Assignment 2

Data-driven Optimization and Decision-making Course (Spring 2024)

max 30 points, Deadline: May 3, 2024

You have been provided with the dataset described during the second part of Lecture 5 (Real world application), taken from B. S. Saini, D. Chakrabarti, N. Chakraborti, B. Shavazipour, and K. Miettinen, "Interactive data-driven multiobjective optimization of metallurgical properties of microalloyed steels using the desdeo framework," *Engineering Applications of Artificial Intelligence*, vol. 120, p. 105918, 2023.

Details about the provided dataset:

The dataset contain information about the physical properties of many alloys of steel. Each row contains information about a particular alloy composition. Columns A-C and F-Z contain information about concentration of various alloying elements in the steel. Note: Niobium (Nb) and Columbium (Cb) are infact, the same element.

Columns AD through BD contain information about various physical properties. The properties of note are: yield strength (in MPa, Column AD), ultimate tensile strength (in Mpa, Column AE), and percentage elongation before fracture (unitless, column AF), as described during the lecture. Other columns are not relevant for this assignment Treat these properties as the objectives of a three objective problem and do the following:

1. Process the data such that you are able to use it for surrogate modelling. Describe the process. (5 + 5p)
2. From a pool of five surrogate modelling techniques of your choice, find out which technique should be used for optimization. Describe your process. (2 + 5p)
3. Optimize the hyperparameters of the chosen surrogate models for each objective. Describe your process. (3 + 3p)
4. Formally implement the problem, with the surrogate models acting as objectives, in a programming language of your choice. Solve it using an interactive multiobjective optimization algorithm. You may use your implementation of the reference point method from Assignment 1. (4 + 3p)

Note: You can use external libraries for these tasks. If you are using Python, for example, you may use `numpy`, `scipy`, `pandas`, `scikit-learn`, `desdeo`, and `matplotlib`.