CSC2421 Spring'24
Assignment 2 Solutions

Q1 [15 Points] Diversity

[This is a non-technical question with no correct answer. Any reasonable answer will receive at least 80% marks, and particularly insightful answers will receive more.]

A commonly advocated position is that companies should hire employees that come from diverse backgrounds (not only in terms of race and gender, but also in terms of birthplace, educational discipline, etc.), with the argument that this will result in the AI systems developed by the companies being more "ethical".

Consider Netflix hiring employees to build a classifier, which classifies each user into "high-value" vs "low-value", depending on how much revenue Netflix is expected to make from the user over the next year. This will in turn be used for decision-making by a number of programs within Netflix (marketing, recommendations, etc.). Netflix wants the classifier to be fair with respect to gender.

Describe some mechanisms by which hiring a diverse team of employees can result in the eventual classifier built being "fairer". For this question, do not restrict yourself to interpreting "fairness" as a concrete definition (e.g., equalized odds) under a specific modeling; after all, the team of employees can affect how the problem is modeled in the first place. But please do interpret "fairness" to still refer to the general aspect of ethics that a definition like equalized odds aims to capture.

Solution to Q1

I have created marking criteria for this on MarkUs.

- (3/15) Blank or poor: The question was left blank or the answer makes no sense.
- (12/15) Reasonable: Lays out at least one plausible mechanism, though the answer is not particularly innovative.
- (14/15) Good: Lays out one or more plausible mechanisms that are described reasonably well.
- (15/15) Excellent: Lays out multiple mechanisms, describes in detail how they are realized, includes innovative ideas.

Q2 [25 Points] Fair Classification

A company designs a classifier $h : \mathcal{X} \to \{0, 1\}$, which takes in an input x describing an applicant (e.g., their work experience, educational experience, race, gender, etc.) and outputs a binary label $y = h(x)$ indicating whether the applicant should be made an offer ($y = 1$) or not ($y = 0$). The company wants the classifier to be fair with respect to the gender of the applicant. For this problem, assume the company is using binary gender (M/F).

The ML engineering team has worked hard to gather data from 40 past applicants along with their "true labels", indicating whether they were truly qualified for the job. Here is a summary:

Gender \ Qualified	Yes	No	Total
M	16	12	28
F	8	4	12

(a) [15 Points] Eight engineers ended up designing eight different classifiers. Below, you can see the number of applicants from each of the four categories in the above table that these classifiers would make offers to in the training data. We use the abbreviation Q = qualified and U = unqualified. First, write down what demographic parity and equalized odds constraints are using fractions written in plain words such as $\frac{\text{\# of accepted QM applicants}}{\text{\# of accepted applicants}}$. Then, determine which of the following classifier satisfies each of demographic parity and equalized odds (no need to show your calculation).

- h_1 makes an offer to 7 QM and 3 QF applicants. No unqualified applicants are made an offer.
- h_2 makes an offer to 4 QM, 2 QF, 12 UM applicants.
- h_3 makes an offer to 4 QM, 2 QF, 9 UM, 3 UF applicants.
- h_4 makes an offer to 7 UM and 3 QF applicants.
- h_5 makes an offer to all QM and QF applicants. Like h_1, no unqualified applicants are made an offer.
- h_6 makes an offer to all UM and UF applicants. No qualified applicants are made an offer!
- h_7 makes no offers.
- h_8 makes an offer to exactly half of the applicants from each of the 4 categories (QM, QF, UM, UF), thus essentially acting as a random coin flip classifier.

(b) [10 Points] Describe all classifiers that achieve both demographic parity and equalized odds on the training data by fully characterizing the set of 4-tuples of # of offers that can be made to applicants from the four categories (QM,QF,UM,UF).

Solution to Q2

(a) Demographic parity:

$$\frac{\text{\# of accepted M applicants}}{\text{\# of M applicants}} = \frac{\text{\# of accepted F applicants}}{\text{\# of F applicants}}$$

Equalized odds:

$$\frac{\text{\# of accepted QM applicants}}{\text{\# of QM applicants}} = \frac{\text{\# of accepted QF applicants}}{\text{\# of QF applicants}}$$

and

$$\frac{\text{\# of accepted UM applicants}}{\text{\# of UM applicants}} = \frac{\text{\# of accepted UF applicants}}{\text{\# of UF applicants}}$$

Satisfaction of these notions among the given classifiers:

- DP: h_1, h_4, h_7, h_8
- EO: h_3, h_5, h_6, h_7, h_8

(b) Let (QM,QF,UM,UF) denote the number of accepted applicants from those corresponding four categories. For demographic parity, we need:

$$\frac{QM + UM}{28} = \frac{QF + UF}{12}$$

For equalized odds, we need:

$$\frac{QM}{16} = \frac{QF}{8} \text{ and } \frac{UM}{12} = \frac{UF}{4}.$$

The latter two give us $QM = 2QF$ and $UM = 3UF$. Substituting them in the demographic parity equation gives $QF = 2UF$, so $QM = 4UF$. Hence, we get that

$$QM : UM : QF : UF = 4 : 3 : 2 : 1.$$

This means the possible tuples are

$$(0, 0, 0, 0), (4, 3, 2, 1), (8, 6, 4, 2), (12, 9, 6, 3), \text{ and } (16, 12, 8, 4).$$

The higher tuples are infeasible due to there not being enough applicants in any of the categories.

Q3 [25 Points] Fair Clustering

Recall the clustering problem from class. We are given a metric space $M = (X, d)$, where d is a distance metric (satisfying the triangle inequality) over a set of points X, a set of $P \subseteq X$ with $|P| = n$, and a desired number of cluster centers k. Our goal is to find the set of locations $C \subseteq X$ of $|C| = k$ cluster centers. Once C is chosen, each data point $x \in P$ is assigned to its nearest cluster center (pardon the slight abuse of notation) $C(x) = \arg\min_{c \in C} d(x, c)$.

Recall that a clustering C is said to be in the α-core if there is no set of data points $S \subseteq P$ with $|S| \geq n/k$ and a new center $c' \in X$ such that $\alpha \cdot d(x, c') < d(x, C(x))$ for all $x \in S$.

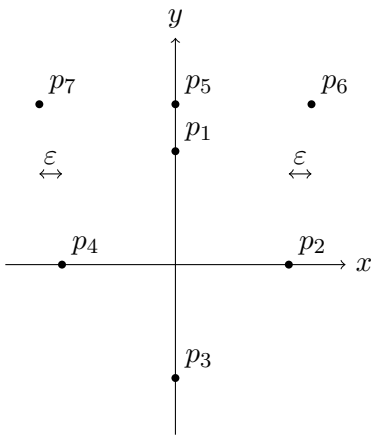
Recall the Greedy Capture algorithm: it grows balls at each point in X simultaneously and with equal speed. As soon as any ball captures at least n/k data points from P, its center c is added to C. Subsequently, any data point captured by an already existing ball centered at a location in C is assigned to that ball, and for any new c' to be added to C, its ball must capture at least n/k data points from P that are not already captured by any ball centered at any location in C. The process terminates when all the points in P are captured. In class, you saw a proof that Greedy Capture satisfies $(1 + \sqrt{2})$-core for any metric.

(a) [15 Points] Prove that it satisfies 2-core for the metric $(\mathbb{R}^p, d = L^2)$ (i.e., L^2 norm in the p-dimensional Euclidean space) for any p. For this, assume the following fact about this metric: For any $x, y \in \mathbb{R}^p$, the set of points $\{z \in \mathbb{R}^p : 2 \cdot d(z, y) \leq d(z, x)\}$ is a ball of radius $(2/3) \cdot d(x, y)$.

[Hint: Suppose for contradiction that the clustering returned is not in 2-core, so there exists a violation S, c' as per the definition. Let $i \in S$ be the first point of S that was captured (say, by a ball centered at $c \in C$) during the execution of Greedy Capture. Use the fact provided to find a ball of radius less than $(2/3) \cdot d(c, c')$ that must contain all of S. Prove $(2/3) \cdot d(c, c') < d(i, c)$ to derive a contradiction.]

(b) [10 Points] Now, consider the metric $(\mathbb{R}^2, d = L^1)$ (i.e., L^1 norm in the 2-dimensional Euclidean space). Consider the following 7 points:

$p_1 = (0, 1), p_2 = (1, 0), p_3 = (0, -1), p_4 = (-1, 0),$
$p_5 = (0, \sqrt{2}), p_6 = (1 + \epsilon, \sqrt{2}), p_7 = (-1 - \epsilon, \sqrt{2}).$



Consider an instance with $n = 28$ points divided into four sets of 7 points, each set isomorphic to the set of 7 points above, and each set located sufficiently far from all the other sets. Let $k = 7$. Prove that the approximation ratio to the core achieved by Greedy Capture on this instance is $1 + \sqrt{2}$ as $\epsilon \to 0$. That is, the improvement we see for L^2 in part (a) does not apply to L^1.

[Hint: There will be one of the four sets of points where Greedy Capture will only place one cluster center

at its $(0,0)$. In this set of points, find four that can deviate and be significantly better.]

Solution to Q3

(a) Suppose for contradiction that the clustering is not in 2-core, and there exist $S \subseteq P$ and $c' \in X$ such that $2 \cdot d(x,c') < d(x,C(x))$ for all $x \in S$.

Let $i \in S$ be the first point of S that was captured during the execution of Greedy Capture, and suppose it was captured by a ball centered at $c \in C$.

Note that for all $x \in S$, we have

$$2 \cdot d(x,c') < d(x,C(x)) \leq d(x,c).$$

So, $S \subseteq \{z \in \mathbb{R}^p : 2 \cdot d(z,c') \leq d(z,c)\}$, and using the given fact, there must be a ball of radius at most $(2/3) \cdot d(c,c')$ that contains it.

We now prove that $(2/3) \cdot d(c,c') < d(i,c)$, which yields the desired contradiction: when i was captured by a ball centered at c (which must have had radius at least $d(i,c)$), all the points in S were uncaptured, and since $|S| \geq n/k$ and there was a ball of radius *less than* $d(i,c)$ that contained them, this ball must have been selected by Greedy Capture instead, regardless of any tie-breaking. To see the desired inequality,

$$\begin{aligned} d(c,c') &\leq d(c,i) + d(i,c') && (\because \text{ triangle inequality}) \\ &< d(i,c) + \frac{1}{2} d(i,c) && (\because \text{ violation of 2-core}) \\ &= (3/2) d(i,c). \end{aligned}$$

Rearranging yields the desired inequality.

(b) Note that $n/k = 28/7 = 4$. It can be verified that Greedy Capture will place the first four cluster centers at $(0,0)$ of the four copies that covers their $\{p_1, p_2, p_3, p_4\}$. Regardless of where it places the next three centers, at least one copy of the 7 points will be sufficiently far away from these additional three centers. In this copy, we can consider a deviation by the points $\{p_1, p_5, p_6, p_7\}$.

With a center at p_5, the improvements for the four points would be as follows:

$$\begin{aligned} p_1 : &\frac{d(p_1,(0,0))}{d(p_1,p_5)} = \frac{1}{\sqrt{2}-1} = 1+\sqrt{2}, \\ p_5 : &\frac{d(p_5,(0,0))}{d(p_5,p_5)} = \infty, \\ p_6 : &\frac{d(p_6,(0,0))}{d(p_6,p_5)} = \frac{1+\epsilon+\sqrt{2}}{1+\epsilon} \xrightarrow{\epsilon \to 0} 1+\sqrt{2}, \\ p_6 : &\frac{d(p_7,(0,0))}{d(p_7,p_5)} = \frac{1+\epsilon+\sqrt{2}}{1+\epsilon} \xrightarrow{\epsilon \to 0} 1+\sqrt{2}. \end{aligned}$$

Thus, as $\epsilon \to 0$, this shows an improvement by a factor of $1+\sqrt{2}$, so the approximation ratio to the core achieved by Greedy Capture on this instance is no better than $1+\sqrt{2}$. The reason it is equal to $1+\sqrt{2}$ is because we have seen in class that it is at most $1+\sqrt{2}$ on any instance (for any metric).