**This assignment covers Modules 1–4. It is due on Fri, Feb 14, 2003.**

**Problem 1 (Literature search; 8 marks)** Consider the well-known *$k$-means clustering problem*.

**(a)** Give a concise definition of the problem, as found in the standard literature, and show how it fits the definition of a combinatorial optimisation problem as stated in Chapter 1, Section 1.

**(b)** State the optimisation and decision variants of this problem (see Ch. 1, Sec. 1).

**(c)** Based on a basic literature search, what can you say about the computational complexity of this problem? (Give at least one reference to back up your statements. I expect no more than 5–6 sentences overall, but try to be as precise and informative as possible. Don't give references to papers that you haven't seen – there are many different variants of clustering problems!).

**(d)** List at least two stochastic local search algorithms that have been applied to the $k$-means clustering problem (with references) and briefly describe the neighbourhood relation used in these approaches.

*Note:* The $k$-means clustering problem is also known under a few other names which you will likely come across during your literature search.

**Problem 2 (Problem solving; 5 marks)** Consider the following argument. For Euclidean TSPs, polynomial algorithms exist for arbitrary approximation ratios. Hence, the associated decision problems can be solved in polynomial time for arbitrary solution quality bounds, which implies that the search variant is also efficiently solvable. Why is this argument flawed? (*Hint:* Think carefully about the nature of the solution quality bounds.)

**Problem 3 (Problem solving; 3 marks)** Show formally that Iterative Improvement and Randomised Iterative Improvement can be seen special cases of Probabilistic Iterative Improvement.

**Problem 4 (Problem solving; 5 marks)** Which role do 2-exchange steps play in the Lin-Kernighan procedure for the TSP? (Hint: Is there a way of obtaining complex LK steps based on 2-exchange steps?)