ML and Society

Apr 21, 2022

Discussion summaries graded

note @57

stop following

0 views

Discussion Summary graded

The discussion summary for tomorrow has been graded! Please do look at my comments/feedback on Autolab.

Also the in-class discussion scores have not been added since that has not happened yet.

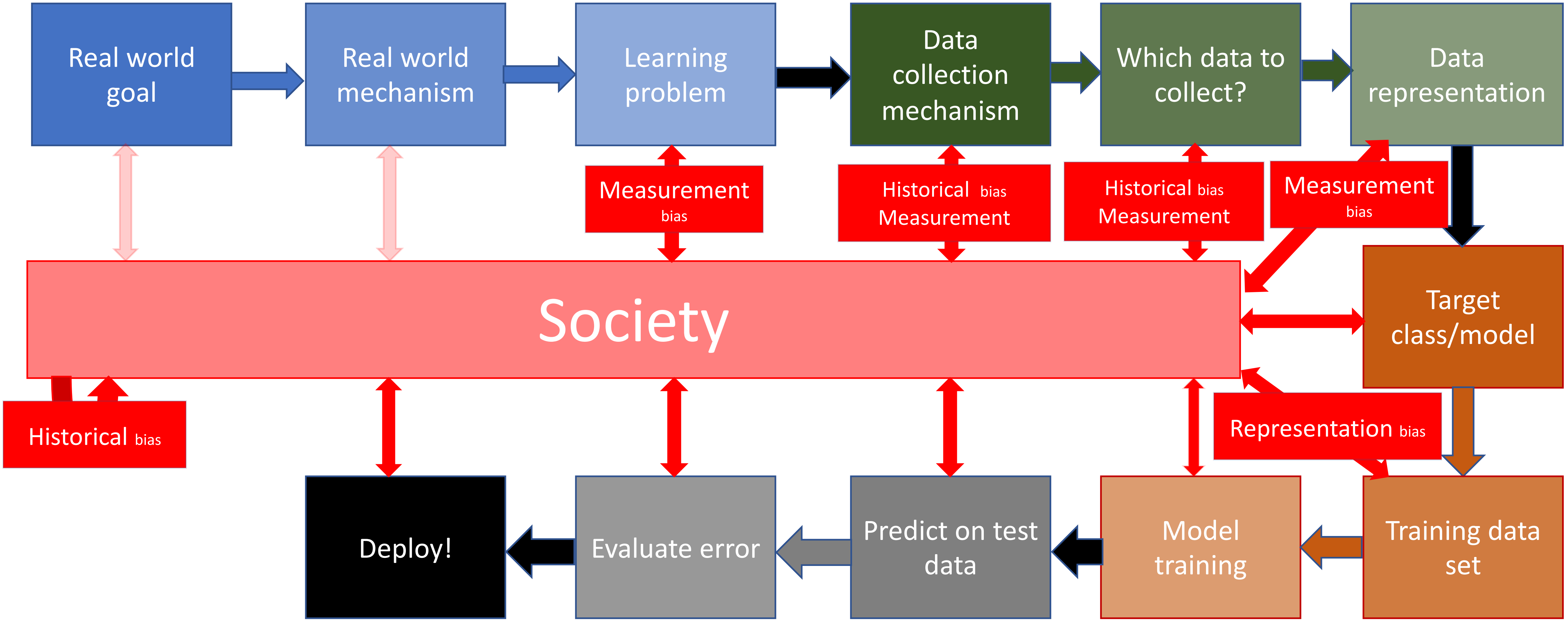
Based on your submissions, I think tomorrow's discussion is gonna be great: I'm looking forward to it :-)

autolab discussion_summary

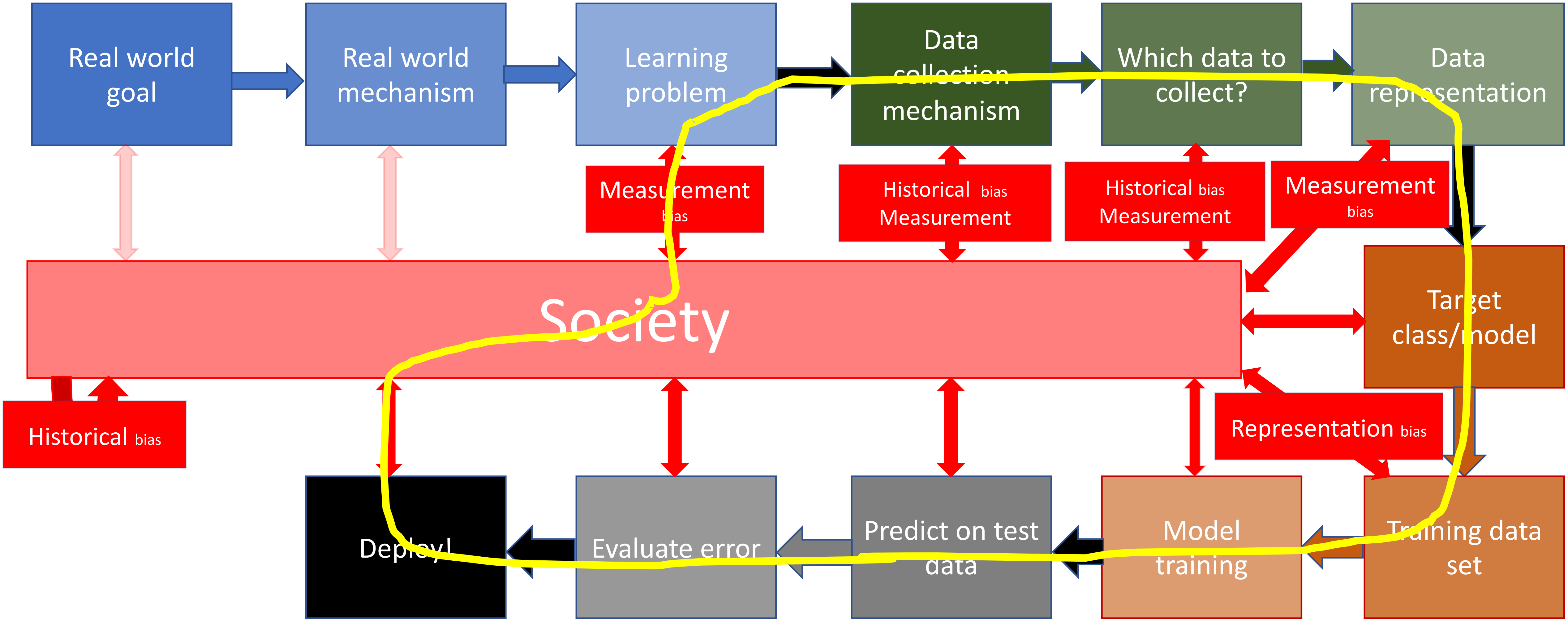
edit · good note | 0

Updated Just now by Atri Rudra

What is a feedback loop



The "loop" in feedback loop



This happened last time I taught the course..

Kate Crawford
@katecrawford

Big news: LAPD will end the use of the broken predictive policing system known as PredPol, citing budget concerns under COVID-19. This is thanks in large part to community groups like @stoplapdspying pushing back against its use.

LAPD will end controversial program that aimed to predict where crimes woul...
Chief Moore says, due to financial constraints caused by the pandemic, the LAPD will end a program that predicts where property crimes could occur.
latimes.com

Do feedback loops exist?

How do we “prove” that feedback loops can exist in predictive policing?

Simulation results

Theoretical modeling results

A simulation result

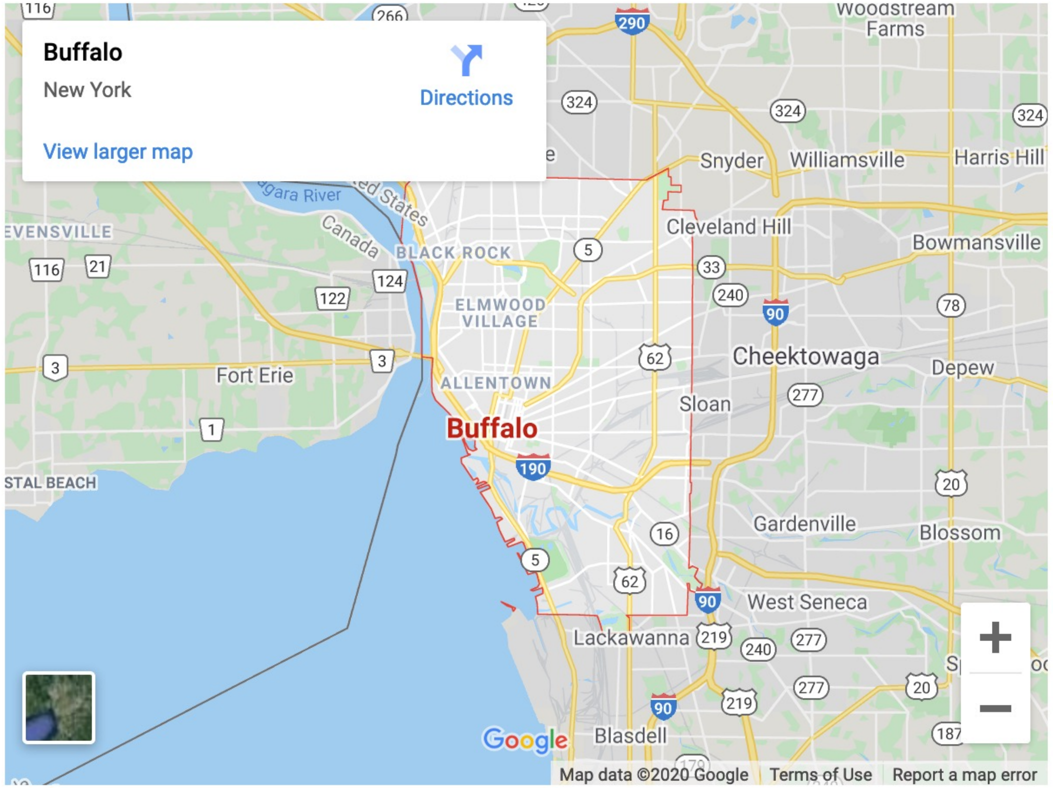
IN DETAIL

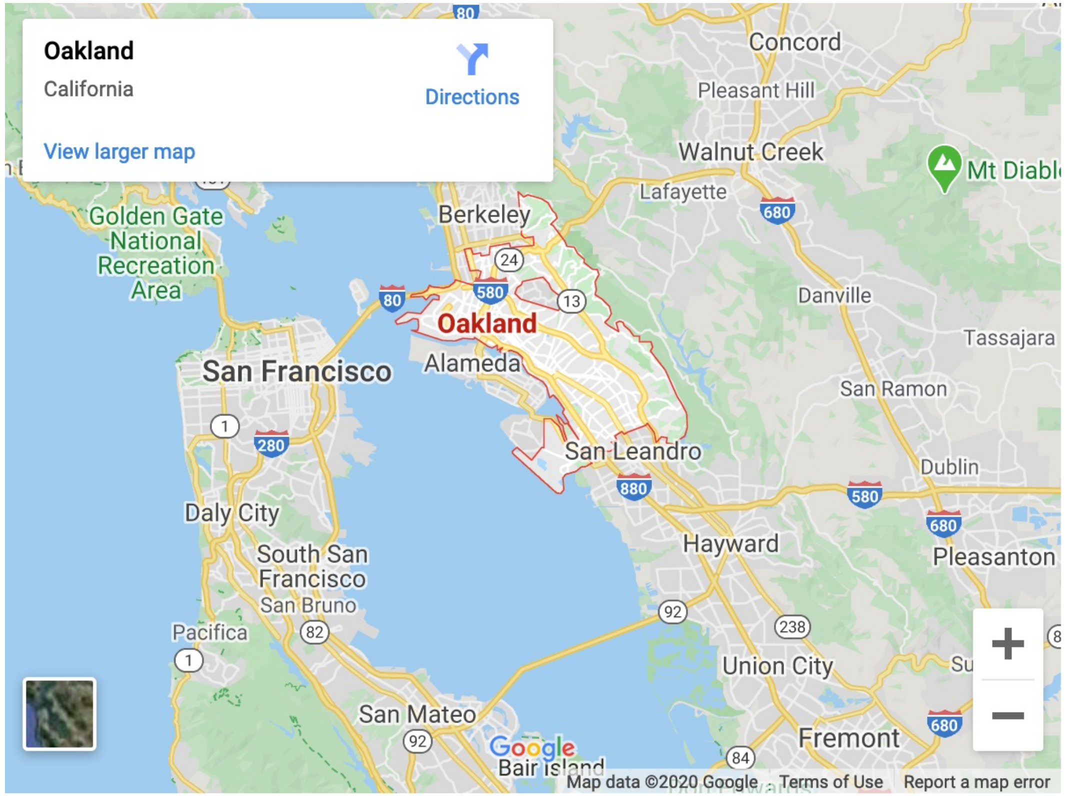
To predict and serve?

Predictive policing systems are used increasingly by law enforcement to try to prevent crime before it occurs. But what happens when these systems are trained using biased data? **Kristian Lum** and **William Isaac** consider the evidence – and the social consequences

How would you run this simulation?

Step 1: Pick a city





Run predpol on historical crime data and see if there is indeed a feedback loop

Step 2: Pick a crime type

Drug related crimes

For the rest of the section **let us assume you have access to drug crimes arrests for say year 2019 in Buffalo**. This is a realistic assumption since many cities in the US
male crime incidents data public: e.g. here is crime incidents reported in Buffalo.

Access to predpol?

What about access to predpol?

We will **assume that you have access to the predpol algorithm**. It turns out that there is enough information about the algorithm (see e.g. overview from predpol that this is not an unreasonable assumption.

One of the video above briefly talked about how predpol works but here is a quick summary:

1. The geographical areas are divided into "grids" and all the crime information is aggregated in each cell.
2. predpol (as per its founders and via Lum and Isaac) "only three data points in making predictions: past type of crime, place of crime and time of crime. It uses no personal information about individuals or groups of individuals, eliminating any personal liberties and profiling concerns."

Questions we want answered

Initial State Question

Here the question basically is the following. If we were to use predpol to say decide on where potential drug crimes were to happen on January 1, 2019 in Buffalo, then how does predpol's prediction compare to "actual" drug crimes on the same day.

`Note` The use of quotes in "actual" is on purpose-- we will come to this shortly.

Feedback Loop State Question

Here the question is actually to test whether predictive policing could lead to feedback loop. The idea is to simulate the use of predpol in Buffalo from January 1, 2019 to December 31, 2019 in Buffalo and use the observed drug crime data for a given day to generate predpol's prediction for the next day. The goal here is to check if the the odds of location being "targeted" by predpol goes up (say when compared to the reported 2019 drug crimes as a baseline).

Initial state question

Exercise (using crime reports data)

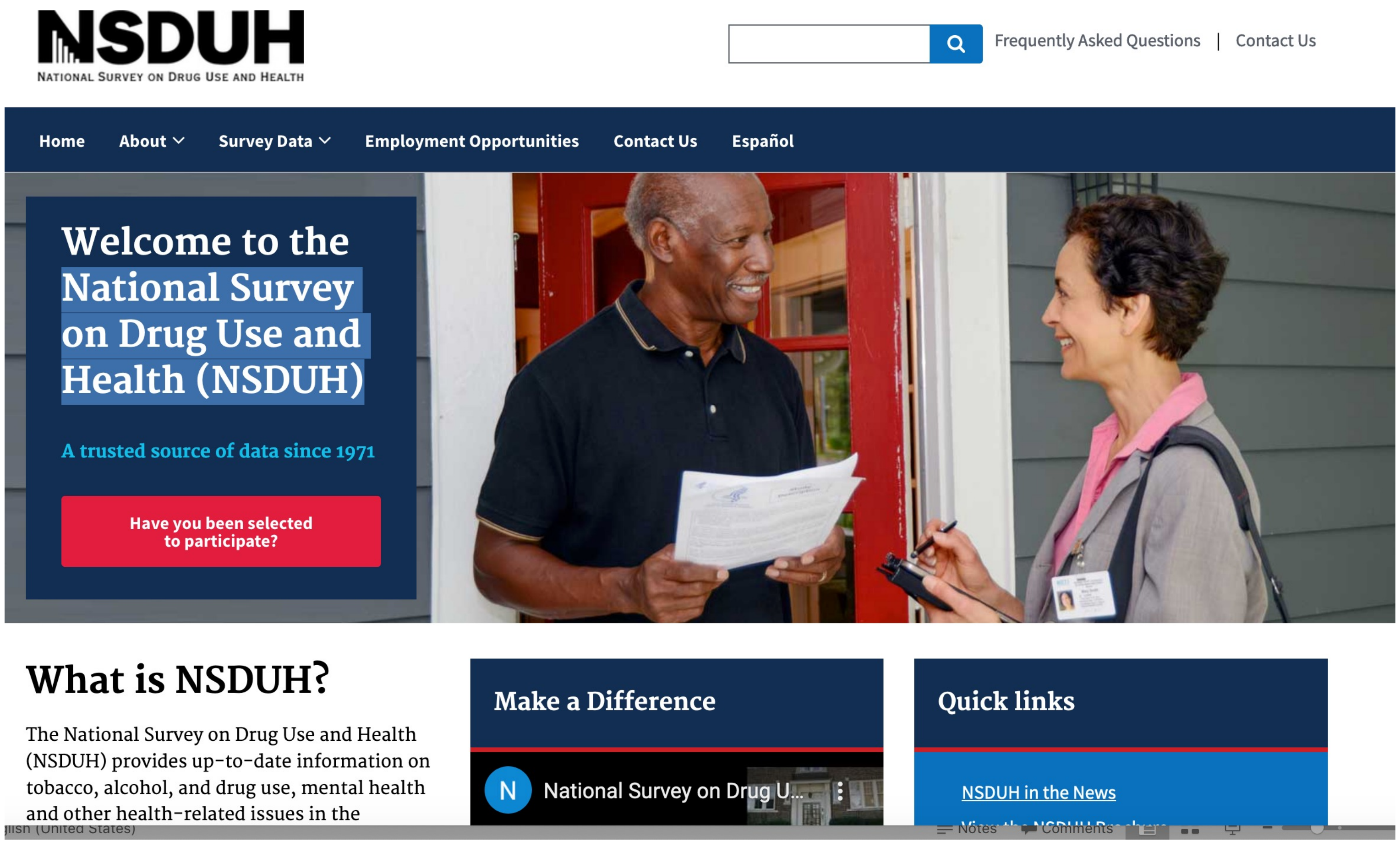
You already have access to reported drug crimes data (which will also be used by predpol): one option would be to use this option as a proxy for the actual occurrences of drug crimes. Is this fine? If so, why would this not be biased? If not, why not?

Click here for the answer

As we have seen in a video earlier, there is representation bias and measurement bias in police drug crime/arrest records, so if we use the reported crime data then we will **not** be getting close to the ground truth.

Actual drug use data?

Go door to door and ask in each household in Buffalo about their drug usage



Exercise: figure out “ground truth”!

Exercise (ground truth on actual drug use)

Figure out a mechanism by which to compute a reasonably accurate count of drug use in various grids in Buffalo.

`Hint` Think of how to use an ML pipeline to predict drug usage based on the information above on NSDUH.

Click here for the procedure used by Lum and Isaac

- They first created a synthetic population. In other words, there created a dataset of individuals in Oakland such that the dataset is **demographically accurate**. Specifically, they represented each individual by their sex, household income, age, race, and the geo-coordinates of their home. These values were assigned so that the demographic aggregates of each component matched US census data to the most granular level possible.

- They used the NSDUH survey, which has information on the responder's demographic characteristics (i.e. sex, household income, age, and race) as well as their (reported) drug usage.

- They fit a model to predict an individual drug usage based on an individual demographic characteristics. In other words, they used the NSDUH as their training data set.

- They then used their model to predict drug usage on their synthetic dataset and add up the number of drug usage (i.e. number of drug crimes) in each cell of the grid.

Back to initial state question

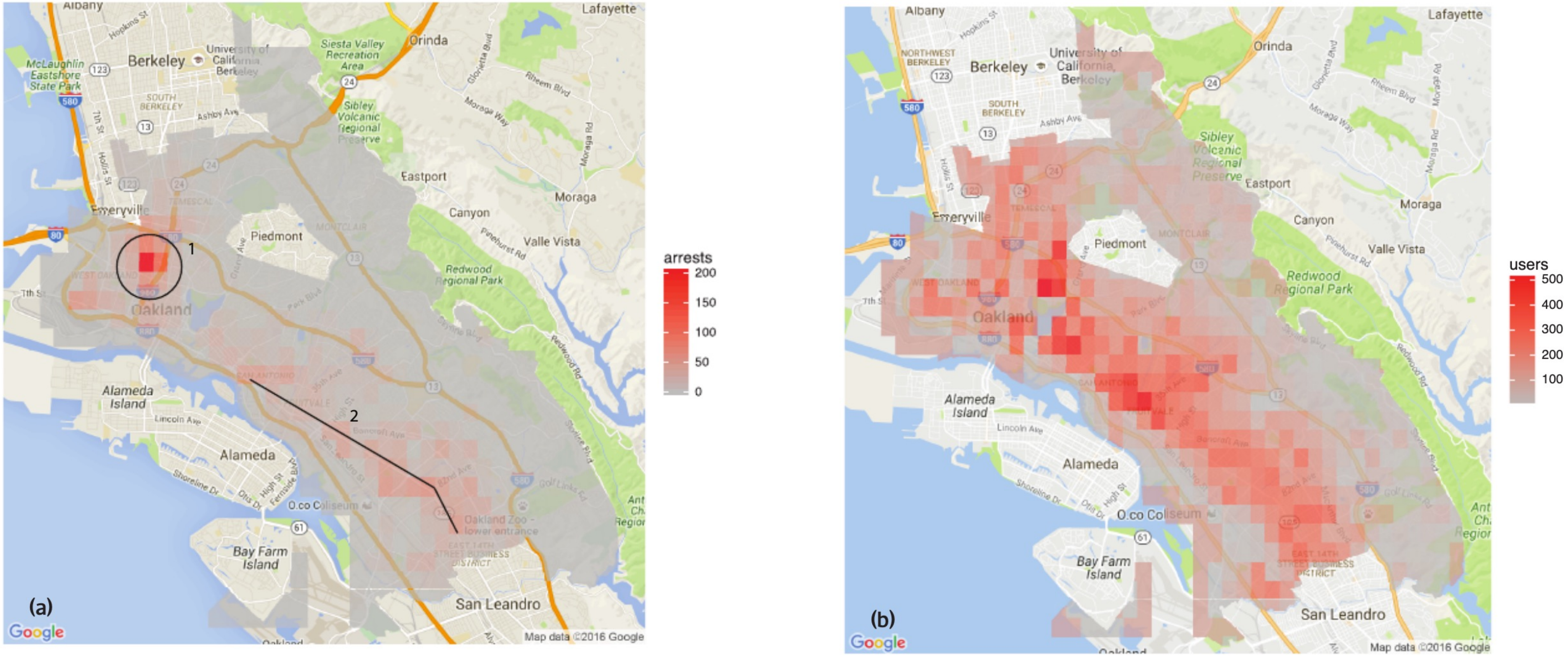
Exercise (initial state question)

Now that you have solved the issue of figuring out the ground truth, go ahead and solve the initial state question.

Click here for the answer

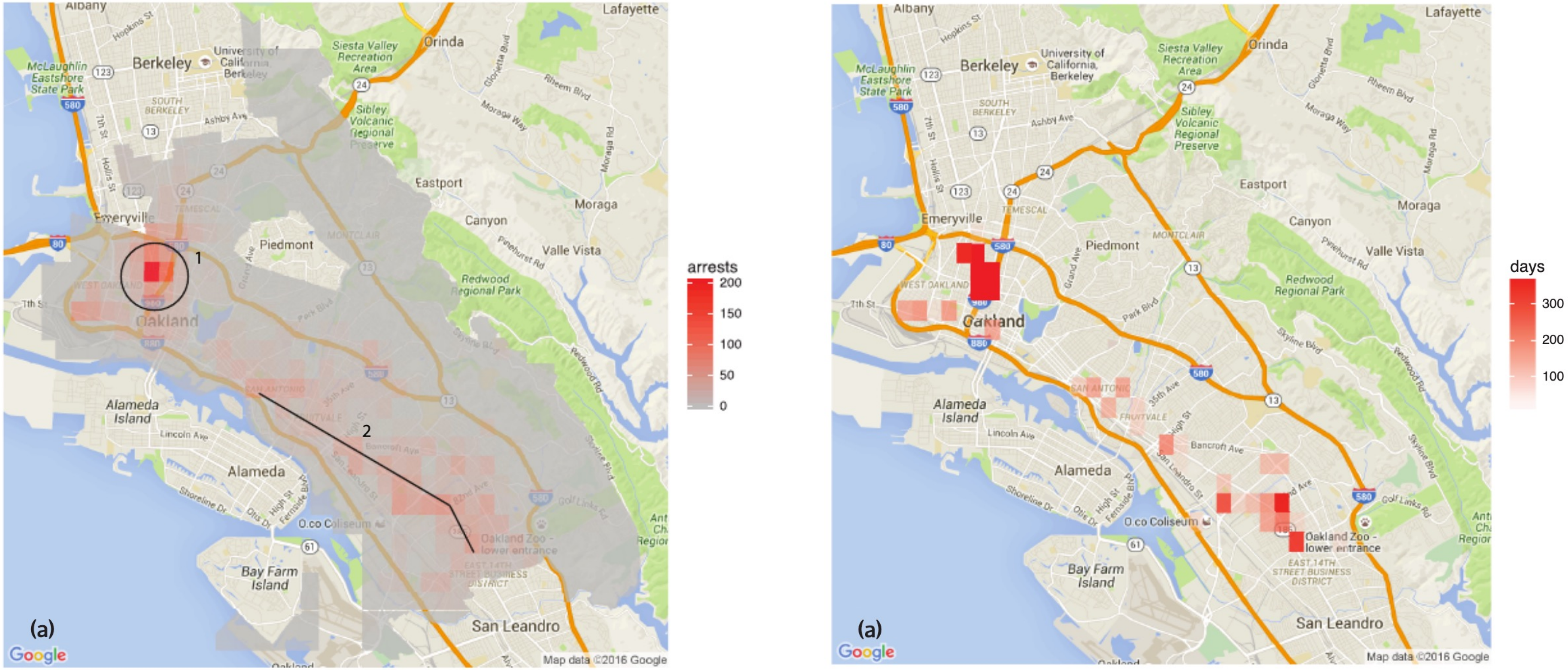
At this point we have all the information to answer the initial state question. In particular, we have the ground truth about drug crimes as well the police data on reported drug crimes on each grid cell. Since we also assumed we had access to predpol algorithm, we can run it on the reported drug crime data and see which grid cells predpol predicts will be high crimes areas. Then it is just a matter of comparing the results.

Lum and Isaac results: arrests vs ground truth



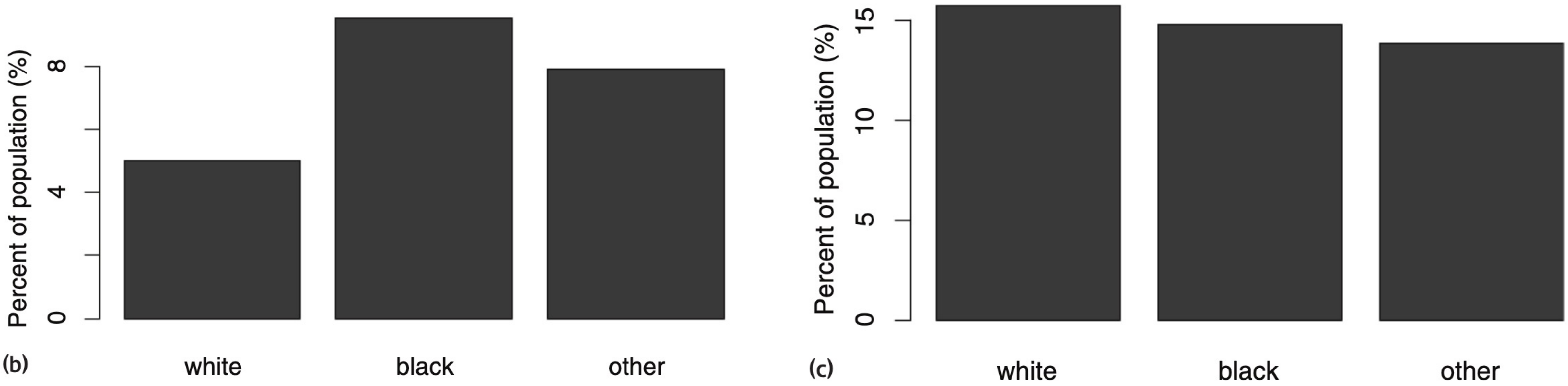
Figures taken from the Lum and Isaac paper

Lum and Isaac result: predpol output



Figures taken from the Lum and Isaac paper

Lum and Isaac result



Figures taken from the Lum and Isaac paper

Back to feedback loop state question

Feedback Loop State Question

Here the question is actually to test whether predictive policing could lead to feedback loop. The idea is to simulate the use of predpol in Buffalo from January 1, 2019 to December 31, 2019 in Buffalo and use the observed drug crime data for a given day to generate predpol's prediction for the next day. The goal here is to check if the the odds of location being "targeted" by predpol goes up (say when compared to the reported 2019 drug crimes as a baseline).

Exercise (using predpol only on historical data)

Consider the following simulation. From January 1, 2019 to December 31, 2019, we use the **original** drug crime data for a given day to generate predpol's prediction for the next day.

Do you expect the above simulation result to show that the odds of location being "targeted" by predpol go up (when compared to the reported 2019 drug crimes)? If so, why? If not, why not?

Click here for the answer

Since the effect of predpol prediction is not fed back into predpol, we should expect the odds of predpol targeting an area to track the historical reported drug crimes.

Now really back to the Q

Exercise (feedback loop state question)

How would you go about incorporating the effect of predpol's predictions into reported drug crimes? I.e. how would you incorporate the "feedback"?

Click here for how Lum and Isaac answered this question

One simple way to do this to increase the number of crimes reported in areas that predpol recommend that police be sent. E.g., one could bump up the reported drug crimes in a area by $x\%$ (for some value of x) in all areas were police were sent by predpol. In areas not targeted by predpol, use the historical reported drug crime data. The idea here is that if more police are sent to an area then more crimes will be reported/arrests will be made. This updated crimes data is then fed into predpol for get predictions for next day.

Lum and Isaac result (20% increase)

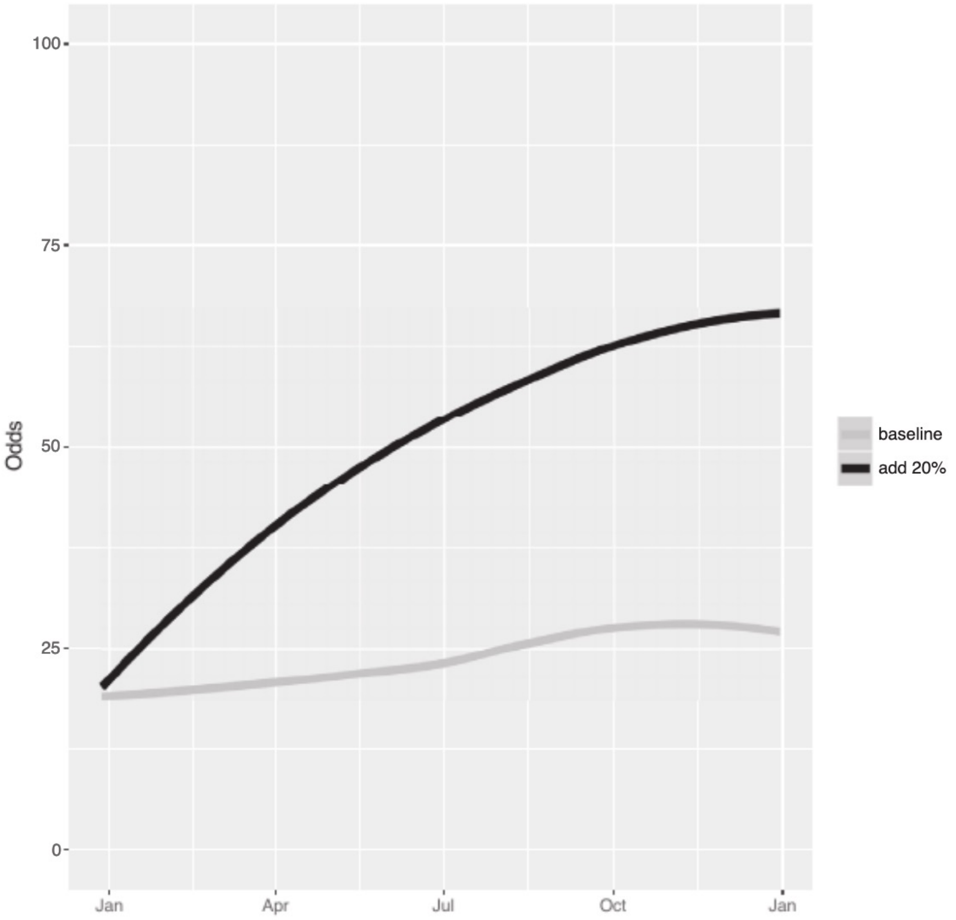


FIGURE 3 Predicted odds of crime in locations targeted by PredPol algorithm, relative to non-targeted locations. 'Baseline' is original Oakland police data. 'Add 20%' simulates the effect of additional crimes being observed in targeted locations

Figure taken from the Lum and Isaac paper

Passphrase for today: Cynthia Rudin

Cynthia Rudin

Professor of Computer Science, Electrical and Computer Engineering, and Statistical Science
PI, Prediction Analysis Lab
Duke University

Office: LSRC D342
Research Drive
Durham NC, 27708 USA

cynthia@cs.duke.edu

Research:

My research focuses on machine learning tools that help humans make better decisions, mainly interpretable machine learning. This also includes variable importance measures, causal inference methods, new forms of decision theory, uncertainty quantification, and methods that can incorporate domain-based constraints and other types of domain knowledge into machine learning. These techniques are applied to critical societal problems in criminology, healthcare, and energy grid reliability. The interpretable machine learning algorithms heavily rely on efficient discrete optimization techniques and Bayesian hierarchical modeling.

Bio:

Cynthia Rudin is a professor of computer science, electrical and computer engineering, and statistical science at Duke University, and directs the Prediction Analysis Lab, whose main focus is in interpretable machine learning. She is also an associate director of the Statistical and Applied Mathematical Sciences Institute (SAMSI). Previously, Prof. Rudin held positions at MIT, Columbia, and NYU. She holds an undergraduate degree from the University at Buffalo and a PhD from Princeton University. She is a three time winner of the INFORMS Innovative Applications in Analytics Award, was named as one of the “Top 40 Under 40” by Poets and Quants in 2015, and was named by Businessinsider.com as one of the 12 most impressive professors at MIT in 2015.

In-class discussion

Big Data's Disparate Impact

Solon Barocas* & Andrew D. Selbst**

Advocates of algorithmic techniques like data mining argue that these techniques eliminate human biases from the decision-making process. But an algorithm is only as good as the data it works with. Data is frequently imperfect in ways that allow these algorithms to inherit the prejudices of prior decision makers. In other cases, data may simply reflect the widespread biases that persist in society at large. In still others, data mining can discover surprisingly useful regularities that are really just preexisting patterns of exclusion and inequality. Unthinking reliance on data mining can deny historically disadvantaged and vulnerable groups full participation in society. Worse still, because the resulting discrimination is almost always an unintentional emergent property of the algorithm's use rather than a conscious choice by its programmers, it can be unusually hard to identify the source of the problem or to explain it to a court.

This Essay examines these concerns through the lens of American antidiscrimination law—more particularly, through Title

DOI: http://dx.doi.org/10.15779/Z38BG31

California Law Review, Inc. (CLR) is a California nonprofit corporation. CLR and the authors are solely responsible for the content of their publications.

Thing to keep in mind

You are expected to participate ☺

Discussion Participation

During the in-class discussion, y'all will form groups of size three (3) with perhaps one of two exceptions (to form groups of size two).

What happens in the group discussion

The goal of the group discussion is to come up with two top group responses for each part of the discussion summary: `Thoughts`, `Questions` and `Epiphanies`. *Ideally*, these responses should come from one of the group members discussion summary submission. However, it is OK to come up with a new response if e.g. if the group felt it would be better to synthesize the individual group member's responses.

After the group discussion is done, each group member will present two group responses. (It is up to the group on how to divide among the `Thoughts`, `Questions` and `Epiphanies`.) I will be keeping track of individual participation and you will be graded as follows.

Discussion participation grading rubric

- `Level 0`: No participation.
- `Level 1`: Exactly one non-trivial question asked or one non-trivial answer given.
- `Level 2`: At least two non-trivial questions asked or one non-trivial answers given.

What is a non-trivial question/answer?

I do not want to formally define what questions/answers are non-trivial since this is somewhat subjective. But just to give an idea: If the question was "What did you think about the paper assigned for today's in-class discussion?". An answer "Great!" will be considered trivial whereas a non-trivial answer would be one that goes into the specifics of what part(s) of the paper you though were great. Perhaps a better phrase for non-trivial would be *thoughtful*.

Discuss!

Big Data's Disparate Impact

Solon Barocas* & Andrew D. Selbst**

Advocates of algorithmic techniques like data mining argue that these techniques eliminate human biases from the decision-making process. But an algorithm is only as good as the data it works with. Data is frequently imperfect in ways that allow these algorithms to inherit the prejudices of prior decision makers. In other cases, data may simply reflect the widespread biases that persist in society at large. In still others, data mining can discover surprisingly useful regularities that are really just preexisting patterns of exclusion and inequality. Unthinking reliance on data mining can deny historically disadvantaged and vulnerable groups full participation in society. Worse still, because the resulting discrimination is almost always an unintentional emergent property of the algorithm's use rather than a conscious choice by its programmers, it can be unusually hard to identify the source of the problem or to explain it to a court.

This Essay examines these concerns through the lens of American antidiscrimination law—more particularly, through Title

---

DOI: http://dx.doi.org/10.15779/Z38BG31

Thoughts

Big Data's Disparate Impact

Solon Barocas* & Andrew D. Selbst**

Advocates of algorithmic techniques like data mining argue that these techniques eliminate human biases from the decision-making process. But an algorithm is only as good as the data it works with. Data is frequently imperfect in ways that allow these algorithms to inherit the prejudices of prior decision makers. In other cases, data may simply reflect the widespread biases that persist in society at large. In still others, data mining can discover surprisingly useful regularities that are really just preexisting patterns of exclusion and inequality. Unthinking reliance on data mining can deny historically disadvantaged and vulnerable groups full participation in society. Worse still, because the resulting discrimination is almost always an unintentional emergent property of the algorithm's use rather than a conscious choice by its programmers, it can be unusually hard to identify the source of the problem or to explain it to a court.

This Essay examines these concerns through the lens of American antidiscrimination law—more particularly, through Title

DOI: http://dx.doi.org/10.15779/Z38BG31

Questions + Epiphanies

Big Data's Disparate Impact

Solon Barocas* & Andrew D. Selbst**

Advocates of algorithmic techniques like data mining argue that these techniques eliminate human biases from the decision-making process. But an algorithm is only as good as the data it works with. Data is frequently imperfect in ways that allow these algorithms to inherit the prejudices of prior decision makers. In other cases, data may simply reflect the widespread biases that persist in society at large. In still others, data mining can discover surprisingly useful regularities that are really just preexisting patterns of exclusion and inequality. Unthinking reliance on data mining can deny historically disadvantaged and vulnerable groups full participation in society. Worse still, because the resulting discrimination is almost always an unintentional emergent property of the algorithm's use rather than a conscious choice by its programmers, it can be unusually hard to identify the source of the problem or to explain it to a court.

This Essay examines these concerns through the lens of American antidiscrimination law—more particularly, through Title

---

DOI: http://dx.doi.org/10.15779/Z38BG31

California Law Review, Inc. (CLR) is a California nonprofit corporation. CLR and the authors are solely responsible for the content of their publications.

Guest lecture on Tuesday: Kenny Joseph

There will be attendance!