Lab Assignment 4

Machine Learning Due on: 23rd April 2024 11:59 PM | M.M.:50 CS 503

Instructions:

1. The research assignment is for students not doing the project.
2. This lab assignment needs to be done individually, and it carries 13% weightage of your total lab requirements.
3. Start by going through the research paper first before you start coding.
4. For queries, you should comment on Google Classroom. Alternatively, you may email the course TAs or course instructors.
5. All code must be written in python using a jupyter/colab notebook.
6. Submission is only through Google Classroom. The code and the accompanying observations must be included in the same jupyter/colab notebook for each task separately. You can also prepare a separate report pdf if you want but observations for each task should be highlighted in separate sections.
7. Code Readability is very important. Modularize your code using classes and functions that can be flexibly reused wherever necessary. Also, use self-explanatory variable names and add comments to describe your approach wherever necessary.
8. **Students are expected to follow the honour code of the class**. Discussions and interactions with your classmates to take help in this assignment are not allowed.

This assignment offers two options. The first focuses on fairness in machine learning, while the second focuses on explainable AI. You may select one of these directions to pursue. If you opt for the clustering task, you must exclusively complete all tasks in that section. Similarly, if you choose explainable AI, you should only answer tasks related to that topic. Mixing topics is not permitted.

Option 1: Fairness in Unsupervised Machine Learning

Advances in machine learning research have resulted in the development of increasingly accurate models. While improving the accuracy is the primary objective of these algorithms, their use to allocate social goods and opportunities such as access to healthcare and job and educational opportunities warrants a closer look at the societal impacts of their outcomes. Recent studies have exposed a discriminatory outlook in the outcomes of these algorithms, leading to treatment disparity towards individuals belonging to marginalized groups based on gender and race in real-world applications like automated resume processing, loan application screening, and criminal risk prediction. Designing fair and accurate machine learning models is thus an essential and immediate requirement for these algorithms to make a meaningful real-world impact.

Lab Assignment 4

To emphasize the importance of fairness in unsupervised learning, we consider the following hypothetical scenario: An employee-friendly company is looking to open branches at multiple locations across the city and distribute its workforce in these branches to improve work efficiency and minimize overall travel time to work. The company has employees with diverse backgrounds based on, for instance, race and gender and does not prefer any group of employees over other groups based on these attributes. The company's diversity policy dictates hiring a minimum fraction of employees from each group in every branch. Thus, the natural question is: where should the branches be set up to maximize work efficiency, minimize travel time, and maintain diversity. In other words, the problem is to devise an unsupervised learning algorithm for identifying branch locations with the fairness (diversity) constraints applied to each branch. This problem can be naturally formulated as a clustering problem with additional fairness constraints on allocating the data points to the cluster centers.

0.1 Fairness Notion for Group Fair Clustering

Typically, fairness is measured by the algorithm's performance over different groups based on protected(sensitive) attributes such as gender, race, and ethnicity. The first fairness notion for clustering was proposed by Chierichetti et al. [2], wherein each cluster is required to exhibit a Balance defined as follows :

Definition 1 (Balance [2])**.** *It is the ratio of protected attribute and non-protected attribute in each cluster to the level of this ratio in the entire dataset. Mathematically, for a binary-valued protected attribute $S = \{a, b\}$, balance of clustering $\mathcal{C}$ with clusters $\{C_1, C_2, \ldots, C_k\}$ is computed as*

$$Balance = \min_{C_j \in \mathcal{C}} \left(\frac{\sum_{x \in C_j} \mathbb{I}(S(x) = a)}{\sum_{x \in C_j} \mathbb{I}(S(x) = b)}, \frac{\sum_{x \in C_j} \mathbb{I}(S(x) = b)}{\sum_{x \in C_j} \mathbb{I}(S(x) = a)} \right) \tag{1}$$

Here, x refers to points assigned in cluster C_j where j ranges from 1 to k and $\mathbb{I}$ is an indicator function which outputs one whenever the condition inside is evaluated true. So here $\sum_{C_j \in \mathcal{C}} \mathbb{I}(S(x) = a)$, counts the number of points belonging to group value a in cluster C_j.

Balance can be extended to multi-valued protected groups, i.e., groups taking more than two values. This is a simple extension where we take minimum over all possible pairs of group values, i.e.,

$$Balance = \min_{C_j \in \mathcal{C}} \left(\frac{\#a}{\#b}, \frac{\#b}{\#a}, \frac{\#c}{\#b}, \frac{\#b}{\#c}, \frac{\#c}{\#a}, \frac{\#a}{\#c} \right) \tag{2}$$

Here, $\#a$ is shorthand for the number of points in cluster j for group value a.

Also, to evaluate fairness, there is another metric called fairness error, which is defined as follows:

Definition 2 (Fairness Error [6]). *It is the Kullback-Leibler (KL) divergence between the required protected group proportion and achieved proportion within the clusters. Here, the required group proportion is the dataset composition of different protected group values and achieved is the obtained composition of different group values in a cluster.*

For example, if the dataset has 100 red points and 50 blue points, the algorithm A outputs two clusters, each of which has 25 blue and 50 red points. Then, the required proportion of the dataset is $50/100 = 1/2$, and the achieved proportion in each cluster is also $1/2$, so the fairness error is zero.
Next, to estimate the quality of clustering, we define clustering cost as follows:

Definition 3 (Clustering Cost [3]). *Given X and an assignment function $\phi : X \to C$ that assigns each point $x \in X$ to a center. The objective cost of clustering is*

$$L_p(X, \phi) = \left(\sum_x d(x, \phi(x))^p \right)^{1/p} \tag{3}$$

The cases with $p = \{1, 2, \infty\}$ norms represent standard k-median, k-means, and k-center vanilla clustering respectively.

0.2 Task 1

Note that, in summary, existing fairness metrics were designed in a way that they were able to provide us with Linear programming-based or optimization-based approaches. No polynomial time algorithm could achieve group fair clustering for multi-valued protected groups. To this, the work in [3] propose a new notion of fairness, which they call τ-ratio fairness, that strictly generalizes the Balance property to multi-valued protected groups. This notion also helped the authors propose a polynomial time algorithm called $FRAC_{OE}$.

Definition 4 (τ-ratio fairness [3]). *A clustering $\mathcal{C}$ is said to obey τ-ratio fairness, if for all $a \in S, C_j \in \mathcal{C}$,*

$$\sum_{x \in C_j} \mathbb{I}(S(x) = a) \geq \tau_a \sum_{x \in X} \mathbb{I}(S(x) = a). \tag{4}$$

Here, X is the complete dataset. The intuition behind the notion is that we need to divide the $\tau_a * k$ fraction of points from group value a in the dataset equally among all clusters. This will ensure that the minimum criteria of $\tau_a\%$ is fulfilled.
Algorithm description: For algorithmic idea, refer to Algorithm 1 in paper [3].

0.2.1 Question [Reproducibility Study]:

In this task you need re-run the code provided for $FRAC_{OE}$ in folder Paper 1 of clustering. You need to reproduce the results for k-means for $FRAC_{OE}$ as (a) shown in Figure 7, that is, clustering cost, balance, and fairness error over varying k, and (b) shown in Figure 8

the same study over varying dataset size (n). You must run code only for the Adult and Bank datasets as they require lower computational requirements. You must find mean and standard deviation over 3 independent runs, i.e., three different seed values. **[5 Marks each = 10 Marks]**

0.2.2 Question [Further Study 1]:

In this task you need run the code provided for $FRAC_{OE}$ in folder Paper 1 of clustering for k-center objective cost. The k-center cost is computed by substituting the value of p as infinity, which resolves to calculating the following:

$$\text{Cost } k\text{-center} = \max_{C_j \in \mathcal{C}} \max_{x \in X} d(x, \phi(x)) \tag{5}$$

This means it is the maximum of the farthest distance in each cluster.

Your task is to compute the cost, balance and fairness error value for $k = 10, 20, 40$ for the Adult and Bank dataset and plot these results for the complete dataset size. Thus, to solve this part, you need to write down code for calculating k-center cost; the rest of the same code could be used. **[7.5 Marks]**

0.3 Task 2

Recently, group fair clustering algorithms have been studied under data label poisoning attack [1]. So, the threat model is as follows: the adversary (attacker) can control a small portion of individuals' protected group memberships (either through social engineering, exploiting a security flaw in the system, etc.); by changing their protected group memberships the adversary aims to disrupt the fairness utility of the fair algorithm on other uncontrolled groups. So, in this task, you need to do the following:

0.3.1 Question [Attack]:

In adult dataset randomly flip (0 to 1 and vice versa) gender column for p% of rows. Then find the cost, balance, and fairness error value of $FRAC_{OE}$ at $k = 10$. Vary p as $1, 2, 5, 10, 20\%$. You must also print the number of males and females in the modified dataset (after flipping) for each p. **[7.5 Marks]**

0.4 Task 3

Group fairness does not ensure fair treatment for a particular individual. The trait of human envy might still make an individual discontented. For example, an employee might feel discriminated against or left out if similar employees (who may not belong to the same group value) receive a favourable appraisal. There are algorithms in the literature that guarantee individual fairness [4]. One of the notions of individual fairness is that for each data point, x center should be within a fixed radius r. This r is the radius of the ball drawn with point

x as center such that the ball contains n/k points (including x). Sometimes, this notion is relaxed to having center for each data point x within $\alpha \cdot r$ where $\alpha \geq 1$.

To evaluate individual fairness, usually, we measure the number of data points for which the fairness condition is violated, i.e., the center is not within $\alpha \cdot r$ distance [5]. We call this, say, violation number V.

The code for paper [5] is provided in the Paper 2 folder of clustering. Your task is the following:

0.4.1 Question [Individual Fairness]:

You need to run the code on Adult dataset for $k = 10$ on complete dataset. The α parameter is computed using an inbuilt pre-coded binary search which sets α such that exactly k centers are produced by code. You just need to check how to input the dataset file and reproduce the results. Report the default metrics, i.e., cost and individual fairness, computed by code. These are named output_init_fair, output_init_cost, output_ls_fair, and output_ls_cost in the code. **[10 Marks]**

0.5 Task 4

My preliminary research has observed that a certain level of individual fairness gets automatically satisfied by running a group fair clustering algorithm ($FRAC_{OE}$), i.e., it achieves a balanced clustering with a certain level of individual fairness. This implies that if you consider $\alpha = 1$, the number of violations V over the group fair clusters is quite small. So, your task reduces to the following:

0.5.1 Question [Novelty]:

Can you come up with an algorithmic idea of how one can design an algorithm which takes input the τ_a for all group values $a \in S$ and also takes input α and in return outputs cluster that is satisfying both levels of fairness. You can even give descriptive ideas, and providing preliminary results will be an added advantage, i.e., showing balance and violations metric V. **[15 Marks]**

or

Can you modify the existing $FRAC_{OE}$ so that it becomes somewhat robust to label perturbations? You can present descriptive ideas as well. Again, experimental testing can be an added advantage.

[15 Marks]

Option 2: Explainability in Machine Learning

Paper title- Head Matters: Explainable Human-centered Trait Prediction from Head Motion Dynamics

Github link: https://github.com/MonikaGahalawat11/Head-Matters--Code

Task 1

Replicate the findings: Execute Figure 6 representing the feature fusion of head and facial feature modalities and specifically focus on Table 3, which displays FICS Regression results, focusing solely on columns 2 (AU only), 3(Kineme Only), and 4(Feature fusion).

[20 Marks]

Task 2

Consider exploring additional modalities such as eye-gaze, audio, or transformer-based representations of video frames to enhance prediction accuracy in the proposed model. For instance, if opting to work on audio, explore alternative methodologies apart from those presented in the referenced work (https://arxiv.org/pdf/2302.09817.pdf). Along with modality, focus on building new fusion methodology for performance enhancement.

[15 Marks]

Task 3

Emphasize the development of a new explainability dimension within the paper, focusing on providing detailed explanations for the results obtained. **[15 Marks]**

References

[1] A. Chhabra, P. Li, P. Mohapatra, and H. Liu. Robust fair clustering: A novel fairness attack and defense framework. In *The Eleventh International Conference on Learning Representations*, 2022.

[2] F. Chierichetti, R. Kumar, S. Lattanzi, and S. Vassilvitskii. Fair clustering through fairlets. In *NeurIPS*, pages 5036–5044, 2017.

[3] S. Gupta, G. Ghalme, N. C. Krishnan, and S. Jain. Efficient algorithms for fair clustering with a new fairness notion. *arXiv:2109.00708*, 2021.

[4] S. Mahabadi and A. Vakilian. Individual fairness for k-clustering. In *International conference on machine learning*, pages 6586–6596. PMLR, 2020.

[5] S. Mahabadi and A. Vakilian. Individual fairness for k-clustering. In *ICML*, pages 6586–6596, 2020.

[6] I. M. Ziko, J. Yuan, E. Granger, and I. B. Ayed. Variational fair clustering. In *AAAI*, pages 11202–11209, 2021.