Genetic Algorithm Assignment

- Ratnesh Kumar Rai(MIT2019098)

1. Consider three strings A_1=11101111, A_2= 00010100, A_3= 01000011 and six schemata H_1=1*******, H_2=0*******, H_3= ******11, H_4= ***0*01*, H_5= 1*****1*, and H_6=1*****1*.

    a. Which schemata is matched by which string?
    b. What are the order and defining length of each of the schemata?
    c. Estimate the probability of survival of each schema under mutation when the probability of a single mutation is p_m=0.001.
    d. Estimate the probability of survival of each schema under crossover when the probability of crossover p_c=0.85.

Sol: **Given p_m = 0.001 and p_c = 0.85, the calculation table is given below:**

Schema(H)	Order (o(H))	Defining length (δ(H))	Strings matching to schema	$S_m(p_m) = (1 - p_m)^{o(H)}$ (SURVIVAL UNDER MUTATION)	$S_c(H,p_c) \geq 1 - p_c \frac{\delta(H)}{l-1}$ (SURVIVAL UNDER CROSSOVER)
H_1 = 1*******	1	0	A1	0.999	1
H_2 = 0*******	1	0	A2, A3	0.999	1
H_3 = ******11	2	1	A1, A3	0.998	0.8780
H_4 = ***0*01*	3	3	A3	0.997	0.6357
H_5 = 1***1***	2	6	A1	0.998	0.2714
H_6 = 1***1***	2	6	A1	0.998	0.2714

2. A population contains the following strings and fitness values at generation 0:

#	String	Fitness
1.	10001	20
2.	11100	10
3.	00011	5
4.	00011	15

The probability of mutation is p_m=0.01 and the probability of crossover is p_c=0.7. Calculate the expected number of schemata of the form 1**** in generation 1. Estimate the expected number of schemata of the form 0**1* in generation 1.

Sol. Let H_1 = 1**** and H_2 = 0**1*. Let the given strings be denoted by S1, S2,S3,S4 respectively.

String (S_i)	Fitness(f_i)
S1 = 10001	20
S2 = 11100	10
S3 = 00011	5
S4 = 00011	15
Total Fitness($\sum_{i=1}^{N} f_i$)	50
Avg. Fitness ($\bar{f}$)	12.5

Now,

Schema(H)	Order (o(H))	Defining length (δ(H))	Strings matching to schema	m(H,0)	f(H,0) = $\frac{\left(\sum_{i\in H}^{m(H,0)} f_i\right)}{m(H,0)}$
H_1 = 1****	1	0	S1, S2	2	15
H_2 = = 01***	2	3	S3, S4	2	10

The expected number instance for Schema(H) at next generation is given by :

$$E[m(H,k+1)] \geq m(H,k)\frac{f(H,k)}{\bar{f}}\left(1-p_c\frac{\delta(H)}{l-1}\right)(1-p_m)^{o(H)}$$

Given, p_m = 0.01, p_c = 0.7, Now,

$$\boldsymbol{E[m(H_1,1)] \geq m(H_1,0)\frac{f(H_1,0)}{\bar{f}}\left(1-p_c\frac{\delta(H_1)}{l-1}\right)(1-p_m)^{O(H_1)}}$$

$$\boldsymbol{\geq 2 * \frac{15}{12.5}\left(1-0.7*\frac{0}{4}\right)(1-0.01)^1}$$

$$\boldsymbol{\geq 2.376}$$

Similarly,

$$\boldsymbol{E[m(H_2,1)] \geq m(H_2,0)\frac{f(H_2,0)}{\bar{f}}\left(1-p_c\frac{\delta(H_2)}{l-1}\right)(1-p_m)^{O(H_2)}}$$

$$\boldsymbol{\geq 2 * \frac{10}{12.5}\left(1-0.7*\frac{3}{4}\right)(1-0.01)^2}$$

$$\boldsymbol{\geq 0.7448}$$

So, Expected number of schemata for $H_1 \geq 2.376$ and $H_2 \geq 0.7448$ in generation 1.