Computer Science and Creative Technologies

Coursework or Assessment Specification

Module Details

Module Code	UFCFY3-15-3
Module Title	BioComputation
Module Leader	L Bull
Module Tutors	-
Year	2019/20
Component/Element Number	B1
Total number of assessments for this module	1
Weighting	50%
Element Description	Practical Assignment requiring the production of program code

Dates

Date issued to students	21/10/19
Date to be returned to students	20/12/19
Submission Date	28/11/19
Submission Place	Blackboard
Submission Time	14:00
Submission Notes	Report and source code

Feedback

Feedback provision will be	Blackboard (written), lab sessions (verbal)

Contents

Section 1: Overview of Assessment

This assignment assesses the following module learning outcomes:

1. To apply an appropriate technique(s) to a given problem
2. Formulate a problem such that it is amenable to modern Artificial Intelligence techniques
3. Appraise the usefulness of various techniques for particular situations

The assignment is worth ***50%*** of the overall mark for the module.

Broadly speaking, the assignment requires you to write a report on your attempts to solve a set of classification problems as effectively as possible using any form of evolutionary intelligence covered on the course. This requires you to write your own code, in a language of your choice, building upon your own genetic algorithm code written and developed in the first few lab sessions.

The assignment is described in more detail in section 2.

This is an individual assignment.

Working on this assignment will help you to develop your understanding of how learning can be seen as a search process and how the parameters controlling search techniques affect their ability to solve tasks. If you have questions about this assignment, please email the module leader.

Section 2: Task Specification

There are four text files on Blackboard each of which contains data of the format: input variables - predicted variable. For example, "data1.txt" contains 60 binary data points with 6 input variables, one exemplar per line:

000000 0
000001 1
000010 0
….

Data2.txt contains binary data records and the other two files contain real-valued data.

The task is to create a system that correctly classifies a given set of input variables, that is, give the predicted variable as the output, using simulated evolution. Fitness is therefore how many of the data a given system correctly classifies in a given

file/format. *To pass the assignment, you must implement a system that successfully evolves a classifier and demonstrate the effects of parameter changes, preferably through graphs and including your understanding of what is happening*. Representations you might like to consider include look-up tables, rule-based systems or neural networks. The search algorithm should be an extension of your genetic algorithm, which might include immune system inspired mechanisms. How you split the data for training and testing is important – no one is interested in a system than can give the correct output for an input it has seen during learning! There is a separate handout "getting started" on Blackboard with further guidance.

Include a research section which briefly reviews data mining in general and then, using two or three examples from the media, considers some of the ethical consequences which can arise from classification mining tasks, eg, face recognition. In an experimentation section describe the encoding used for the induction, show example runs and solutions found. *More marks will be given to the effective use of more sophisticated approaches, particularly for the real-valued data.* Alternatively, once you have solved the problems with your own code, you may use freely available software, eg Weka, to compare performance with other approaches on benchmark data sets.

Section 3: Deliverables

Depending on font size, and line spacing, around 5 pages is a reasonable target length. The intention is your hand-in approximates to a research paper – please use the template provided. Please include commented source code as a printed Appendix. A demonstration will be required during the lab sessions before the end of term.

Section 4: Marking Criteria

	0-40%	40-60%	60-100%
General approach - quality of writing and visual impression (10%)	Use of template and basic word processing skills.	Coherent structure in presentation, including some graphs.	Well structured, results presented in multiple/suitable ways.
Research – relevance and level of understanding shown (20%)	Brief identification and discussion of aspects of data mining, with mention of some ethical issues.	Description of data mining field in general, with summaries of ethical consequences of some applications.	Review of data mining and classification, summaries of related advanced work such as rforrest. Clear identification and discussion of poignant ethical issues.
Experimental Method – no. of experiments, systematic parameter changes, etc.(30%)	Presentation of attempts on the first data set.	Presentation of attempts on the binary data sets.	Presentation of attempts on all data sets.

Analysis and Discussion – presentation and discussion of learning behaviour (30%)	Brief discussion of apparent effects of varying one or more parameters.	Discussion of apparent effects of varying parameters clearly supported by results.	Demonstration of clear insight of effects from parameter sweeps.
Conclusions (5%)	Concise summary.	Summary showing wider understanding.	Demonstration of clear understanding and implications of results.
Citation and Reference Scheme (5%)	A small number of relevant refs.	Some data mining and rule mining refs.	Sets of refs for data mining, class, rules, etc.

Section 5: Feedback mechanisms

The lab sessions will be used for the assignment from the hand-out date enabling formative feedback from then onwards. Feedback will be given in Blackboard during the marking process.