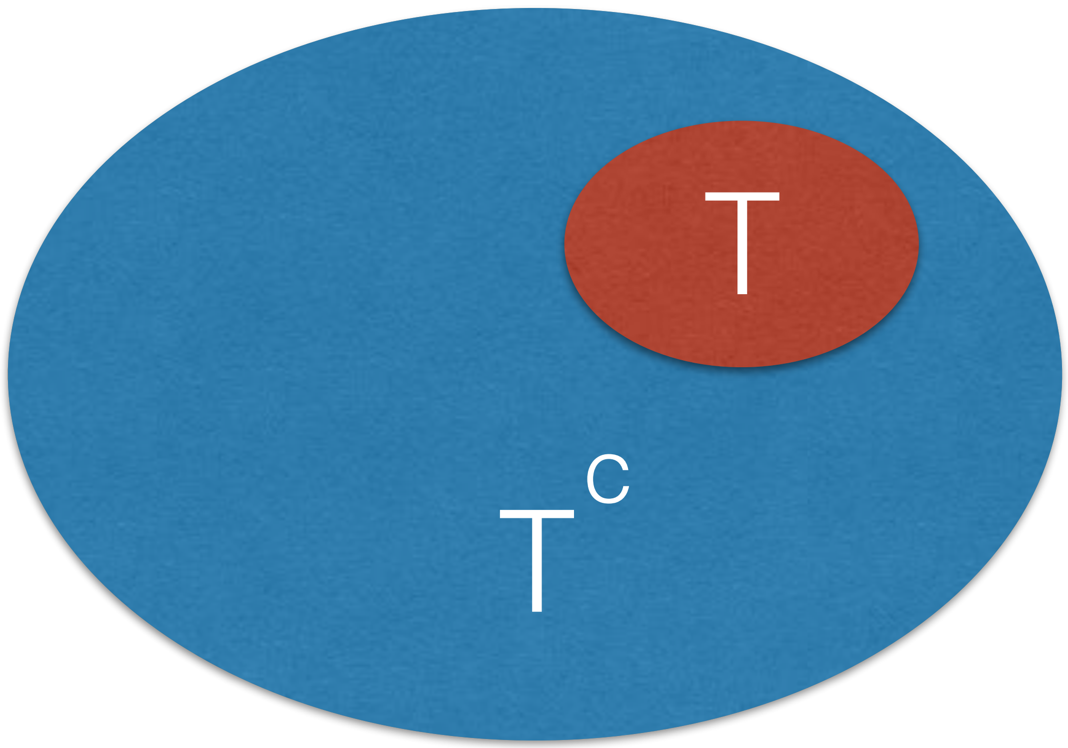
Population

$T:$ Protected subset

$T^c:$ Rest of the population

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

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!

Equalized Odds

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)

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

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

Sufficiency or Predictive Rate Parity

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

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