MATH70098 Ethics in Data Science and AI

Exercise Sheet - Multi-objective optimisation and Pareto Fronts

Zak Varty

Exercise sheet: Multi-objective optimisation and Pareto Fronts

The questions on this sheet are designed to let you test your own understanding of the course content on multi-objective optimisation and Pareto fronts. Some questions will test basic notions, while others will encourage you to think more deeply about some of the concepts introduced this week.

Q1: Wiki contributions

The Wikipedia article on multi-objective optimisation provides a brief but comprehensive introduction to the topic. The "Examples of applications" section of this page lists application areas for multi-objective optimisation including economics, finance and optimal design. Along with each example application is a short one or two paragraph explanation of how multi-objective optimisation can be used in that area. Ethical AI is not currently on this list.

Write one to two paragraphs, in language suitable for this format, describing how multi-objective optimisation may be applied in the context of ethical AI.

Model Solution:

Ethical AI aims to understand, quantifiy and make explicit the ethical implications of artificial intelligence and other data-lead methods. It also aims to develop methods which lead to outcomes that are in some sense more ethical.

Standard AI problems can often be viewed as single-objective optimisation, where some loss function (such as predictive accuracy) is maximised. This is extended by ethical AI to introduce additional objectives, for example to ensure that the resulting algorithm treats different sub-groups fairly or maintains the anonymity of individuals within the dataset.

Multi-objective optimisation can be used to balance the trade-off between these desirable properties, which are often in conflict with one another. For example, it is known that not all fairness metrics may be satisfied at once, and so MOO could be used to find the optimal model for an individual's given set of preferences over these metrics.

Q2: Formal definitions of a Pareto-optimal and a dominated solution

Formally define and explain in plain language what is meant by a dominated and a Pareto-optimal solution to the multi-objective optimisation problem:

$$\max_{x=(x_1,\dots,x_n)\in X} g(f_1(x), f_2(x), \dots f_k(x)) \text{ s.t. } x \in \mathcal{X}.$$

Solution:

The set $\mathcal{X} \subseteq X$ denotes the set of all feasible solutions to the optimisation problem. Consider two feasible solutions to the optimisation problem $x, x^* \in \mathcal{X}$.

The solution x^* is said to dominate the solution x (or x is dominated by x^*) if and only if x^* is at least as good for each individual objective as x, and improves on at least one objective as compared to x. Formally:

- $f_i(x^*) \geq f_i(x)$ for all $i \in \{1, \ldots, k\}$;
- $f_i(x^*) > f_i(x)$ for some $i \in \{1, \ldots, k\}$.

A Pareto-optimal solution is a feasible solution that is not dominated by any other feasible solution. Formally, $x' \in \mathcal{X}$ is Pareto optimal if and only if for all $x \in \mathcal{X} \setminus x'$

- $f_i(x') \geq f_i(x)$ for all $i \in \{1, \ldots, k\}$;
- $f_i(x') > f_i(x^)$ for some $i \in \{1, \ldots, k\}$.

Q3: A concrete example

Consider the size classifiers given in the table below.

|	classifier	k_anon	equalised_odds	error_rate	false_positive_rate
1	A	2	FALSE	0.11	0.51
2	B	4	TRUE	0.13	0.22
3	C	3	TRUE	0.05	0.61
4	D	5	FALSE	0.20	0.17
5	E	3	TRUE	0.13	0.86
6	F	2	TRUE	0.08	0.64

(i) Consider the four pairs of objectives given below. Visualise the the bi-objective performance of the six classifiers for each of these pairs. Identify which classifiers are Pareto-optimal and which are dominated in each case.

(a) false positive rate vs error rate,
(b) k-anonymity vs equalised odds,
(c) k-anonymity vs error rate,
(d) Equalised odds vs error rate.

(ii) When considering all four objectives at once, which (if any) of the classifiers are Pareto optimal?

Solution

(i) It is conventional to present Pareto fronts so that each objective is to be minimised. This means that the preferred points are toward the bottom-left of the plot. Since "good" values of k-anonymity and equalised odds take larger values, we first transform these before plotting.

(It would be equally valid to consider accuracy = 1 - error rate, in which case dominant solutions would be to the top-right of the plots.)

(a) false positive rate vs error rate
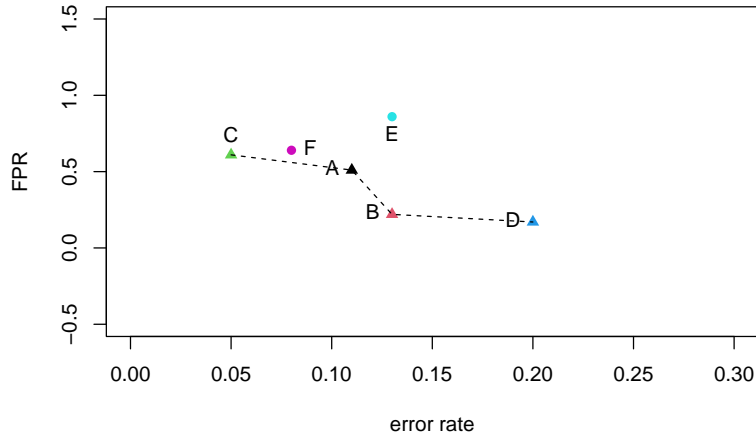
In this plot the classifiers E and F are dominated (shown as circles) and the remaining classifiers are Pareto optimal (shown as triangles). Because error rate and false positive rate are continuous valued quantities, we can add a linear interpolation of the Pareto-optimal points to aid the visualisation of the Pareto-frontier.

(b) k-anonymity vs equalised odds,
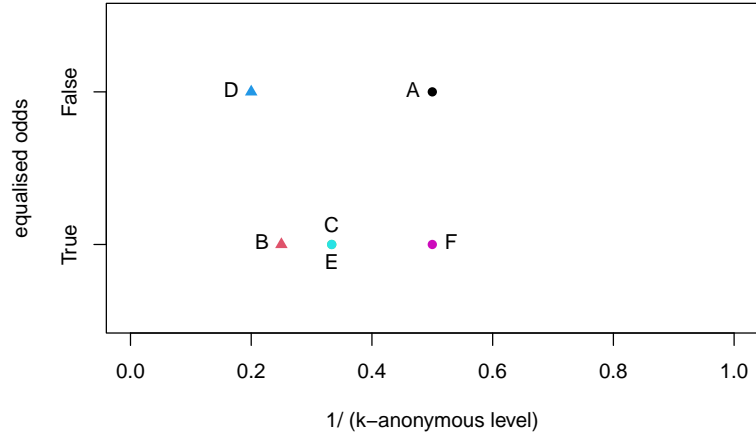
In this case we can see that classifiers B and D dominate the other three solutions. Without knowing the end-users preferences between fairness and anonymity we can not say which should be preferred.

Note that it does not make sense to have fractional levels of k-anonymity or to partially achieve equalised odds. (Note: the interested reader may want to research epsilon-equalised odds, which relaxes equalised odds condition.)

(c) k-anonymity vs error rate,
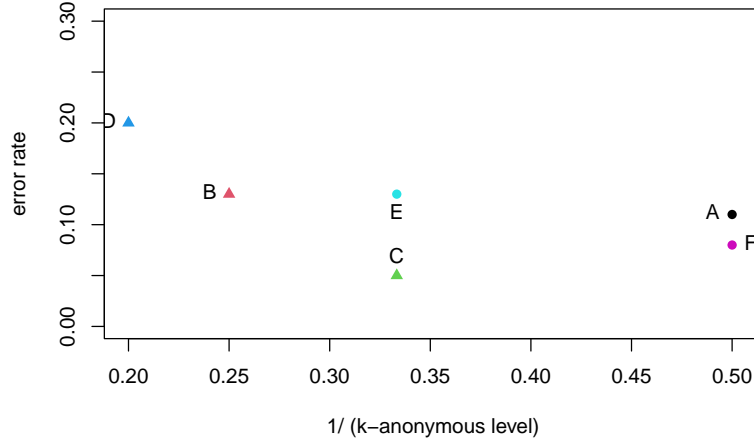
In this comparison, classifiers D B and C are all Pareto-optimal and dominate the remaining classifiers.

(d) Equalised odds vs error rate
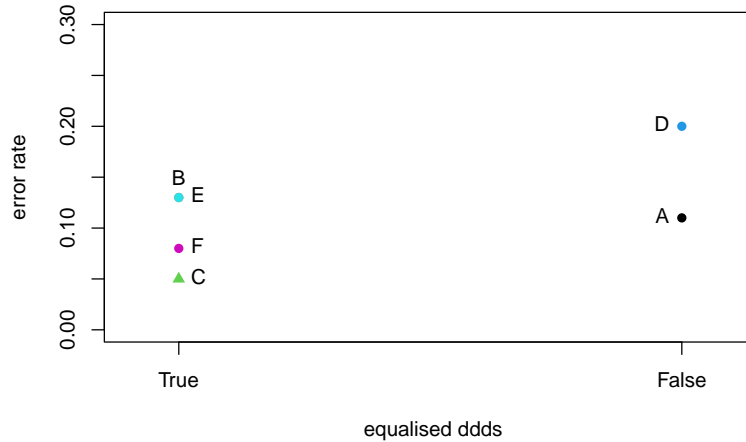
In this final comparison, classifier C dominates all other classifiers.

(ii) When considering all four outcomes none of the classifiers are dominated and so all are Pareto Optimal.

A is non-dominated. (B is worse on error rate, C is worse on FPR, D is worse on equalised odds, E is worse on error rate, and F is worse on false positive rate.)

B is non-dominated. (A,C,E and F are worse than B on k-anonymity. D is worse than B on error rate.)

C is non-dominated. (All classifiers are worse than C on error rate.)

D is non-dominated. (All classifiers are worse than D on k-anonymity.)

E is dominated by both B and C. (A and F have worse k-anonymity, D does worse than E on equalised odds. However, B is better than E on k-anonymity and false positive rate, and at least as good on all other objectives. Similarly, C is better than E on error rate and false positive rate, and at least as good on all other objectives)

F is dominated by C. (A, B D and E are worse than F on error rate. C is at least as good as F in all attributes and is better on k anonymity, error rate and false positive rate.)

Therefore our Pareto front is the set of Pareto optimal classifiers: $\{A, B, C, D\}$.

Q4: Scalarisation

(a) Define in your own words what it means to scalarise a multi-objective optimisation problem.

4

(b) Linearisation and epsilon-constraint are two methods of linearisation. Explain in plain language how each method works.

(c) Give linearised and epsilon-constrained formulations for the multi-objective optimisation in Question 2.

Solution:

A scalarisation of a multi-objective optimisation problem is any method that reduces the dimensionality of the response surface to a single dimension, such that the maxima/minima in that single-objective optimisation correspond to Pareto-optimal solutions from the original problem.

Linearisation is one method which scalarises a multi-objective optimisation problem by replacing a collection of separate objective functions by a weighted sum of those objectives.

Epsilon-constrained linearisation reduces multiple objectives into a single objective by selecting one of the objective functions to be the primary objective and replacing each of the other objectives with a constraint. Specifically, this is the constraint that each objective should have a (potentially objective specific) value less/greater than some specified value ϵ.

The optimisation in Question 2 may be linearised using a vector of individual-specific weights $w = (w_1, \ldots, w_k) \in \mathbb{R}^k$ as:

$$\max_{x=(x_1,\ldots,x_n)\in X} \{w_1 f_1(x) + w_2 f_2(x) + \ldots + w_k f_k(x)\} \text{ s.t. } x \in \mathcal{X}.$$

The optimisation in Question 2 may be reformulated as an epsilon-constrained single objective problem by selecting $j \in \{1, \ldots, k\}$ to be the primary objective and introducing constraint parameters $\epsilon_i \in \mathbb{R}$ for $i \in \{1, \ldots, k\} \setminus \{j\}$. The resulting scalarisation is given by:

$$\max_{x=(x_1,\ldots,x_n)\in X} f_j(x) \quad \text{such that} \quad x \in \mathcal{X} \quad \text{and} \quad f_i(x) > \epsilon_i \text{ for } i \in \{1, \ldots, k\} \setminus \{j\}.$$