



The University of Vermont

CS352 Evolutionary Computation: Homework 2

Ayat Ospanov

September 25, 2018

Contents

1	Task 1	1
2	Task 2	2
3	Task 3	2
4	Task 4	2
5	Task 5	3
6	Task 6	3

1 Task 1

The equation for the basic Schema Theorem is the next:

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

where

- $m(H, t)$ is the proportion of individuals representing schema H at time-step t
- $f(H)$ is the fitness of the schema H
- \bar{f} is the mean fitness of the population
- $o(H)$ is the order of the schema H
- $\delta(H)$ is the defining length of the schema H
- L is the length of a genotype
- p_c is the probability of applying crossover
- p_m is the bitwise mutation probability

The first factor $m(H, t) \frac{f(H)}{\bar{f}}$ is the probability of a schema being selected. It is obvious that it depends on the relative fitness of the schema (because the selection is fitness proportionate) and on the proportion of individuals.

The second factor $1 - \frac{\delta(H)}{L-1}p_c - o(H)p_m$ is the probability of surviving variation operators. $\frac{\delta(H)}{L-1}p_c$ is the probability of disrupting examples by crossover and $o(H)p_m$ is the probability of disrupting examples by mutation.

Thus, the theorem states “short, low-order (derives from the second factor) schemata with above-average fitness (the first factor) increase exponentially over generations”

2 Task 2

The fundamental message of the Schema Theorem, or Building Block Hypothesis, is that GAs can begin by selecting short, low-order schemata examples, and then combine them to create higher order schemata, repeating until a schema of length $L - 1$ and order L is created and selected.

3 Task 3

There are 3 types of problem “difficulty”.

- intra-BB
- inter-BB
- extra-BB

The most important is an intra-BB difficulty, or deceptiveness. This means at least one optimal schema is outcompeted by non-optimal schema.

4 Task 4

$$\begin{aligned} S1 &= *0**11***0**, & o(S1) &= 4, \delta(S1) = 8, L = 12 \\ S2 &= *****0*1****, & o(S1) &= 2, \delta(S1) = 2, L = 12 \end{aligned}$$

a) The probability p_s of surviving both crossover and mutation is

$$p_s(H) = (1 - p_m)^{o(H)} \left(1 - p_c \frac{\delta(H)}{L - 1} \right)$$

Thus,

$$\begin{aligned} p_s(S1) &= (1 - p_m)^4 \left(1 - \frac{8p_c}{11} \right) \\ p_s(S2) &= (1 - p_m)^2 \left(1 - \frac{2p_c}{11} \right) \end{aligned}$$

It is obvious, that $p_s(S1) \leq p_s(S2)$. And the probability of surviving of both of them is $p_s(S1) \cdot p_s(S2)$

b) As we know, “building blocks” are short (in terms of defining length) and low-order schematas. $S1$ cannot be called a “building block” as its order and length are high, while $S2$ can be called a “building block”, because it has the length of 2 and the order of 2 which makes possible to describe $S2$ as short and low-order schemata.

5 Task 5

string	fitness
100	10
111	20
011	15
010	15

a)

$$\begin{aligned}
 f(1**) &= \frac{N_{100}f_{100} + N_{111}f(111)}{N_{100} + N_{111}} = \\
 &= \frac{25 * 10 + 25 * 20}{25 + 25} = \frac{10 + 20}{2} = 15
 \end{aligned}$$

b) The estimated number of samples of a schemata H at the next step is:

$$n(H, t+1) = n(H, t) \frac{f(H)}{\bar{f}}$$

Thus, for the schemata 1** we can get the estimated number of samples on the next step:

$$n(1**, t+1) = 50 * \frac{15}{15} = 50$$

But we need the estimated fitness. To count, we need to estimate the number of survivors of guys from the schema 1**. As we know those guys and can estimate numbers for a scheme, we can put $H_1 = 100$ and $H_2 = 111$. Now let's estimate the numbers:

$$\begin{aligned}
 n(100, t+1) &= n(100, t) \frac{f(100)}{\bar{f}} = 25 \cdot \frac{10}{15} = 25 \cdot \frac{2}{3} \\
 n(111, t+1) &= n(111, t) \frac{f(111)}{\bar{f}} = 25 \cdot \frac{20}{15} = 25 \cdot \frac{4}{3}
 \end{aligned}$$

Now we can estimate the fitness value for the 1** scheme.

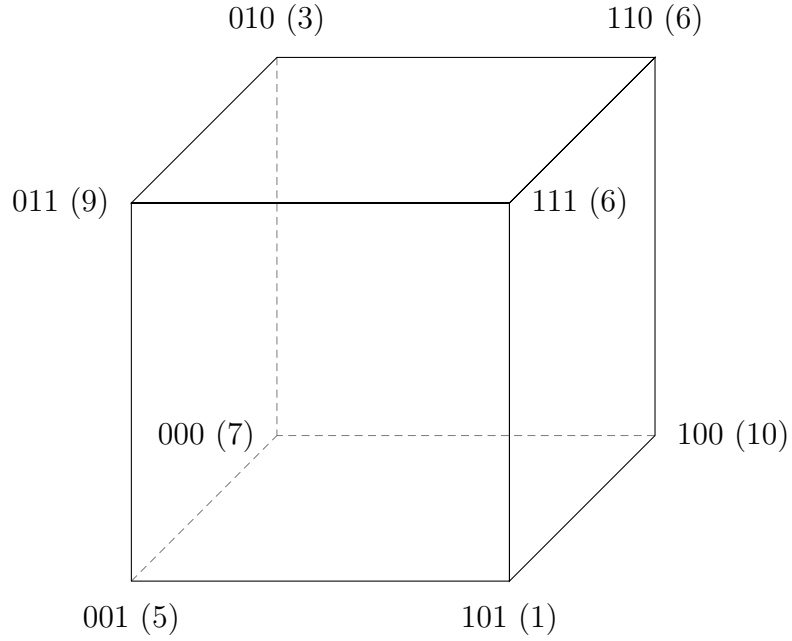
$$\begin{aligned}
 f_{t+1}(1**) &= \frac{n(100, t+1) * f(100) + n(111, t+1) * f(111)}{n(100, t+1) + n(111, t+1)} = \\
 &= \frac{25 \cdot \frac{2}{3} \cdot 10 + 25 \cdot \frac{4}{3} \cdot 20}{50} = \frac{\frac{2}{3} \cdot 10 + \frac{4}{3} \cdot 20}{2} = \\
 &= \frac{10 + 2 \cdot 20}{3} = \frac{100}{6} \approx 16.7
 \end{aligned}$$

6 Task 6

phenotype (integer)		0	1	2	3	4	5	6	7
genotype	binary	000	001	010	011	100	101	110	111
	gray	000	001	011	010	110	111	101	100
fitness		7	5	3	9	10	1	6	6

Table 1: Schema Analysis for binary code and gray code

	order	3	2			1			0
binary	schema	111	11*	1*1	*11	**1	*1*	1**	***
	fitness	6	6	3.5	7.5	5.25	6	5.75	5.875
	schema		01*	0*1	*01	**0	*0*	0**	
	fitness		6	7	3	6.5	5.75	6	
	schema		10*	1*0	*10				
gray	fitness		5.5	8	4.5				
	schema		00*	0*0	*00				
	fitness		6	5	8.5				
	schema	100	10*	1*0	*00	1**	*0*	**0	***
	fitness	6	6	8	6.5	5.75	6	8	5.875
gray	schema		11*	1*1	*11	0**	*1*	**1	
	fitness		5.5	3.5	2	6	5.75	3.75	
	schema		01*	0*1	*01				
	fitness		6	4	5.5				
gray	schema		00*	0*0	*00				
	fitness		6	8	6.5				

**Figure 1:** Hamming cube diagram for binary encoding

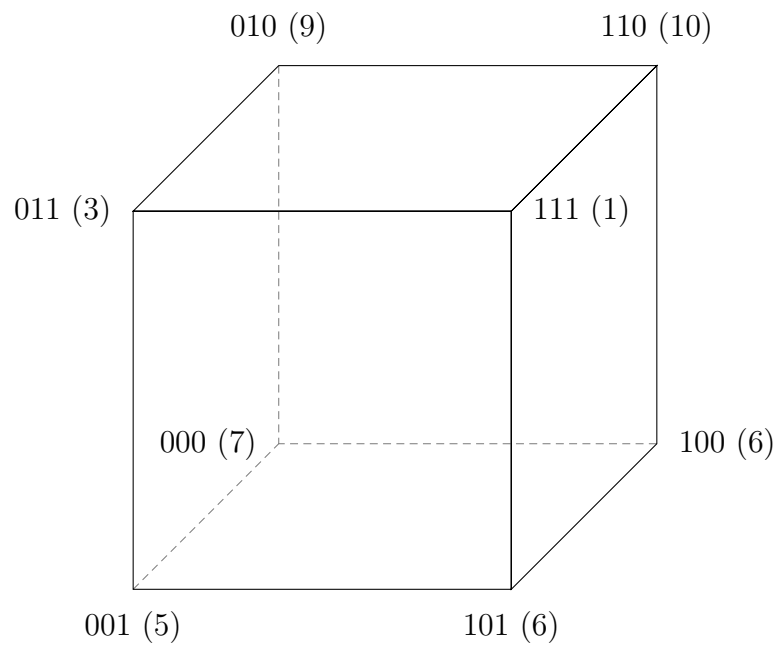


Figure 2: Hamming cube diagram for gray encoding