1. (20 points) We consider the following clustering problem: given $k \in \mathbb{N}$ and a set $Y = \{y_1, \ldots, y_n\} \subseteq \mathbb{R}^d$ of $n$ points, find a partition $C_1, \ldots, C_k$ of $Y$ and representatives $c_i \in C_i$ that minimize
\[ \sum_{i=1}^{k} \sum_{x \in C_i} \|x - c_i\|^2. \]

The difference to $k$-means clustering is that the representatives have to be data points. This is in contrast to $k$-means clustering, where the cluster centers can be arbitrary points in $\mathbb{R}^d$.

We consider the following local search algorithm, similar to the $k$-means method: given $k$ initial representatives, assign data points to their nearest representative. Then, within each cluster obtained in this way, select the point as new representative that minimizes the objective function. Proceed until convergence.

Our goal is to analyze this algorithm in the framework of smoothed analysis. To do this, we assume that $Y$ is obtained from an arbitrary set $X \subseteq [0, 1]^d$ of $n$ points as described in the lecture. Prove that the probability that this local search method needs more than $t$ iterations is bounded from above by $p(n)q(\sigma)\cdot\sqrt{t}$ for two polynomials $p$ and $q$.

Remark, hint: No distinction in sparse and dense iterations is needed here. The analysis of the decrease of the potential caused by reassigning data points suffices, as the positions of the hyperplanes do not depend on the locations of the points that are reassigned. It is also possible to prove a polynomial expected running-time using the chi distribution instead of Part (1a) for the distance between representatives.

a) Prove that the probability that there are two points in $Y$ within a distance of at most $\eta$ is bounded from above by $n^2 \cdot (\eta/\sigma)^d$.

b) Prove that the probability that there exist three distinct points $a, b, x \in Y$ such that $x$ is within a distance of at most $\eta$ of the hyperplane bisecting $a$ and $b$ is bounded from above by $n^3 \cdot \eta/\sigma$.

c) Combine Parts (1a) and (1b) to prove that the probability that there is an iteration in which the potential decreases by at most $\varepsilon$ is bounded from above by $O(n^3 \sqrt{\varepsilon}/\sigma)$.

d) Prove that the objective value is initially bounded by some polynomial in $n$ with a probability of at least $1 - n^{-3kd}$. Use this to finish the proof.
2. (10 points) We consider again the knapsack problem, but this time both weights and profits are random. Our goal is to show that correlations between weights and profits seem to make the problem harder.

Specify probability distributions for the weights \( w_1, \ldots, w_n \) and profits \( p_1, \ldots, p_n \) of items 1, \ldots, \( n \), respectively, with the following properties:

- the random vectors \((p_1, w_1), \ldots, (p_n, w_n)\) are all independent,
- there can be arbitrary correlations between \( w_i \) and \( p_i \),
- the marginal distribution of \( w_i \) is uniform in \([0, 1]\) and the marginal distribution of \( p_i \) is uniform in \([0, 1]\), and
- the expected size of the Pareto curve of solutions is at least \( c^n \) for some constant \( c > 1 \).

3. (20 points) For a graph \( G = (V, E) \), let \( \Delta \) be the maximal degree of \( G \), let \( n = |V| \), and let \( m = |E| \). Let \( \phi \geq 1 \), and let \( f_e : [0, 1] \to [0, \phi] \) be a density function for \( e \in E \). Let \( w_e \) be drawn according to \( f_e \). We consider the flip heuristic for Max-Cut on the instance \( (G, w) \). Let \( \delta_{\text{min}} \) be the smallest possible improvement caused by any possible iteration of the flip heuristic. Let \( T \) be the maximum number of iterations that the flip heuristic needs on the instance \( (G, w) \).

a) Prove that \( \mathbb{P}(\delta_{\text{min}} \leq \varepsilon) \leq \frac{2^\Delta n \phi \varepsilon}{\varepsilon} \).

b) Prove that \( \mathbb{P}(T \geq t) \leq \frac{2^\Delta n m \phi}{t} \).

c) Prove that \( \mathbb{E}(T) = O(2^\Delta n^2 m \phi) \).