VRIJE UNIVERSITEIT AMSTERDAM
EXAM EVOLUTIONARY COMPUTING
04.06.2002

Note 1 Your name must be written on each sheet in CAPITALS.

Note 2 You can answer the questions in English or in Dutch.

Points to be collected: 72.

Your grade: points collected divided by 72 and rounded up.

1. We have a data mining task (classification problem) to solve with evolutionary computing. That is, there is a data set containing 5000 records each of them having 31 variables. The first 30 variables are all real valued, these are the independent ones. The 31st is binary, this is the dependent variable representing the class of each record to be predicted. We are looking for a predictive model using (some of) the 30 independent variables to tell the class (the value of the 31st variable) of any given data record. Specify an EA suitable[1] for solving this problem. In particular, give
   (a) **(4p)** a representation (the syntax of the chromosomes representing a prediction model),
   (b) **(4p)** a fitness function (how the 5000 records are used for model evaluation),
   (c) **(2p)** an appropriate selection mechanism,
   (d) **(2p)** an appropriate crossover operator,
   (e) **(2p)** an appropriate mutation operator,
   (f) **(2p)** a stop condition.
2. (a) **(2p)** Give the definition of uniform crossover in GAs.
   (b) **(2p)** Give the definition of global discrete recombination in ESs.
   (c) **(2p)** Explain the difference between these operators.
3. **(4p)** Under what conditions can it be guaranteed that an EA will find a solution with probability 1? Present the corresponding formal theorem (including prerequisites).
4. (a) **(3p)** Give the pseudocode of the (1+1) Evolution Strategy (ES) using also σ's.
   (b) **(3p)** Give the pseudocode of Simulated Annealing (SA).
   (c) **(3p)** Mention three important differences between the (1+1) ES and SA.
   (d) **(3p)** Can these algorithms be seen as evolutionary algorithms? Give pro and/or contra arguments.

[1] The EA does not have to be "smart" (efficient). But the representation and the operators should be such that a solution can be found.

5. **(6p)** Suppose you have a genetic algorithm for solving time-tabling problems. (That is, some representation of time-tables, crossover and mutation on time-tables, selection, etc.) Additionally, you have a heuristic operator that can resolve one conflict in any given time-table (providing there is one). In particular, the operator takes a time-table, looks for a conflict, and returns a new time-table without this conflict. In an effort to possibly improve GA performance you need to hybridize the GA with this heuristic. Explain how can you use this operator in the GA, that is how can you extend/modify the GA so that it will (sometimes? always?) use this operator? Describe the resulting algorithm in pseudo code.

6. (a) **(2p)** Explain what order-based representation is.
   (b) **(2p)** Sketch a problem that can be well treated by order-based representation.
   (c) **(2p)** Give the description of an order-based crossover operator.
   (d) **(2p)** Give the description of an order-based mutation operator.

7. Sketch the differences and similarities between Genetic Algorithms, Evolution Strategies and Evolutionary Programming with respect to
   (a) **(3p)** representation,
   (b) **(3p)** mutation,
   (c) **(3p)** crossover,
   (d) **(3p)** self-adaptation.

   Make a table.

8. (a) **(2p)** What is a constraint satisfaction problem (CSP)?
   (b) **(2p)** Why is it not straightforward to solve CSPs with EAs?
   (c) **(4p)** Outline one possible approach to solve a CSP with an EA. Explain the way constraints are handled in detail.