Homework 2

Gurmehr Sohi, gs3541@nyu.edu

April 14, 2023

1 Question 1

1.1 Part.a

1. One way racial disparities can arise if historically there was always more arrests made in certain specific community (like Black neighbourhoods). If the historical data-set is already biased against them, it might lead to pre-existing bias and unfair system for further policing.

2. Second we can have racial unfairness due to factors like poverty, race, unemployment and education level which can cause the policing to do more policing towards certain demographic groups (Blacks) and lead to more drug arrests. In this the system can be unfair due to correlations between features.

3. Also racial bias can arise due to the system being created by bias individuals. If the individuals who created the system for prediction are already biased or unfair towards specific categories in features (like being unfair to a race, sex, education category) , it can cause racial disparities.

1.2 Part.b

1. One way to achieve fairness in this scenario would be to remove race as an input feature and instead include other variables that are highly predictive of drug activity. Alternatively, the system could be modified to actively seek out and correct for any bias that arises in its predictions. This could be achieved through techniques such as adversarial debiasing.

2. We can take regular feedback of the system usage (how correct the system is predicting) , interviewing police personals about their personal experiences with the usage of system. We can also reduce racial bias by removing unnecessary features that do not correlate to the drug usage and add new features to the data-set like more information about the neighbourhoods (hospital information regarding cases of drug overdose).

1.3 Part.c

1. Firstly the report by Lum Isaac are very limited as it shows only the data of 2010 and 2011. We need to analyse years before 2010 to get more in-depth dynamics of the arrests made. This is a very small sample to get a proper picture of the actual representation of the drug usage in other regions. Also the findings are only specific to Oakland and therefore cannot be justifiable in other areas. Their findings cannot be used to make prediction system in other regions which have different demographics, social and economic settings.

2. Secondly in their findings we cannot state that racial bias is the only reason for more drug arrests. We need more studies and in-depth analyses of the drug arrest and this might lead to finding more reasons for bias in arrests made in 2011.

2 Question 2

2.1 Part.a

If any category of individuals have a distribution different than the overall feature distribution, they will face a disadvantage when replacing null values with overall mean value.

In this case, the mean experience years for white females is 7.40 and non-white females is 7.91 are significantly higher than the overall mean experience years which is 6.12.

While we can see that males will get an advantage as their mean is smaller than the overall mean. This would create a bias in our final data set after imputation.

Thus, by imputing the missing values with the overall mean, Alex is assuming that the experience distribution is the same across all demographic groups, which is not the case in this data-set.

Our model would rank all females, especially non white females with the lower ranking.

2.2 Part.b

An appropriate imputation method would be to use the mean value for each demographic group separately. This would account for the differences in experience distribution across groups and provide a more accurate representation of an applicant's experience level.

Like for white males it would be better to use the overall mean of all white males and similarly for all other demographic groups.

This would also reduce any bias in the ranking model that can be created during imputation and would create a fair ranking model for all applicants.
This approach would result in a more precise and fair ranking of all applicants, as it accounts for the variation in experience distribution across demographic groups.

2.3 Part.c

Here in our case if we use the overall mean to impute values into the null values in our dataset, we can further extend a pre-existing bias that was already there in the dataset. This pre-existing bias is due to the under-representation of females in the dataset.

In our data set which has 4000 total individuals, we just have 934 females and 3066 males which is a huge imbalance and the cause of pre-existing bias. Due to this the overall mean is skewed towards the mean of males. This pre-existing bias is the representation of our world today we live in where women are always under-representated in jobs. Even though the mean experience of females is higher than males we see a biased being introduced during imputation and females being ranked lower.

Emergent bias can occur during model training on imputed data. If the imputed data is already mis-representating the experience levels of individuals, it would incorrectly train the model. The trained model would be biased and this would further create more unfair outcomes for future predictions.

3 Question 3

3.1 Part.a

In class we saw that for this flip coin problem we have the following steps:

1. flip a coin C1
   - if C1 is tails, then respond truthfully
2. if C1 is heads, then flip another coin C2
   - if C2 is heads then Yes
   - else C2 is tails then respond No

1. Truth=Yes by P
2. Response=Yes by A
3. C1=tails by T
4. C1=heads and C2=tails by HT
5. C1=heads and C2=heads by HH

Now we know that a Randomized algorithm M provides ϵ (epsilon is a privacy parameter) if, for all neighboring databases D1 and D2, and for any set of outputs S :

$$Pr[M(D1) \in S] <= e^{\epsilon} Pr[M(D2) \in S]$$

$$Pr[A|P] = Pr[T] + Pr[HH] = 3/4$$

$$Pr[A|\ P] = Pr[HH] = 1/4$$

$$Pr[A|P] = 3Pr[A|\ P]$$

$$\epsilon = ln(3)$$

This shows that is $ln(3)$ - differentially private.

3.2 Part.b

In this part we have 2 dices D1 and D2 with faces labelled 1,2,3,4,5,6. The mechanism here is

1. Roll a dice
   - if value is smaller than 4 (1,2,3) we will respond truthfully.
2. Otherwise we will roll dice D2
   - if D2 is smaller than 3 (1,2) we respond yes
   - otherwise we respond no

Probabilities

1. Truth=Yes by P
2. Response=Yes by A
3. D1= smaller than 4 by S
4. D1 = bigger than or equal 4 and D2 = smaller than 3 by BS
5. D1 = bigger than or equal 4 and D2 = bigger than or equal 3 by BB

Now we know that a Randomized algorithm M provides ϵ (epsilon is a privacy parameter) if, for all neighboring databases D1 and D2, and for any set of outputs S :

$$Pr[M(D1) \in S] <= e^{\epsilon} Pr[M(D2) \in S]$$

$$Pr[A|P] = Pr[S] + Pr[BB] = 3/6 + (3/6 * 2/6) = 1/2 + 1/6 = 4/6$$

$$Pr[A|\ P] = Pr[BB] = (3/6 * 2/6) = 1/6$$

$$Pr[A|P] = 4Pr[A|\ P]$$

$$\epsilon = ln(4)$$

This shows that is $ln(4)$ - differential private.

3.3 Part.c

Now we know that a Randomized algorithm M provides ϵ (epsilon is a privacy parameter) if, for all neighboring databases D1 and D2, and for any set of outputs S :

$$Pr[M(D1) \in S] <= e^{\epsilon} Pr[M(D2) \in S]$$

$$Pr[BB] = 3/6 * (1/12 + 1/12) = 3/6 * 1/6 = 3/36 = 1/12$$

$$Pr[A|P] = Pr[S] + Pr[BB] = 3/6 + 1/12 = 6 + 1/12 = 7/12$$

$$Pr[A|\ P] = Pr[BB] = 1/12$$

$$Pr[A|P] = 7Pr[A|\ P]$$

$$\epsilon = ln(7)$$

This shows that is $ln(7)$ - differentially private. This is more than the epsilon value obtained in the previous part.b and therefore it would guarantee less privacy to the respondent because lower epsilon value gives stronger privacy.

4 Question 4

4.1 Question 4.a