DDA4210 - 24 Spring Assignment Guru: Fanzeng Xia

Assignment 3 Solution

1. Suppose we have the following structural causal model. Assume all exogenous variables are independent and that the expected value of each is 0. 20 points

$$V = \{X, Y, Z\}, \quad U = \{U_X, U_Y, U_Z\}, \quad F = \{f_X, f_Y, f_Z\} \tag{1}$$

$$f_X : X = U_X \tag{2}$$

$$f_Y : Y = \frac{X}{3} + U_Y \tag{3}$$

$$f_Z : Z = \frac{Y}{16} + U_Z \tag{4}$$

1. Draw the graph that complies with the model. 5 points

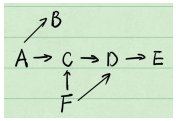

Figure 1: SCM

2. Determine the best guess of the value (expected value) of Z, given that we observe Y = 3. 5 points

   Answer:

$$E[Z|Y=3] = E[\frac{Y}{16} + U_Z|Y=3] = \frac{1}{16}E[Y|Y=3] + E[U_Z|Y=3] = \frac{3}{16} + 0 \tag{5}$$

$$= \frac{3}{16} \tag{6}$$

3. Determine the best guess of the value of Z, given that we observe X = 3. 5 points **Answer:**

$$E[Z|X=3] = E[\frac{1}{16}Y + U_Z|X=3] = E[\frac{1}{16}(\frac{1}{3}X + U_Y)|X=3] = \frac{1}{16} \tag{7}$$

4. Determine the best guess of the value of Z, given that we observe X = 1 and Y = 3. 5 points

Answer:

$$E[Z|X=1,Y=3] = E[\frac{1}{16}Y + U_Z|X=1,Y=3] \tag{8}$$
$$= \frac{1}{16}E[Y|Z=1,Y=3] + E[U_Z|X=1,Y=3] \tag{9}$$
$$= \frac{3}{16} + 0 = \frac{3}{16} \tag{10}$$

2. In the central model of differential privacy, a trusted curator aggregates all data and randomizes responses to queries. However, in the local model of differential privacy, users do not trust the aggregator, so they randomize their data locally before sending it to the aggregator. Consider the Randomized Response (RR) mechanism, proposed by Warner in 1965, aiming to collect sensitive statistics while providing each participant some deniability.

Each of the n users holds a private bit $b_i \in 0, 1$, and we want to estimate the average $a := \frac{1}{n}\sum_{i=1}^{n} b_i$. The RR mechanism, executed independently by each user, consists of flipping two unbiased coins and following these steps:

If the first coin is heads, send bi to the aggregator. Otherwise, look at the second coin:

- If heads, send 0 to the aggregator.
- If tails, send 1 to the aggregator.

30 points

1. Demonstrate that the RR mechanism ensures ϵ-differential privacy for each user's individual bit, with $\epsilon = \ln(3)$. 5 points

Answer:

For the RR mechanism, we need to consider the maximum ratio between the probabilities of sending 1 given the true bit is 1 or 0:

$$\frac{P(\hat{b}_i = 1|b_i = 1)}{P(\hat{b}_i = 1|b_i = 0)} = e^{\ln 3}$$

Therefore, the RR mechanism ensures ϵ-differential privacy with $\epsilon = \ln(3)$.

2. Let $\hat{b}_i$ be the i-th user's randomized response. Prove that the untrusted aggregator that receives all these noisy bits can compute an unbiased estimate $\hat{a}$ of a (i.e., $\mathbb{E}[\hat{a}] = a$). 5 points

Answer:

Consider the random variable $\hat{b}_i$ which represents the randomized response of user i. Thus, we have the conditional expected value of $\hat{b}_i$ given b_i as

$$\mathbb{E}[\hat{b}_i|b_i = 1] = 1 * \mathbb{P}(\hat{b}_i = 1|b_i = 1) + 0 * \mathbb{P}(\hat{b}_i = 0|b_i = 1) = \frac{3}{4}$$
$$\mathbb{E}[\hat{b}_i|b_i = 0] = 1 * \mathbb{P}(\hat{b}_i = 1|b_i = 0) + 0 * \mathbb{P}(\hat{b}_i = 0|b_i = 0) = \frac{1}{4}$$

Therefore, we have

$$\begin{aligned}\mathbb{E}[\hat{b}_i] &= \mathbb{E}[\hat{b}_i|b_i = 1] * \mathbb{P}(b_i = 1) + \mathbb{E}[\hat{b}_i|b_i = 0] * (1 - \mathbb{P}(b_i = 1)) \\ &= P(b_i = 1) \times 3/4 + P(b_i = 0) \times 1/4 \\ &= a \times 3/4 + (1 - a) \times 1/4 \\ &= a/2 + 1/4\end{aligned}$$

So we can estimate a as

$$\hat{a} = \frac{2}{n}\sum_{i=1}^{n}\hat{b}_i - \frac{1}{2}$$

which is unbiased as $\mathbb{E}[\hat{a}] = a$.

3. Show that the estimation error $\hat{a} - a$ has a standard deviation of $O(\frac{1}{\sqrt{n}})$. 10 points

Answer:

Since it is clear that $Var(\hat{b}_i)$ is $O(1)$ and

$$\text{Var}[\hat{a}] = \text{Var}[\frac{2}{n}\sum_{i=1}^{n}\hat{b}_i - \frac{1}{2}] = \frac{4}{n}\text{Var}[\hat{b}_i]$$

Thus, by the Central Limit Theorem, the standard deviation of $\hat{a} - a$ is $O(1/\sqrt{n})$.

4. Compare this result with the central model, where all users send their bits b_i to a trusted curator using the Laplace mechanism to output a noisy estimate $\hat{a}c$ of a that is ln(3)-differentially private. Show that the estimation error $\hat{a}c - a$ has a standard deviation of $O(\frac{1}{n})$. 10 points

Answer:

In the central model, the curator adds Laplace noise with scale parameter $\Delta f/\epsilon$ to the true average a, where Δf is the sensitivity of the query function. The sensitivity of computing the average of bits is 1/n, so the scale parameter of the Laplace noise is $(1/n)/\ln(3)$. The standard deviation of the Laplace distribution is $\sqrt{2}$ times the scale parameter, so the standard deviation of the estimation error is $O(\frac{1}{n})$.

5. Design a generalized RR mechanism that provides ϵ-differential privacy for each user's individual bit, for any fixed $\epsilon > 0$. Show that your mechanism satisfies ϵ-DP in the local model, and that the standard deviation of the untrusted aggregator's estimation error is $O(\frac{1}{\epsilon\sqrt{n}})$. Bonus: 5 point

Answer:

To achieve ε-differential privacy, we can adjust the probabilities of the coins. Let $p = e^{\varepsilon}/(1+e^{\varepsilon})$ and $q = 1 - p$. The user sends their bit with probability p and a random bit with probability q. It can be shown that this mechanism satisfies ϵ-DP.

$$\frac{P(\hat{b}_i = 1|b_i = 1)}{P(\hat{b}_i = 1|b_i = 0)} = \frac{e^{\varepsilon}/(1+e^{\varepsilon})}{1/(1+e^{\varepsilon})} = e^{\varepsilon}$$

The variance of the user's response is $O(1)$, so by the Central Limit Theorem, the standard deviation of the aggregator's estimation error is $O(1/\epsilon\sqrt{n})$.

For references, please check:

http://researchers.lille.inria.fr/abellet/teaching/ppml_lectures/lec6.pdf

3. Suppose you have two algorithms to predict whether a student can obtain an offer from Harvard University based on some features such as GPA, number of credits, internship, and research experience. The following table shows the ground-truth label and the predictions made by the two algorithms. The sensitive attribute considered in this fairness study is gender.

a Determine if the algorithms satisfy demographic parity. Show the derivation.

b Determine if the algorithms satisfy equal opportunity. Show the derivation.

c Determine if the algorithms satisfy equalized odds. Show the derivation. 15 points

Gender	Label (truth)	Algorithm 1	Algorithm 2
Male	Yes	No	Yes
Male	No	Yes	Yes
Male	Yes	Yes	No
Male	Yes	No	No
Female	Yes	Yes	Yes
Female	No	No	Yes
Female	Yes	No	No
Female	Yes	Yes	No

Answer:

(a) Algorithm I and II: demographic parity holds.

(b) Algorithm 1: equal opportunity does not hold; Algorithm II: equal opportunity holds.

(c) Algorithm I: equalized odds does not hold; Algorithm II: equalized odds holds.

4. In machine learning, fairness is an important consideration when building models that are used to make decisions affecting people. Two common fairness metrics are demographic parity and equalized odds. Demographic parity requires that the classification rates (e.g., acceptance rates) for different demographic groups should be equal or close to equal. Equalized odds requires that the true positive rate (TPR) and false positive rate (FPR) are equal across groups. Consider a binary classification task for a loan approval system. The system predicts whether to approve or reject a loan application based on several features. You are given the following information about the loan applicants:

- There are 1000 applicants in total
- 600 applicants belong to Group A and 400 applicants belong to Group B
- 200 applicants from Group A are approved for loan
- The overall loan approval rate for all applicants is 25%
- The true positive rate (TPR) for Group A is 50%
- The false positive rate (FPR) for Group A is 20%

Your task is to analyze fairness in this loan approval system using demographic parity and equalized odds. Answer the following questions: 30 points

1. Calculate the loan approval rate for Group A. 5 points

   Answer:

   The loan approval rate for a group is the number of loans approved for that group divided by the total number of applicants from that group.

   Given that 200 applicants from Group A are approved for loans and there are 600 applicants in Group A, the loan approval rate for Group A is 200 / 600 = 0.333 or 33.3%.

2. Calculate the loan approval rate for Group B. 5 points

   Answer:

   First, we need to determine the total number of loans approved. The overall loan approval rate is 25%, and there are 1000 total applicants, so the total number of loans approved is 0.25 * 1000 = 250.

   Given that 200 of these approved loans are for Group A, the remaining 50 must be for Group B. Therefore, the loan approval rate for Group B is 50 / 400 = 0.125 or 12.5%.

3. Determine whether demographic parity is achieved in this loan approval system. If not, what is the difference in approval rates between Group A and Group B? 10 points

   Answer:

   Demographic parity is achieved when the loan approval rates for different groups are equal or close to equal. In this case, the loan approval rate for Group A is 33.3%, while the approval rate for Group B is 12.5%. Thus, demographic parity is not achieved. The difference in approval rates between Group A and Group B is 33.3% - 12.5% = 20.8%.

4. Calculate the true positive rate (TPR) and false positive rate (FPR) for Group B, assuming equalized odds are achieved. 10 points

   Answer:

   If equalized odds are achieved, then the true positive rate (TPR) and false positive rate (FPR) for Group B should be the same as those for Group A. Therefore, the TPR for Group B is 50% and the FPR for Group B is 20%.

5. Suppose you want to achieve both demographic parity and equalized odds by adjusting the loan approval thresholds for Group A and Group B. Calculate the number of applicants from each group that need to be approved for loans to achieve both demographic parity and equalized odds while keeping the overall loan approval rate constant. If it is not possible to achieve both demographic parity and equalized odds simultaneously, explain why.

Answer:

To achieve demographic parity, the approval rates for Group A and Group B must be the same. Given that the overall approval rate is 25%, both groups must have this approval rate to achieve demographic parity. This means 150 (*i.e.*, $0.25 * 600$) applicants from Group A and 100 (*i.e.*, $0.25 * 400$) applicants from Group B should be approved.

However, It's impossible to determine the exact figures without knowing the proportion of truly eligible applicants in each group.

5. For a set of d players represented by $D = \{1, \ldots, d\}$ and a cooperative game $v : 2^d \mapsto \mathbb{R}$, the Shapley value for each player $i \in D$ is defined as:

$$\phi_i(v) = \sum_{S \subseteq D \backslash \{i\}} \frac{|S|!(d - |S| - 1)!}{d!} (v(S \cup \{i\}) - v(S)) \tag{11}$$

(a) Prove the null player axiom of the Shapley value, which states that if a player contributs no value in a game $v : 2^d \mapsto \mathbb{R}$, or $v(S \cup \{i\}) - v(S) = 0$ for all $S \subseteq D \backslash \{i\}$, then $\phi_i(v) = 0$. 3 points

Answer:

It clearly follows from the formula of the Shapley value that

$$\phi_i(v) = 0$$

for all i such that $v(S \cup \{i\}) = v(S)$ for all $S \subseteq D \backslash \{i\}$.

(b) Prove the linearity axiom of the Shapley value, which states that given two cooperative games $u : 2^d \mapsto \mathbb{R}$ and $v : 2^d \mapsto \mathbb{R}$ and a third game w defined as $w(S) = u(S) + v(S)$ for all $S \subseteq D$, the following holds for all players $i \in D$: 5 points

$$\phi_i(w) = \phi_i(u) + \phi_i(v)$$

Answer:

$$\begin{aligned}
\phi_i(w) &= \sum_{S \subseteq D \backslash \{i\}} \frac{|S|!(d - |S| - 1)!}{d!} (w(S \cup \{i\}) - w(S)) \\
&= \sum_{S \subseteq D \backslash \{i\}} \frac{|S|!(d - |S| - 1)!}{d!} ((u(S \cup \{i\}) + v(S \cup \{i\})) - (u(S) + v(S))) \\
&= \sum_{S \subseteq D \backslash \{i\}} \frac{|S|!(d - |S| - 1)!}{d!} (u(S \cup \{i\}) - u(S)) - \sum_{S \subseteq D \backslash \{i\}} \frac{|S|!(d - |S| - 1)!}{d!} (v(S \cup \{i\}) - v(S)) \\
&= \phi_i(u) + \phi_i(v)
\end{aligned}$$

(c) Prove the efficiency axiom of the Shapley value, which states that the following holds for all games $v : 2^d \mapsto \mathbb{R}$: 8 points

$$\sum_{i=1}^{d} \phi_i(v) = v(D) - v(\phi).$$

Answer:

Recall that the Shapley value for player i is defined as

$$\phi_i(v) = \sum_{S \subseteq D \backslash \{i\}} \frac{|S|!(d - |S| - 1)!}{d!} (v(S \cup \{i\}) - v(S))$$

Now, let's sum over all players:

$$\sum_{i=1}^{d} \phi_i(v) = \sum_{i=1}^{d} \sum_{S \subseteq D \backslash \{i\}} \frac{|S|!(d - |S| - 1)!}{d!} (v(S \cup \{i\}) - v(S))$$

Then, we can swap the summations and factor out the common terms:

$$\sum_{i=1}^{d} \phi_i(v) = \sum_{S \subseteq D} \sum_{i \in S} \frac{|S \backslash \{i\}|!(d - |S \backslash \{i\}| - 1)!}{d!} (v(S) - v(S \backslash \{i\}))$$

By the properties of the characteristic function v, we know that $v(S \backslash i) = 0$ for $i \notin S$, so we can rewrite the above as:

$$\sum_{i=1}^{d} \phi_i(v) = \sum_{S \subseteq D} \frac{|S|!(d - |S| - 1)!}{d!} v(S)$$

If you sum over all subsets S of D, you'll find that each S appears exactly $d!$ times (once for each ordering of players). Thus, the coefficients cancel out:

$$\sum_{i=1}^{d} \phi_i(v) = \sum_{S \subseteq D} v(S) = v(D)$$

which is the efficiency axiom of the Shapley value.

(d) Prove that the Null Player, Symmetry, and Linearity axioms can be replaced by a single property

$$\phi_i(D) - \phi_i(D\backslash\{j\}) = \phi_j(D) - \phi_j(D\backslash\{i\})$$

for all $i, j \in D$ with $i \neq j$ where $\phi_i(D\backslash\{j\})$ denotes the Shapley value of the player i with player j removed. Bonus: 5 points

Answer:

Prove it by definition or directly check the orginal article:

https://www.jstor.org/stable/1911054

6. We consider some example of cooperative games and calculate their Shapley values.

(a) Define a specific characteristic function of a cooperative game:

Table 1

S	ϕ	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	$\{1,2,3\}$
$v(S)$	0	2	3	4	5	6	7	8

Calculate the Shapley value for all players $i \in \{1,2,3\}$ for the following cooperative game characterized by $v(S)$. 5 points

Answer:

For the player 1, we have the computation table as in Table 3.

Table 2: Intermediate computation for the Shapley value of node 1

S	$\lvert S\rvert$	d	$d - \lvert S\rvert - 1$	$\frac{\lvert S\rvert!(d-\lvert S\rvert-1)!}{d!}$	$S \cup \{1\}$	$v(S \cup \{1\})$	$v(S)$	$v(S \cup \{1\}) - v(S)$
ϕ	0	3	2	$\frac{1}{3}$	$\{1\}$	2	0	2
$\{2\}$	1	3	1	$\frac{1}{6}$	$\{1,2\}$	5	3	2
$\{3\}$	1	3	1	$\frac{1}{6}$	$\{1,3\}$	6	4	2
$\{2,3\}$	2	3	0	$\frac{1}{3}$	$\{1,2,3\}$	8	7	1

Hence, the Shapley value for the player 1 is

$$\phi_1(v) = \frac{1}{3} \cdot 2 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 2 + \frac{1}{3} \cdot 1 = \frac{5}{3}.$$

Similarly, we have the Shapley values for players 2 and 3 as follows.

$$\phi_2(v) = \tfrac{8}{3} \quad \phi_3(v) = \tfrac{11}{3}$$

(b) Calculate the Shapley value for all players $i \in \{1, 2, 3\}$ in the game $v(S)$ given by 5 points

$$v(S) = 2x_1 + 3x_2 + 4x_3.$$

where x_i are binary variables that are equal to 1 if $i \in S$ and 0 otherwise.

Answer:

For the player 1, we have the computation table as in Table 3.

Table 3: Intermediate computation for the Shapley value of node 1

S	$\lvert S\rvert$	d	$d - \lvert S\rvert - 1$	$\frac{\lvert S\rvert!(d-\lvert S\rvert-1)!}{d!}$	$S \cup \{1\}$	$v(S \cup \{1\})$	$v(S)$	$v(S \cup \{1\}) - v(S)$
ϕ	0	3	2	$\frac{1}{3}$	$\{1\}$	2	0	2
$\{2\}$	1	3	1	$\frac{1}{6}$	$\{1, 2\}$	5	3	2
$\{3\}$	1	3	1	$\frac{1}{6}$	$\{1, 3\}$	6	4	2
$\{2, 3\}$	2	3	0	$\frac{1}{3}$	$\{1, 2, 3\}$	9	7	2

Hence, the Shapley value for the player 1 is

$$\phi_1(v) = \frac{1}{3} \cdot 2 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 2 + \frac{1}{3} \cdot 2 = 2.$$

Similarly, we have the Shapley values for players 2 and 3 as follows.

$$\phi_2(v) = 3 \quad \phi_3(v) = 4$$

(c) Calculate the Shapley value for all players $i \in \{1, 2, 3, 4, 5\}$ in the game $v(S)$ given by 5 points

$$v(S) = 2x_2 + 3x_3 + 4x_4 + 5x_1x_3 + 7x_2x_5 - 12x_1x_2x_3.$$

where x_i are binary variables that are equal to 1 if $i \in S$ and 0 otherwise.

Hint: write a Python function to calculate the Shapley value automatically, using the formula in Eq. 1.

Answer:

For the player 1, we have the computation table as in Table 4.

Hence, the Shapley value for the player 1 is

Table 4: Intermediate computation for the Shapley value of node 1

S	$\lvert S\rvert$	d	$d-\lvert S\rvert-1$	$\frac{\lvert S\rvert!(d-\lvert S\rvert-1)!}{d!}$	$S\cup\{1\}$	$v(S\cup\{1\})$	$v(S)$	$v(S\cup\{1\})-v(S)$
ϕ	0	5	4	$\frac{1}{5}$	$\{1\}$	0	0	0
$\{2\}$	1	5	3	$\frac{1}{20}$	$\{1,2\}$	2	2	0
$\{3\}$	1	5	3	$\frac{1}{20}$	$\{1,3\}$	8	3	5
$\{4\}$	1	5	3	$\frac{1}{20}$	$\{1,4\}$	4	4	0
$\{5\}$	1	5	3	$\frac{1}{20}$	$\{1,5\}$	0	0	0
$\{2,3\}$	2	5	2	$\frac{1}{30}$	$\{1,2,3\}$	-2	5	-7
$\{2,4\}$	2	5	2	$\frac{1}{30}$	$\{1,2,4\}$	6	6	0
$\{2,5\}$	2	5	2	$\frac{1}{30}$	$\{1,2,5\}$	9	9	0
$\{3,4\}$	2	5	2	$\frac{1}{30}$	$\{1,3,4\}$	12	7	5
$\{3,5\}$	2	5	2	$\frac{1}{30}$	$\{1,3,5\}$	8	3	5
$\{4,5\}$	2	5	2	$\frac{1}{30}$	$\{1,4,5\}$	4	4	0
$\{2,3,4\}$	3	5	1	$\frac{1}{20}$	$\{1,2,3,4\}$	2	9	-7
$\{2,3,5\}$	3	5	1	$\frac{1}{20}$	$\{1,2,3,5\}$	5	12	-7
$\{2,4,5\}$	3	5	1	$\frac{1}{20}$	$\{1,2,4,5\}$	13	13	0
$\{3,4,5\}$	3	5	1	$\frac{1}{20}$	$\{1,3,4,5\}$	12	7	5
$\{2,3,4,5\}$	4	5	0	$\frac{1}{5}$	$\{1,2,3,4,5\}$	9	16	-7

$$\phi_1(v)=\frac{1}{20}\cdot 5+\frac{1}{30}\cdot(-7)+\frac{1}{30}\cdot 5+\frac{1}{30}\cdot 5+\frac{1}{20}\cdot(-7)+\frac{1}{20}\cdot(-7)+\frac{1}{20}\cdot 5+\frac{1}{5}\cdot(-7)=-\frac{3}{2}.$$

Similarly, we have the Shapley values for players 2, 3, 4 and 5 as follows.

$$\phi_2(v)=\tfrac{3}{2}\quad \phi_3(v)=\tfrac{3}{2}$$
$$\phi_4(v)=4\quad \phi_5(v)=\tfrac{7}{2}$$