CS6271 Evolutionary Algorithms and Humanoid Robotics
Mid-Term Exam 2022
Time: 50 minutes.

Instructions
- Answer **all** questions;
- Question 1 is worth **40 marks**, and each of Questions 2, 3 and 4 are worth **20 marks**;
- Hand up any roughwork;
- Calculators **can** be used.

Question 1 (Evolutionary Algorithms)

a) Explain briefly what the difference is between **tournament selection** and **roulette wheel selection**. (*2 marks*)

b) Why is it useful to have a *Hall of Fame* when running an Evolutionary Algorithm? (*3 marks*)

c) Why is it recommended to conduct multiple runs in an EA experiment? Suggest a situation where someone using an EA would be satisfied with a small (possibly even a single) number of runs. (*7 marks*)

d) Explain briefly what is meant by **overfitting** and give a course of action to avoid it. (*7 marks*)

e) What is **one-hot encoding** and why is it necessary in Machine Learning? Give a simple example. (*5 marks*)

f) Given the following population (size 5) and associated fitness values where we want to select a new population one individual at a time using **Tournament Selection**, what would happen if the tournament size was 5? What if it was set to 1? You can keep your answer high level. Assume that this is a **maximisation problem**; that is, higher fitness is better. (*8 marks*)

Individual	Fitness
1	23
2	7
3	10
4	15
5	20

g) Using the same individuals and fitness values, what is the percentage chance each individual will be selected using **Roulette Wheel Selection**? (*8 marks*)

Question 2 (Genetic Algorithms)

a) In terms of schemata, what is the difference between **order** and **defining length**? Use examples in your response. (*5 marks*)

b) Explain, with an example, the difference between **two-point** and **uniform** crossover. (*5 marks*)

c) Given the following individuals, which ones contain the schemata 1*****0**1****? It is enough to note the ID number corresponding to the individuals that cannot be created. (*5 marks*)

ID	Individual
1	00110011001100
2	01010101011101
3	11001100110101
4	10110011001101
5	01010101100110
6	10010101100110
7	01100101100111
8	01001010010101

d) Which of the following offspring can **not** be created by one point crossover from two parents 000000 and 111111? It is enough to note the ID number corresponding to the individuals that cannot be created. (*5 marks*)

ID	Individual
1	111111
2	000000
3	111000
4	110011
5	011110
6	001111
7	011111

Question 3 (Genetic Programming)

a) Explain in your own words how the mean square error operates. (*4 marks*)

b) Show an example of crossover in GP using the following two individuals:

(* (+ x y) (+ z y)) and (* (+ (* x x) y) (+ z y)) (*4 marks*)

c) If you were asked to apply GP to a symbolic regression problem with three input variables, what would your function and terminal sets be? (*6 marks*)

d) If you were asked to apply GP to a classification problem with three input variables, all of which are numeric and which needs to return either **True** or **False**, what would your function and terminal sets be? (*6 marks*)

Question 4 (Grammatical Evolution)

a) Under what circumstances is it generally preferable to use GE rather than GP? (*3 marks*)

b) Explain the difference between *actual length* and *effective length*. (*5 marks*)

c) Given the following grammar and codon values, what would the individual map to? You can assume that there is **no** wrapping, so if you get to the end of the genome, you can stop and assume that is your answer. (*6 marks*)

```
<e>::= (<o><e><e>) | (<u><e>) | <v>

<o>::=+|-|/|*

<u>::=Sin|Cos|Tan

<v>::=x|y
```

7, 176, 22, 72, 77, 74, 177, 215, 200, 183, 229, 111, 77, 124, 66, 94, 91, 232, 44, 53, 114, 71, 234, 98, 72, 106, 71, 231, 97, 203, 32, 15

d) Write a grammar for Grammatical Evolution that could be used to produce a classifier (that returns either **True** or **False**) for a data set with the following inputs and types. (*6 marks*)

Input Name	Type
Age	Integer
Gender	Binary (0=Female; 1=Male)
Experience	Integer
PreviousAccidents	Integer