1 Ranking Universities (UTA)

Marc works in the human resources department of a large group and wants to rank universities to identify the best in which to recruit future collaborators. He identified a number of indicators that he considers relevant for assessing the relative quality of higher education institutions:

- Median salary of graduates one year after graduation (K€/year),
- Median salary of graduates five year after graduation (K€/year),
- Percentage of international students,
- Percentage of graduates who found a job in less than 3 months,
- Proportion of teaching staff holding a PhD.

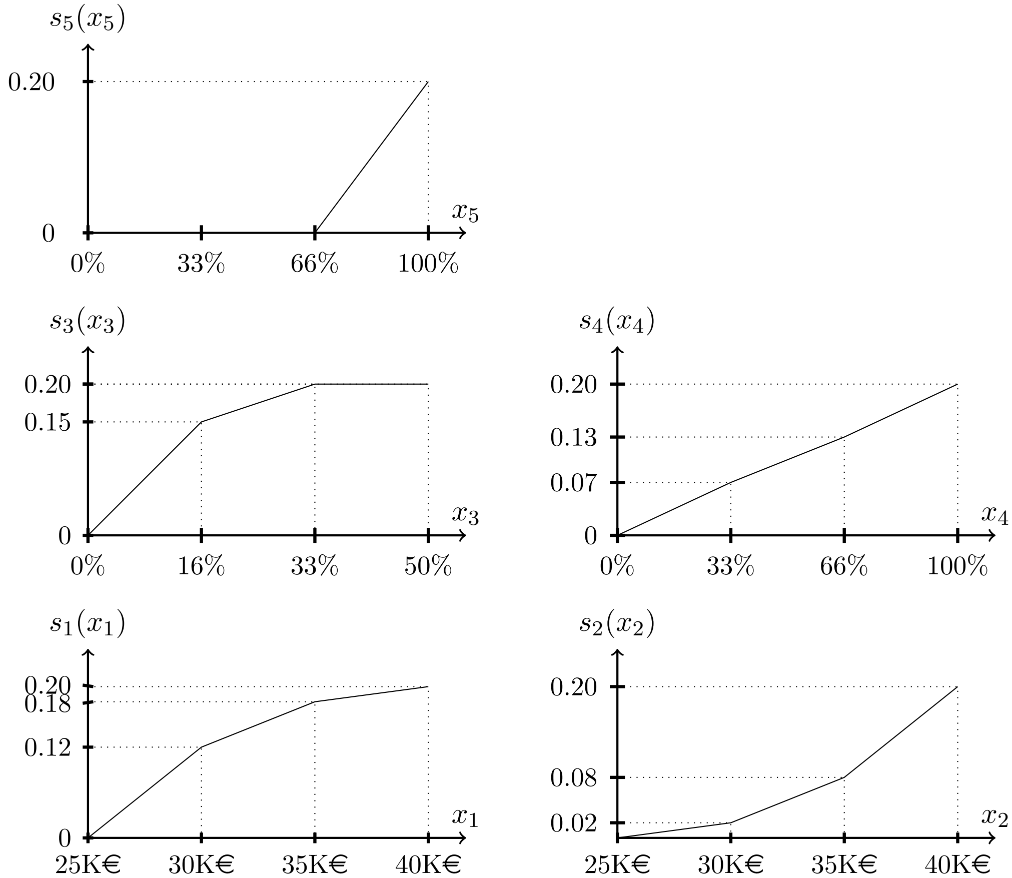
Marc prepared for Sylvie, his department head, a table summarizing these five piece of information for renowned engineering schools/universities in Europe and especially those in which collaborators were recruited in recent years. Part of this information is reported in the table below.

	Salary 1 year (K€/year)	Salary 5 year (K€/year)	% international students	% graduates with job -3 months	% teaching staff with PhD
1	27.5K€	30K€	8%	83%	55%
2	32.5K€	37.5K€	45%	45%	91.5%
3	25K€	32.5K€	16%	90%	25%
4	30K€	35K€	4%	75%	85%
5	25K€	32.5K€	24%	100%	100%
6	39K€	40K€	8%	100%	15%
⋮	⋮	⋮	⋮	⋮	⋮

To develop a ranking, Sylvie suggested to use, for this multicriteria problem, an additive aggregation. It consists in, so she told, formulating a score that specifies the quality of each university; this score then defines a ranking. This score takes the following additive form: $s(j) = \sum_{i=1}^{5} s_i(x_{i,j})$, where $x_{i,j}$ represents the evaluation of University j on criterion i, and $s_i(\cdot)$ is a monotone non-decreasing function such that $s_i(min_i) = 0$ and $s_i(max_i) = w_i$, where w_i is the "weight" of criterion i (with $\sum_i w_i = 1$), and $[min_i, max_i]$ is the evaluation scale on criterion i. This score is defined so that: $j \succ j' \Leftrightarrow s(j) > s(j')$, and $j \sim j' \Leftrightarrow (s(j) = s(j')$ (where $\succ$ and $\sim$ represent the preference and indifference relations among universities).

At first, Marc seeks to directly determine the functions $s_i(\cdot)$ that (he guesses) correspond to the group point of view as to the quality of universities. To make it simple, he considered that $s_i(\cdot)$ functions are piece-wise linear with three linear pieces. The functions designed by Marc are represented below.

1. Can you interpret these functions and clarify the point of view of the group implicitly embedded in these functions?



2. Compute the scores of Universities 1 and 2. Which of the two has the best rank ?

Marc feels that setting directly the parameters of the model is uneasy, and he is not convinced that marginal functions he designed accurately define the group's point of view regarding the quality of universities. Rather than directly specifying the parameters of the additive model, Marc plans to proceed "indirectly": he can give a ranking of a set of Universities he knows, and then determine the score function $s(\cdot)$ which best restores his ranking, allowing then to rank order all Universities.

Before going further, Marc wants to better formalize his approach, as he feels that this work on universities ranking could be of some use for other decision problems... He defines a problem involving n criteria ($n \geq 2$), and considers that the number of linear pieces of $s_i(\cdot)$ functions can vary on each criterion. The scale $[min_i, max_i]$ in criterion i breaks down into the L_i equal subintervals $[x_i^0, x_i^1]$, $[x_i^1, x_i^2]$, ..., $[x_i^{L_i-1}, x_i^{L_i}]$, with $x_i^0 = min_i$ and $x_i^{L_i} = max_i$). Thus, each function $s_i(\cdot)$ is fully defined by the value of the function $s_i(\cdot)$ on "breakpoints", $s_i(x_i^0)$, $s_i(x_i^1)$, ..., $s_i(x_i^{L_i})$.

3. Compute x_i^k, $k \in \{0, ..., L_i\}$, as a function of min_i, max_i, and L_i ? Deduce how to determine $s_i(x_{i,j})$ (the score of University j on criterion i) as a function of $x_{i,j}$, min_i, max_i, L_i and the values $s_i(x_i^k)$, $k \in \{0, ..., L_i\}$.

Marc is familiar with the top 6 universities listed above; he thinks that Univ. 1 precedes Univ. 2, which precedes Univ. 3, ... ($s(1) > \ldots > s(6)$). Marc defines for each university j ($j \in [1, 6]$), σ_j two positive error variable related to the score of University j (σ_j^+, σ_j^-, over, and under-estimation errors), posing: $s'(j) = \sum_{i=1}^{5} s_i(x_{i,j}) - \sigma_j^+ + \sigma_j^-$.

4. Write the linear program that minimizes the sum of errors associated with the 10 ranked universities. The program should include constraints that specify *(i)* Marc's ranking, *(ii)* the monotony of s_i functions, and *(iii)* normalisation constraints $s_i(.) \in [0,1]$.

5. How would you interpret the fact that the optimal value of the objective function is null? What use can you make of the optimal solution of this program in term of the University ranking problem

6. Suppose that Marc states that the intensity by which he prefers Univ. 1 over Univ. 2 is higher than the intensity by which he prefers Univ. 3 over Univ. 4. How can you integrate such statement in the inference program?

7. Marc wants to implement this methodology. More specifically, he want to write a Python program which takes as input:

   – a set of vectors (list of Universities to compare),

   – a rank order on these vectors (a ranking of Universities)

   – a number of linear pieces (for the piecewise value functions)

   which computes a piecewise linear value function that best restores this ranking (assume that $L_i = L$ for every criterion i and use L as a parameter). Use your program to compute the value function inferred from the ranking of the 6 universities (with 3 linear pieces). What is the value of university 7 whose evaluation are (32K€, 35K€, 25%, 85%, 12%); how does it compares with the other universities ?

   The table bellow corresponds to the dataset of the numerical example provided in section 3 in the paper cited in reference. It involves a set of 10 (vintage) cars which are ranked by order of preference. use your program to compute the corresponding additive value function. Check the results with 2, 3 and 4 linear pieces. The Fiat Panda evaluations are (134, 7.4, 6.9, 3, 5.52, 25200), how does it compare to the other cars using the obtained piecewise value functions?

	rank	max speed km/h	cons. city $l/100km$	conso. road $l/100km$	horse power CV	space m^2	price $french\ francs$
Peugoet 505 GR	1	173	11.4	10.01	10	7.88	49500
Opel Record 2000	2	176	12.3	10.48	11	7.96	46700
Citroen Visa Super E	3	142	8.2	7.30	5	5.65	32100
VW Golf GLS	4	148	10.5	9.61	7	6.15	39150
Citroen CX Pallas	5	178	14.5	11.05	13	8.06	64700
Mercedes 230	6	180	13.6	10.40	13	8.47	75700
BMW 520	7	182	12.7	12.26	11	7.81	68593
Volvo 244 DL	8	145	14.3	12.95	11	8.38	55000
Peugeot 104 ZS	9	161	8.6	8.42	7	5.11	35200
Citroen Dyane	10	117	7.2	6.75	3	5.81	24800

Référence

[1] Jacquet-Lagreze, E. & Siskos, J., 1982. "Assessing a set of additive utility functions for multicriteria decision-making, the UTA method", *European Journal of Operational Research*, Elsevier, vol. 10(2), pages 151-164