RDS (DS-UA 202) Spring 2022: Homework 2 Solutions

Problem 1 (10 points): Racial disparities in predictive policing

(a) (5 points) Give **three distinct reasons** why racial disparities might arise in the predictions of such a system.

> Some possible answers are listed below. Any three valid reasons are sufficient for full credit.

1. The system uses historical data in which arrests in Black neighborhoods are overrepresented (relative to the number of drug-related crimes in those neighborhoods).
2. The system is deployed and adjusts its model based on day-to-day data, leading to feedback loops. More drug-related arrests in a neighborhood on one day leads to greater targeting of that neighborhood for drug-related policing activities on subsequent days.
3. Arresting officers may exhibit bias (often unconsciously) and arrest individuals more frequently in Black neighborhoods. Thus, even if police are allocated to other neighborhoods, they may be less likely to make arrests in those neighborhoods.
4. The system may be underutilizing other potentially important datasets and relying too much on historical arrest data.

(b) (5 points) Propose **two mitigation strategies** to counteract racial disparities in the predictions of such a system. Note: It is insufficient to state that we could use a specific pre-, in- or post-processing technique that we covered in class when we discussed fairness in classification. Additional details are needed to demonstrate your understanding of how the ideas from fairness in classification would translate to this scenario.

> Some possible mitigation strategies are listed below. Any two of these are sufficient for full credit.

1. Validate against external data sources, such as estimated drug use statistics that were used by Lum & Isaacs
2. Change policing strategies, for example, by adjusting police department goals. For example, the police could decide not to focus on drug-related crimes and focus on reducing violence-involved crimes. Or, the police could focus on interventions that do not involve making arrests.
3. Break the feedback loop, for example, by randomizing deployment of police vehicles or officers, capping the number of officers sent to Black neighborhoods, or capping the amount of time police are allowed to patrol Black neighborhoods.

Problem 2 (15 points): Randomized response

(a) (15 points) The simplest version of randomized response involves flipping **a single fair coin** (50% probability of heads and 50% probability of tails). Suppose an individual is asked a potentially incriminating question, and flips a coin before answering. If the coin comes up tails, he answers truthfully, otherwise he answers "yes". Is this mechanism differentially private? If so, what epsilon value does it achieve? *Carefully justify your answer.*

This mechanism is not differentially private because it does not randomize – and so does not provide plausible deniability – for all inputs. Specifically, if we observe the answer "yes" – we don't know for sure whether the truth is yes or no. But if we observe that the answer is "no" – we know for sure that the truth is no, and so there is no plausible deniability in this case. Additional details are below.

Let us denote Truth=Yes by P and Response=Yes by A. Observe that, since we are flipping a single coin, the individual whose true answer is Yes (P) will always answer yes (either because coin lands Tails, and so he responds truthfully, or because the coin lands heads and he responds Yes). That is Pr[A|P] = 1. Consequently, Pr[Not A|P] = 0, that is, if the truth is Yes, then the individual will never respond with No. On the other hand, if the true answer is No, then the individual has a 1/2 chance to respond with No (coin lands Tails) and 1/2 chance to respond with Yes (coin lands Heads).

Using the same notation as we used in class, we can write down our observations as follows:

Pr[A|P]=1	Pr[A| Not P] = ½
Pr[Not A| P] = 0	Pr[Not A | Not P] = 1

Based on this, epsilon = ln 2 when the answer is Yes. However, when the answer is No, Pr[Not A| P] < Pr[Not A | Not P] is not true for any value of epsilon.

Note: Answers that do not calculate epsilon are acceptable if the student correctly answered that the mechanism is not differentially private and carefully justified that answer.