CMPE 452: Neural and Genetic Computing

Assignment 4

2023-11-24

Mile Stosic

20233349

Table of Contents

1 Selection (3 points) 1
2 Schema Theory (3 Points) 2
3 Mutation (3 Points) 2
4 GA – Programming (4 Points) 3

1 Selection (3 points)

a) Roulette-wheel selection

a. Total Fitness is the sum of F,

$$\sum F = 5 + 7 + 8 + 10 + 15 = 45$$

Probability for Individual 4,

$$P(F_4) = \frac{F_4}{\sum F} = \frac{10}{45} = \frac{2}{9} = \sim 0.2222 = 22\%$$

b) Tournament Selection, with tournament size equal to 2, and the probability of selecting the best individual (in a given tournament) equal to 0.75, There are two scenarios to consider for individual 4 being selected:

1. Individual 4 is paired with an individual with a lower fitness (1, 2, or 3). The probability of winning is 0.75.
2. Individual 4 is paired with individual 5. The probability of losing is 0.25.

Thus,

$$P(Selecting\ F_4) = \frac{1}{25}\left(1 + 2 * P + 6(1 - P)\right)$$

$$P(Selecting\ F_4) = \frac{1}{25}(6) + 0.24 = 24\%$$

c) Roulette-wheel selection, based on linearly ranked fitness values, where the lowest fitness value is set to 1 and the highest fitness value set to 10,

Ranking the individuals linearly based on the description above,

$$F1' = 1$$

$$F2' = 1 + \frac{(10 - 1)}{4} * 1 = 3.25$$

$$F3' = 1 + \frac{(10 - 1)}{4} * 2 = 5.5$$

$$F4' = 1 + \frac{(10 - 1)}{4} * 3 = 7.75$$

$$F5' = 10$$

Total Fitness is the sum of F,

$$\sum F' = 1 + 3.25 + 5.5 + 7.75 + 10 = 27.5$$

Probability for Individual 4,

$$P(F_4{}') = \frac{F_4{}'}{\sum F'} = \frac{7.75.}{27.5} = \frac{7.75}{27.5} = \sim 0.2818$$

2 Schema Theory (3 Points)

a) The order of the given schema,

$$S = 0 ** 1 * 1 ** 0 ***$$

Is equal to the amount of 0's and 1's, thus the order of the given schema is,

$$Order(S) = 4$$

The length of defining H is equal to,

$$Defining\ Length(H) = (Position\ of\ last\ 0) - (Position\ of\ first\ 0)$$

Thus,

$$Defining\ Length(H) = 8$$

b) The average fitness for 1*** and 0*** can be seen below,

For schema 1*** with L = 4, the possible instances of this schema are,

$$1000, 1001, 1010, 1011, 1100, 1101, 1110.$$

The integer values of these instances are 8, 9, 10, 11, 12, 13, 14, and 15.

The average fitness of the schema 1*** is the average of these integer values,

$$= \frac{(8 + 9 + 10 + 11 + 12 + 13 + 14 + 15)}{8} = \frac{92}{8} = 11.5$$

For schema 0*** with L = 4, the possible instances of this schema are,

$$0000, 0001, 0010, 0011, 0100, 0101, 0110, and\ 0111.$$

The integer values of these instances are 0, 1, 2, 3, 4, 5, 6, and 7.

The average fitness of the schema 0*** is the average of these integer values,

$$= \frac{(0 + 1 + 2 + 3 + 4 + 5 + 6 + 7)}{8} = \frac{28}{8} = 3.5$$

3 Mutation (3 Points)

The probability of mutation events can be analyzed by using the binomial distribution, where the number of trials is the number of genes m, the probability of success is p, and the number of successes is the number of mutations we're interested in.

The binomial probability function is given by,

$$P(X = k) = \binom{m}{k} p^k (1 - p)^{m-k}$$

where X is the random variable representing the number of mutations, k is the specific number of mutations we want to calculate the probability for, m is the total number of genes, p is the mutation rate, and $\binom{m}{k}$ is the binomial coefficient.

Thus,

a) This is the probability that k = 0,

$$P(X = 0) = \binom{m}{0} p^0 (1 - p)^{m-0} = (1 - p)^m$$

b) This is the probability that k = 1

$$P(X = 1) = \binom{m}{1} p^1 (1 - p)^{m-1} = m * p * (1 - p)^{m-1}$$

c) This is the probability that the chromosome undergoes either 0, 1, or 2 mutations. We sum the probabilities of these three separate events.

$$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$P(X < 3) = (1 - p)^m + m * p * (1 - p)^{m-1} + \binom{m}{2} p^2 (1 - p)^{m-2}$$

4 GA – Programming (4 Points)

Code was provided in the submission document but the answers to the following questions can be seen below,

The results show that the genetic algorithm performs better with a combination of crossover and mutation than with mutation alone. This is because:

1. Crossover utilizes the genetic information from two parents, which can create offspring with improved fitness by combining advantageous traits.
2. Mutation introduces new genetic variations that can be beneficial but is less effective on its own for finding optimal solutions, as it relies on random modifications.

When crossover is stopped, the algorithm relies solely on mutation, which significantly increases the number of generations required to find the optimal string. This is less efficient because it lacks the structured exploration and exploitation of the search space that crossover provides.

In essence, crossover allows the GA to search through the solution space by recombining existing good solutions, while mutation ensures diversity and prevents premature convergence. The balance between these two mechanisms is crucial for the effective functioning of a genetic algorithm.