Introduction to AI

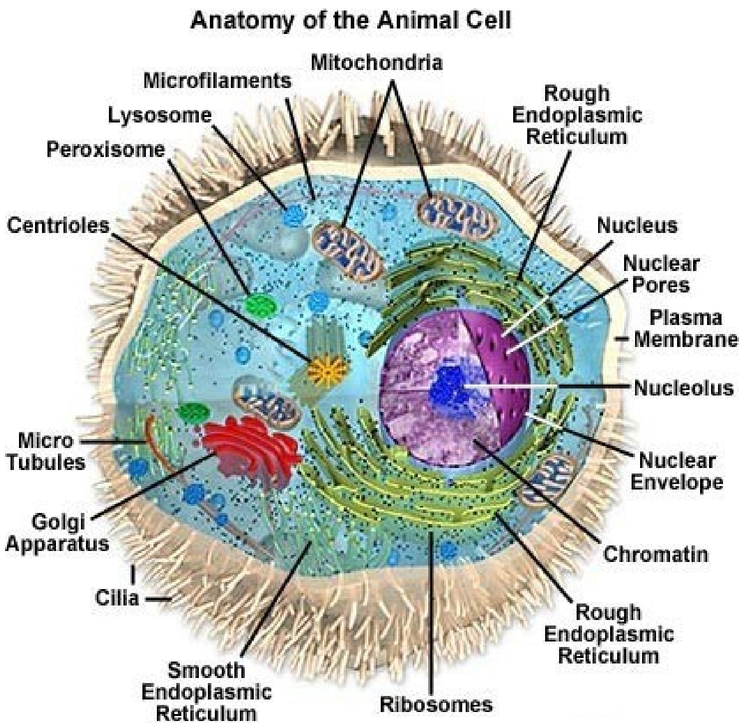
Lecture 7
Genetic Algorithm

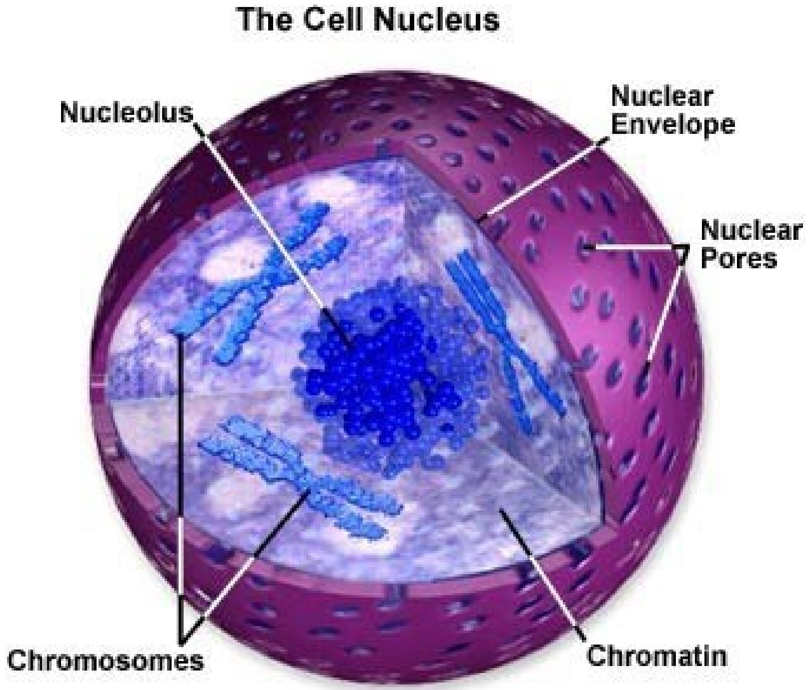
Dr. Tamal Ghosh
Department of CSE
Adamas University

Biological motivation

Biological Background “The cell”

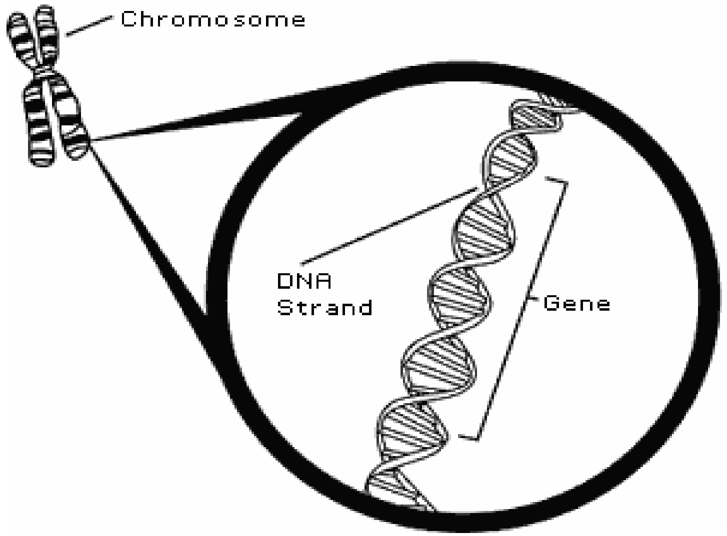
- Every animal cell is a complex system of many small “factories” working together.
- The nucleus in the center of the cell.
- The nucleus contains the genetic information





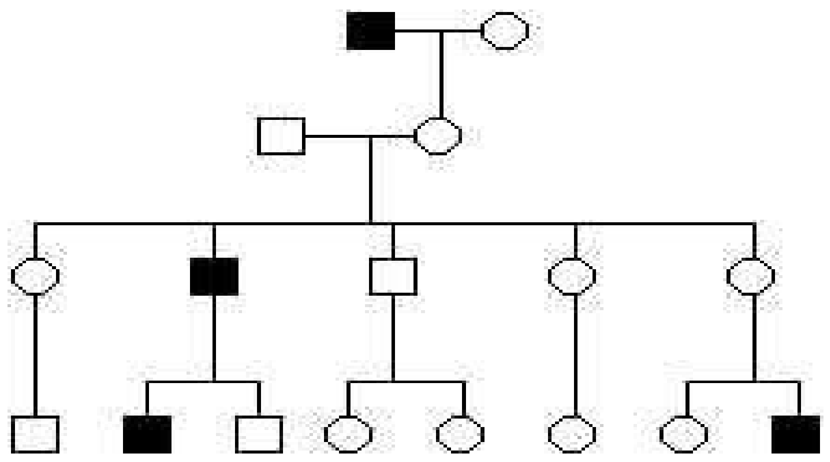
Biological Background "Chromosomes"

- Genetic information is stored in the chromosomes
- Each chromosome is built of DNA
- Genes are encoded in the chromosomes
- Genes code for proteins
- Every gene has a unique position on the chromosome



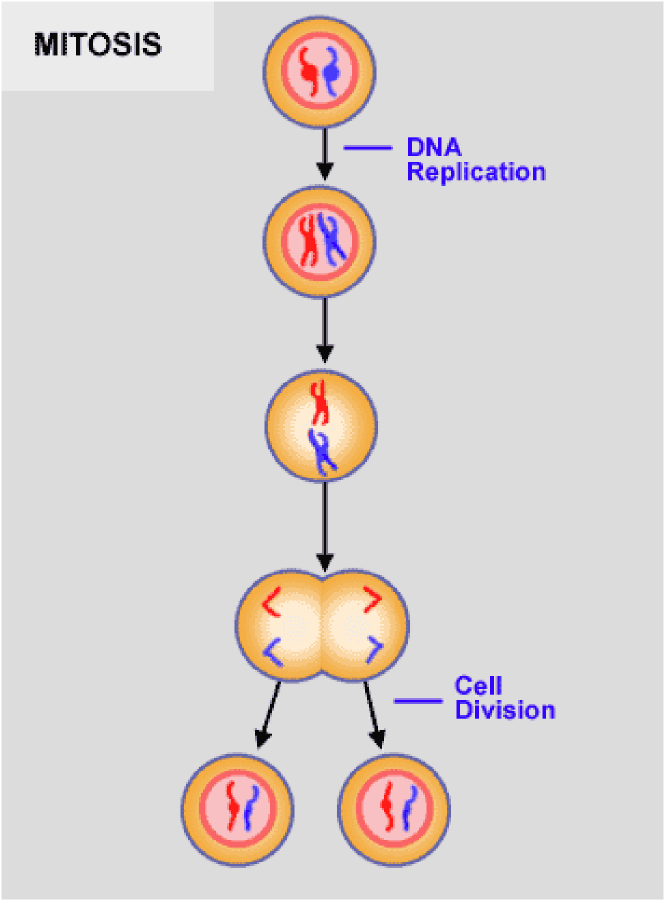
Biological Background: Genotype and phenotype

- The entire combination of genes is called genotype
- A genotype leads to a phenotype (eye color, height, disease predisposition)
- The phenotype is affected by changes to the underlying genetic code



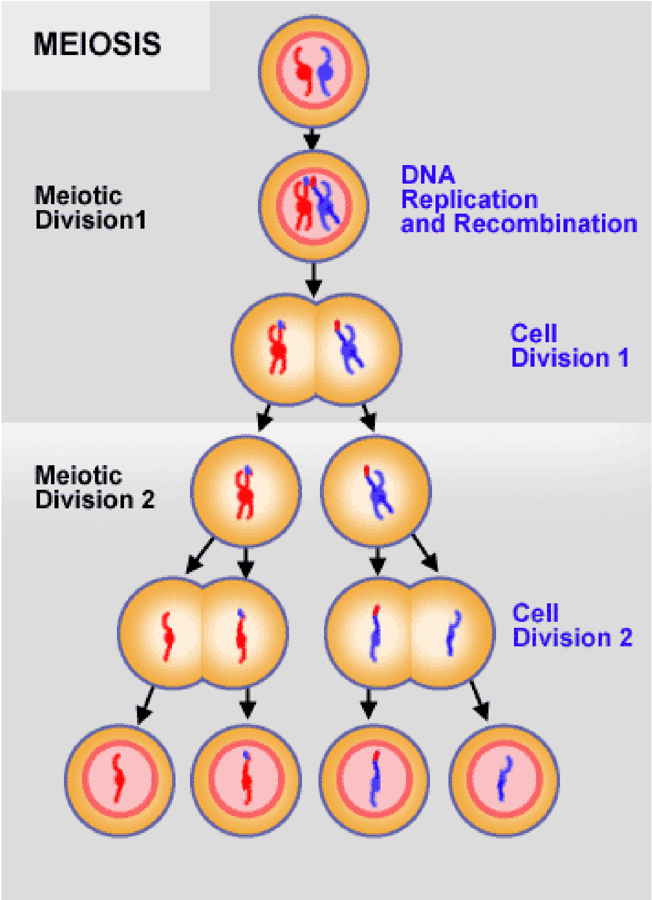
Biological Background “Reproduction ”

- Reproduction of genetical information
  - ***Mitosis***
  - ***Meiosis***
- Mitosis is copying the same genetic information to new offspring: there is no exchange of information
- Mitosis is the normal way of growing of multicell structures, like organs.



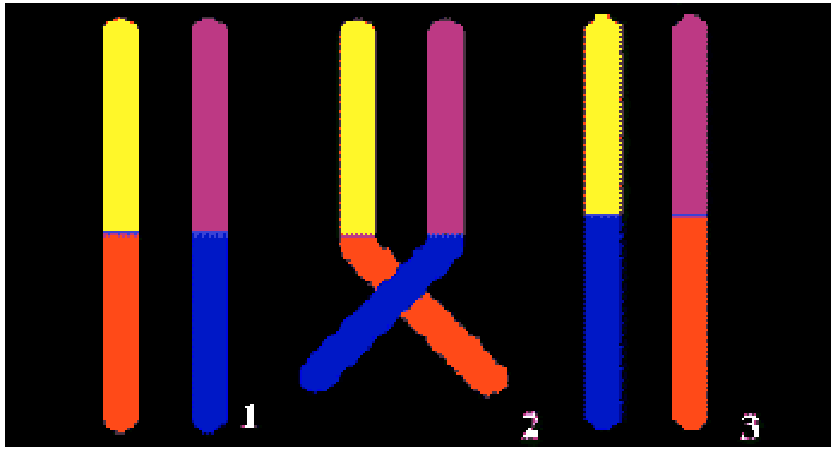
Biological Background Reproduction

- Meiosis is the basis of cross reproduction
- After meiotic division 2 gametes appear
- In reproduction two gametes conjugate to a zygote which will become the new individual
- Crossovers leads to new genotype



Mutations

- In any copying process errors can occur, so single (point) mutations are pretty common.
- Other types of errors, including affecting longer regions (either deletion, inversions, substitutions etc.) can also occur



"Natural selection"

- The origin of species: "Preservation of favourable variations and rejection of unfavourable variations."
- There are more individuals born than can survive, so there is a continuous struggle for life.
- Individuals with an advantage have a greater chance for survive: so survival of the fittest.

History of Algorithm

History of GAs

- As early as 1962, John Holland's work on adaptive systems laid the foundation for later developments.

- By the 1975, the publication of the book *Adaptation in Natural and Artificial Systems*, by Holland and his students and colleagues.

History of GAs

- early to mid-1980s, genetic algorithms were being applied to a broad range of subjects.

- In 1992 John Koza has used genetic algorithm to evolve programs to perform certain tasks. He called his method "genetic programming" (GP).

What is GA

- A genetic algorithm (or GA) is a search technique used in computing to find true or approximate solutions to optimization and search problems.
- (GA)s are categorized as global search heuristics.
- (GA)s are a particular class of evolutionary algorithms that use techniques inspired by evolutionary biology such as inheritance, mutation, selection, and crossover (also called recombination).

What is GA

- The evolution usually starts from a population of randomly generated individuals and happens in generations.

- In each generation, the fitness of every individual in the population is evaluated, multiple individuals are selected from the current population (based on their fitness), and modified to form a new population.

What is GA

- The new population is used in the next iteration of the algorithm.

- The algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population.

No convergence rule or guarantee!

Vocabulary

- **Individual** - Any possible solution
- **Population** - Group of all individuals
- **Fitness** – Target function that we are optimizing (each individual has a fitness)
- **Trait** - Possible aspect (features) of an individual
- **Genome** - Collection of all chromosomes (traits) for an individual.

Basic Genetic Algorithm

- Start with a large “population” of randomly generated “attempted solutions” to a problem
- Repeatedly do the following:
  - Evaluate each of the attempted solutions
  - (probabilistically) keep a subset of the best solutions
  - Use these solutions to generate a new population
- Quit when you have a satisfactory solution (or you run out of time)

GA Operators

- **Methods of representation**
- **Methods of selection**
- **Methods of Reproduction**

Common representation methods

- Binary strings.
- Arrays of integers (usually bound)
- Arrays of letters
- ….

Methods of Selection

There are many different strategies to select the individuals to be copied over into the next generation

Methods of Selection

- *Roulette-wheel selection*.
- *Elitist selection*.
- *Fitness-proportionate selection*.
- *Scaling selection*.
- *Rank selection*.
- ...

Roulette wheel selection

- Conceptually, this can be represented as a game of roulette - each individual gets a slice of the wheel, but more fit ones get larger slices than less fit ones.

Other selection methods

- *Elitist selection*:

  Chose only the most fit members of each generation.

- *Cutoff selection*:

  Select only those that are above a certain cutoff for the target function.

Methods of Reproduction

- There are primary methods:
  - *Crossover*
  - *Mutation*

Methods of Reproduction: Crossover

- Two parents produce two offspring
- Two options:

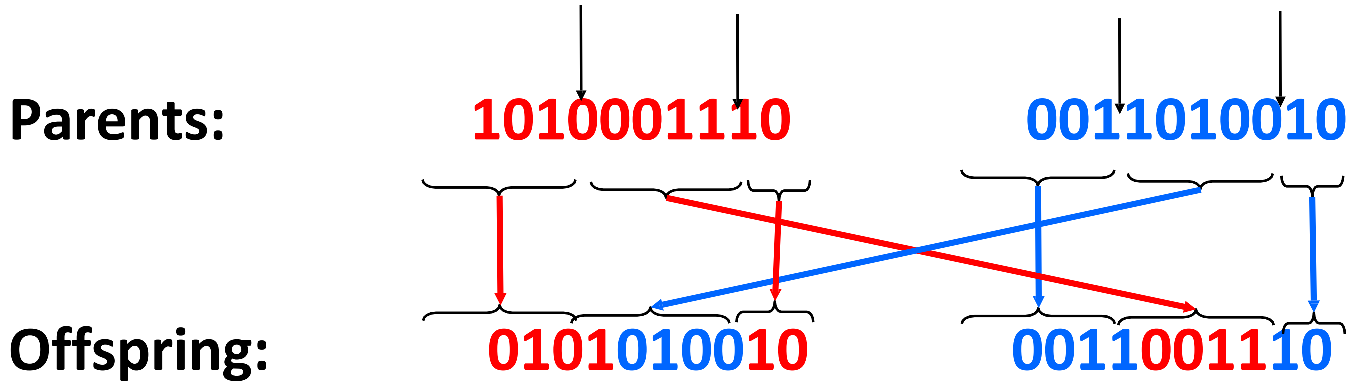
  1. The chromosomes of the two parents are copied to the next generation
  2. The two parents are randomly recombined (crossed-over) to form new offsprings

Several possible crossover strategies

- Randomly select a single point for a crossover
- Multi point crossover
- Uniform crossover

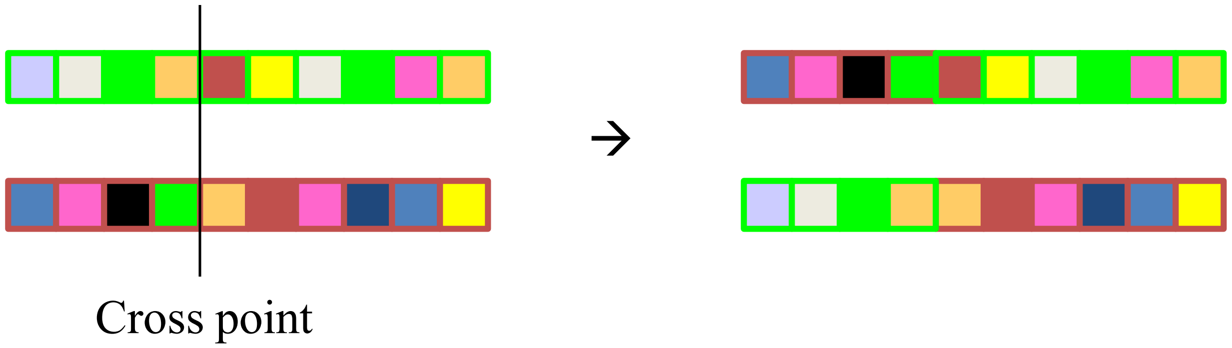
Two-point crossover

- Avoids cases where genes at the beginning and end of a chromosome are always split

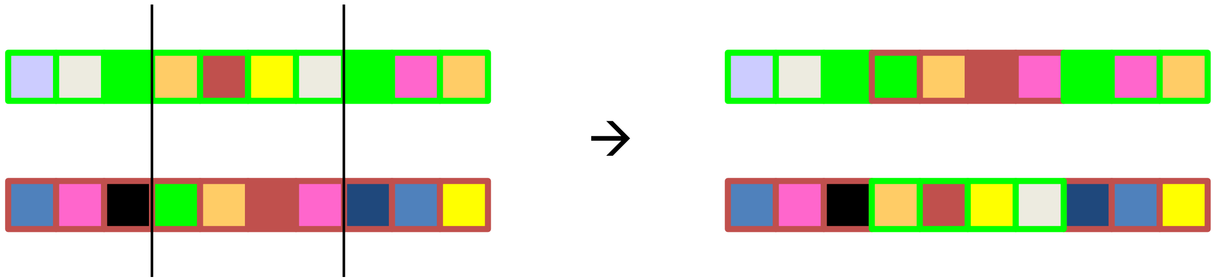


Crossover

- Single point crossover



- Two point crossover (Multi point crossover)



Uniform crossover

- A random subset is chosen
- The subset is taken from parent 1 and the other bits from parent 2.

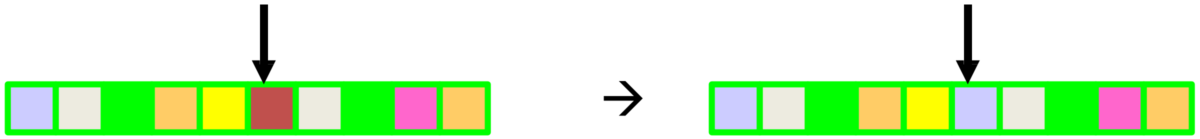
Subset: BAABBAABBB (Randomly generated)

Parents: 1010001110 0011010010

Offspring: 0011001010 1010010110

Methods of Reproduction: Mutations

– Generating new offspring from single parent



GA Applications

Domain	Application Type
Control	Gas pipeline, missile evasion
Design	Aircraft design, keyboard configuration, communication networks
Game playing	Poker, checkers
Security	Encryption and Decryption
Robotics	Trajectory planning

Benefits of Genetic Algorithms

- Concept is easy to understand
- Modular, separate from application
- Supports multi-objective optimization
- Always an answer; answer gets better with time.
- Easy to exploit previous or alternate solutions
- Flexible building blocks for hybrid applications.

Example

Below are the steps to be followed to solve any optimization problem with the help of GA.

- **Step 1-** Choose an encoding technique, a selection operator, and a crossover operator
- **Step 2-** Choose a population size
- **Step 3-** Randomly choose the initial population
- **Step 4-** Select parental chromosomes
- **Step 5-** Perform Crossover (random crossover points)
- **Step 6-** Evaluation of offsprings
- **Step 7-** Repeat the process

Question

Maximize the function $f(x)=x^2$, where x value range from 0-31

Steps

Step 1-

- **Encoding technique-** Binary encoding
- **Selection operator-** Roulette Wheel Selection
- **Crossover operator-** Single point crossover

Step 2-

Population size (n) = 4

Step 3-

Initial population (x value) = 13, 24, 8, 19

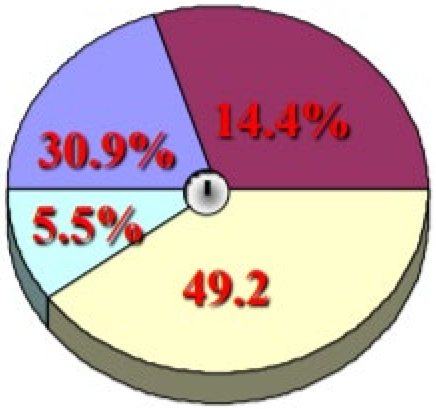
Step 4

defines the roulette wheel
F(x)/Total F(X)

F(x)/Average F(X)

Number of times the parent will get
selected if wheel is spun 'n' times
AC = round(EC)

String No	Initial Population	X Value	F(x) value	Probability count	Expected count	Actual Count
1	01101	13	169	0.14	0.58	1
2	11000	24	576	0.49	1.97	2
3	01000	8	64	0.06	0.22	0
4	10011	19	361	0.31	1.23	1
Total			1170	1	4	
Average			293			



Weighted Roulette wheel

Contd…

We see that if the Roulette wheel is spun four times, we'll get 24 twice and 13 and 19 once. So possible parental combinations are (24,13) and (24,19).

String 2	1 1 0 0 0	Parental combination 1
String 1	0 1 1 0 1	
String 2	1 1 0 0 0	Parental combination 2
String 4	1 0 0 1 1	

Step 5

	Parents	Offsprings
String 2	1 1 0 0 | 0	1 1 0 0 1
String 1	0 1 1 0 | 1	0 1 1 0 0
String 2	1 1 | 0 0 0	1 1 0 1 1
String 4	1 0 | 0 1 1	1 0 0 0 0

Step 6

String No	Offspring 1	X Value	F(x) value
1	0 1 1 0 0	12	144
2	1 1 0 0 1	25	625
3	1 1 0 1 1	27	729
4	1 0 0 0 0	10	256

We can see that the maximum f(x) value has increased from 576 to 729.

Step 7

Now we’ll take these four offsprings as parents and repeat the process until our termination condition is not satisfied.

Q. Solve using GA.

Suppose a genetic algorithm uses chromosomes of the form $x = abcdefgh$ with a fixed length of eight genes. Each gene can be any digit between 0 and 9. Let the fitness of individual x be calculated as:

$$f(x) = (a + b) - (c + d) + (e + f) - (g + h) ,$$

and let the initial population consist of four individuals with the following chromosomes:

x1 = 6 5 4 1 3 5 3 2

x2 = 8 7 1 2 6 6 0 1

x3 = 2 3 9 2 1 2 8 5

x4 = 4 1 8 5 2 0 9 4

Contd…

a) Evaluate the fitness of each individual, showing all your workings, and arrange them in order with the fittest first and the least fit last.

b) Perform the following crossover operations:

i) Cross the fittest two individuals using one–point crossover at the middle point.

ii) Cross the second and third fittest individuals using a two–point crossover (points b and f).

iii) Cross the first and third fittest individuals (ranked 1st and 3rd) using a uniform crossover.

c) Suppose the new population consists of the six offspring individuals received by the crossover operations in the above question. Evaluate the fitness of the new population, showing all your workings. Has the overall fitness improved?

Continue…

d) By looking at the fitness function and considering that genes can only be digits between 0 and 9 find the chromosome representing the optimal solution (i.e. with the maximum fitness). Find the value of the maximum fitness.