Computational Intelligence exam (SOLUTIONS)

All questions were worth 1 point

January 23, 2015

1. **Fuzzy Sets and the Extension Principle**. Let A and B be fuzzy sets defined on $\mathbb{Z}$, with membership functions given by:

$$A(x) = \frac{0.5}{-1} + \frac{0.8}{0} + \frac{1}{1} + \frac{0.2}{2}$$

$$B(x) = \frac{0.5}{1} + \frac{1}{2} + \frac{0.5}{3} + \frac{0.8}{4}$$

Consider a subset $X \subseteq \mathbb{Z}$ and a function $f : X \times X \to X$ defined for all $x_1, x_2 \in X$ as $f(x_1, x_2) = x_1 \cdot x_2$. Calculate $f(A, B)$.

Solution:

$$f(A,B) = \frac{0.5}{-4} + \frac{0.5}{-3} + \frac{0.5}{-2} + \frac{0.5}{-1} + \frac{0.8}{0} + \frac{0.5}{1} + \frac{1}{2} + \frac{0.5}{3} + \frac{0.8}{4} + \frac{0.2}{6} + \frac{0.2}{8}$$

2. **Fuzzy Arithmetic**. Let A and B be two fuzzy numbers whose membership functions are:

$$A(x) = \begin{cases} \frac{x+2}{2} & \text{for } -2 < x \leq 0 \\ \frac{2-x}{2} & \text{for } 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

$$B(x) = \begin{cases} \frac{x-2}{2} & \text{for } 2 < x \leq 4 \\ \frac{6-x}{2} & \text{for } 4 < x \leq 6 \\ 0 & \text{otherwise} \end{cases}$$

Calculate the fuzzy numbers:

Solution:

(a)

$$(A+B)(x) = \begin{cases} \frac{x}{4} & \text{for } 0 < x \leq 4 \\ \frac{8-x}{4} & \text{for } 4 < x \leq 8 \\ 0 & \text{otherwise} \end{cases}$$

(b)

$$(A \cdot B)(x) = \begin{cases} \frac{4-(4-x)^{1/2}}{2} & \text{for } -12 \leq x < 0 \\ \frac{4-(4+x)^{1/2}}{2} & \text{for } 0 \leq x \leq 12 \\ 0 & \text{otherwise} \end{cases}$$

(c)

$$MIN(A(x), B(x)) = A(x)$$

(d)

$$max(A(x), B(x)) = \begin{cases} \frac{x+2}{2} & \text{for } -2 < x \leq 0 \\ \frac{2-x}{2} & \text{for } 0 < x \leq 2 \\ \frac{x-2}{2} & \text{for } 2 < x \leq 4 \\ \frac{6-x}{2} & \text{for } 4 < x \leq 6 \\ 0 & \text{otherwise} \end{cases}$$

3. **Fuzzy Logic systems**. Explain why the Takagi-Sugeno-Kang (TSK) fuzzy logic system is a more compact and computationally efficient representation than a Mamdani fuzzy logic system. In which cases do you think it is more useful to use one or the other? Give explanations and examples to support your opinion.

   Solution: TSK is more compact and computationally more efficient because the representations of rule consequents are constant or linear functions, and therefore the computationally expensive defuzzification process of the Mamdani approach is not required. Both methods work well, Mamdani being very intuitive and TSK being more amenable to mathematical analysis. Mamdani is well suited for control and decision support systems whereas TSK is also a good option for control and lends itself to the use of adaptive techniques for constructing fuzzy models for classification or regression.

4. **Computational Intelligence concepts**. Give a definition of Computational Intelligence (CI) *as you understand it, not copying it from a slide*, and list the three or four fundamental CI paradigms (not specific algorithms) that *best fit into this definition.*

   Sketch an example of a problem requiring CI, where the cost of finding an exact solution to a precisely formulated version of the problem is unfeasible, but we may develop affordable algorithms to find an approximate solution to a (possibly imprecise) formulation for the same problem.

   Solution: CI is a subfield of computer science that draws from the successes of natural systems to develop new ways of solving difficult computational problems.

   The fundamental CI paradigms are Evolutionary Computation, Neural Computation and Fuzzy Computation (Probabilistic and/or Approximate Computation are also useful paradigms).

   The typical scenario would solving a NP problem. Although these problems are largely unfeasible theoretically for non-trivial sizes, very useful algorithms have been developed to find suboptimal solutions using affordable resources (time and space).

   (note: this was a question with a rather free answer)

5. **A problem in inference**. An English-speaking tourist is visiting a country in Europe whose language is not English. At some moment he is out in the street and he needs to ask for some directions. A friend told him that 1 in 10 natives speak English, 1 in 5 people in the street are tourists, and half the tourists speak English. Our visitor then stops someone in the street and finds that this person speaks English. What is the probability that this person is a tourist?

   Solution: Let E denote "English-speaking" and T denote "tourist". We know that $P(E|\overline{T}) = \frac{1}{10}$, $P(T) = \frac{1}{5}$ and $P(E|T) = \frac{1}{2}$. The required probability is

$$P(T|E) = \frac{P(E|T)P(T)}{P(E|T)P(T) + P(E|\overline{T})P(\overline{T})} = \frac{\frac{1}{2} \cdot \frac{1}{5}}{\frac{1}{2} \cdot \frac{1}{5} + \frac{1}{10} \cdot \frac{4}{5}} = \frac{5}{9}$$

   Rather surprisingly, in this country, if you stop someone in the street that speaks English, he/she will most likely be a tourist!

6. **Hopfield networks**. Explain briefly what are the two main fields of application of these networks. Give an example of each application to support your answers. Decide whether the following sentences are true (just answer yes/no):

   (a) The memories are stable states of the network
   (b) The memories are local minima of the energy function of the network
   (c) The energy function always increases (or remains the same)
   (d) The network may not always output one of the memories

(e) The energy function represents the quadratic error between the memories and the actual outputs of the network

Solution: The two main fields of application of a Hopfield network are:

- Content-addressable autoassociative memory, with certain ability to restore or correct memories
- Finding (suboptimal) solutions to combinatorial optimization problems

(a) YES
(b) YES
(c) NO
(d) NO
(e) NO

7. **Early neuron models**. Design a McCullogh & Pitts neuron to compute the logical function:

$$f(x_1, x_2, x_3) = \overline{x}_1 \wedge x_2 \wedge \overline{x}_3, \qquad \text{where } x_i \in \{0, 1\}, i = 1, 2, 3.$$

Solution: The weights for x_1, x_2, x_3 are, respectively, $w_1 = -1, w_2 = 1, w_3 = -1$, with a threshold $\theta = 1$.

8. **A problem for genetic algorithms**. A political map is called k-colorable if each country in the map can be assigned one of k colors such that no two adjacent regions have the same color (two regions are called adjacent if they share a common boundary that is not a corner). In 1976 K. Appel and W. Haken proved the *four color conjecture*: every map is 4-colorable. Describe in detail how to find a solution for a specific map-coloring problem with 4 colors using a genetic algorithm.

Solution: Suppose the map has r regions; we can describe the map using a $r \times r$ adjacency matrix $R = (r_{ij})$, where $r_{ij} = 1$ if regions i, j are adjacent and $r_{ij} = 0$ otherwise; we also set $r_{ii} = 0$. An individual is just a bitstring of length $2r$, because we need 2 bits to code the 4 possible colors. The fitness function of an individual (to be minimized) counts the number of adjacent regions having the same color:

$$F(I) = \sum_{i=1}^{r} \sum_{i=j+1}^{r} r_{ij} [c_i(I) = c_j(I)]$$

where $[z] = 1$ when z is true and 0 otherwise, and $c_i(I)$ is the decoded color in position i of an individual I. It is worth to note that crossover should not break the bits representing a color; the alternative is to code for color using 4 bits, as in `0001,0010,0100,1000`. In this case, both mutation and crossover need to be under control not to generate invalid colors (so this latter code is worse).

9. **Learning and generalization**. Two students are arguing on a conceptual problem: in a certain machine learning task, we have the following six examples (six variables each) available for learning:

	Inputs	Output
Examples	1 0 1 1 0 1	1
	0 0 0 0 0 0	1
	0 0 1 1 0 0	1
	0 1 0 1 1 0	0
	1 1 1 0 1 1	0
	0 0 0 1 1 1	0

Suppose we train a machine learning method on these examples and obtain a prediction model. After training, we present the new example `0 1 1 1 0 1` to the model. Choose one of the answers below and justify the choice:

a. The model will output a prediction of `1` because ...
b. The model will output a prediction of `0` because ...
c. We cannot possibly know, because ...

Solution: We cannot possibly know, because there are many different models (logical functions, in this case) that our learner may have learnt (a nasty consequence of the small number of examples). For example, *even parity* (the number of 1's is even), the *negation of input number 5*, or *symmetry* are all consistent with this training set. However, the prediction for `0 1 1 1 0 1` would be 1, 1 and 0, respectively.

10. **Evolution Strategies and Genetic Algorithms**.

    (a) Explain the differences between an Evolution Strategy (ES) and a Genetic Algorithm (GA). Discuss the role of mutation and recombination, selection mechanisms, data representations and any other fundamental issue that you find relevant.

    (b) Sketch two specific optimization problems, one for which in principle the ES should be superior and another for which the GA should be superior.

Solution: The first question has been largely explained in the lectures. Essentially, ESs explore continuous search spaces, use mutation as the main driving force, selection is deterministic and "comma", there is self-adaptation of strategy parameters and require generation of a surplus offspring. Differently, GAs explore discrete (usually binary) search spaces, use recombination as the main driving force, selection is stochastic and "comma", there is no self-adaptation of strategy parameters and do not require the generation of a surplus offspring.

The second question is more difficult to answer than what it seems at first glance. The obvious answer would be "ESs should be superior when its distinctive characteristics (those listed above) are favourable for the problem to be solved; and same for GAs.". A precise answer is actually a research problem and therefore I will treat your answers with special care.