COSC 5P71 Genetic Programming - Term Test

B. Ross

Name: _____ Student number: _____

There are 6 questions on 8 pages totaling 93 marks. You may use a calculator. No other aids are permitted. Please use the backs of pages if you need more space for your answers.

1. [24] Multiple choice [+2 correct; 0 empty; -0.5 wrong guess; one free wrong guess permitted]
Select the single best answer for each question.

1. Sufficiency refers to the idea that:

a. all functions can execute error-free on all argument values
b. the GP language is adequate for defining a solution to a given problem
c. fitness does not result in over-training
d. bloat is minimal

2. Multi-objective optimization can be considered when:

a. there are several competing problem criteria
b. it is not obvious how to combine different criteria together
c. it may not be desirable to combine different criteria together
d. all of the above

3. Given individuals A and B with multiple objectives, A dominates B when:

a. A is better than B in some objectives, and worse than B in other objectives
b. A is better than B in all objectives
c. A is better than B in a least one objective, and equivalent to B in the remaining objectives
d. A is apples, and B is bananas

4. If you increase the size K of a tournament, you are:

a. promoting ephemeral constants
b. reducing premature convergence
c. adding elitism
d. increasing the selective pressure

5. In order to create electrical circuits, Koza used:

a. ADFs
b. Cellular encoding
c. Architecture-altering operations
d. parallelism

6. Island-model parallelism can improve evolutionary performance primarily due to:

a. promoting genetic diversity
b. GPU's
c. multi-objective optimization
d. reducing program bloat

7. A deme is:

a. a gene
b. a hill in the search space
c. an island in parallel evolution
d. a bloated sub-expression that can be removed

8. The No Free Lunch theorem states that:

a. NP-complete problems cannot be solved efficiently by evolutionary search
b. all search algorithms are equivalent in performance, if one considers the universe of possible problems to solve.
c. the larger the GP language used, the better quality of programs are evolved
d. evolutionary computation is the most efficient search algorithm for most problems of interest

9. The main technical "trick" that the No Free Lunch theorem uses is:

a. the adversary can create random (fitness) functions that have no exploitable structure
b. hiding information about an NP-complete problem should make it harder to solve, not easier
c. premature convergence always happens with any heuristic search algorithm
d. GP trees are kept within maximum depth limits

10. Which is true of competitive co-evolution:

a. powerful, dynamic fitness strategies can be evolved
b. it is often naturally adapted to competitive game applications
c. there is a Free Lunch with it
d. all of the above.

11. The course assignments always required at least 10 runs per experiment because:

a. we can statistically evaluate GP runs in general, rather than accept accidental solutions
b. there is no free lunch, and so you have to do at least 10 runs to get the most optimal answer
c. this undoes the complexity of NP-complete problems
d. none of the above

12. Which of the following does not belong with the rest?

a. depth-first search
b. hill-climbing
c. genetic algorithm
d. particle-swarm optimization

2. [15] Define and briefly discuss the following terms.

(a) Protected division

(b) Many-objective optimization

(c) Tournament selection

(d) ADF

(e) Strongly-typed GP

3. [12] Discuss the steps required to perform a crossover operation. Use illustrations to help your discussion. Include the selection step for the parent(s). Also include references to test(s) to ensure the resulting trees do not violate any restrictions (what restriction is usually enforced?).

4. [12] Define the terms “training” and “testing”. Discuss the similarities and differences between them, and the importance of each. Also discuss overtraining, and how it can be detected during a run.

5. [10+2+2+4 = 18] Consider the following population of 4 individuals having the raw fitness in column (a). High values represent stronger fitness.

	(a) Raw Fitness (high is better)	**(b) Fitness squared**	**(c) Roulette for (a)**	**(d) Roulette for (b)**	**(e) Rank for (a) (low is better**	**(f) Inverse rank (high is better)**	**(g) Roulette for (f)**
1	1						
2	2						
3	4						
4	8						
∑							

a) [10] Compute the rest of the table as follows: (b) represents the squared raw fitness; (c) and (d) are the Roulette wheel areas for (a) and (b) respectively; (e) is the raw fitness converted to ranks ("1" is the most fit individual); (f) inverts the ranks ("4" is the most fit individual); and (g) is the inverted rank converted to Roulette wheel area. (The bottom row is to help your own calculations).

b) [2] Which Roulette wheel area column (c, d, or g) has the strongest selective pressure, and why?

c) [2] Which column, if any, is the most effective column to use with a tournament selection? Explain your answer.

d) [4] Given the above, identify a strength and weakness of both of the Roulette Wheel and Tournament selection strategies.

6. [6+6=12] (a) Define the term “program bloat”. What are its charactistics? Describe 2 ways in which it is detrimental to genetic programming. Also give an example of a program segment or sub-tree, which enables the growth of bloat code in a GP tree.

(b) Identify 2 strategies in which the effects of bloat might be reduced or controlled within a GP run.

The End