Genetic Algorithms

Anirban Mukhopadhyay

Department of Computer Science and Engineering

University of Kalyani

anirban@klyuniv.ac.in, www.anirbanm.in

Today's Talk

- Genetic Algorithm
- Multiobjective Genetic Algorithm
- Genetic Algorithm based clustering
- Microarray Gene Expression data clustering

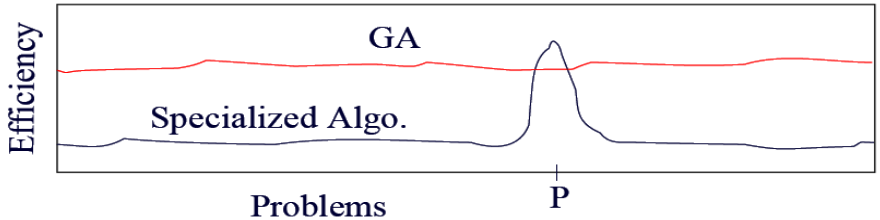
Genetic Algorithm (GA)

- Randomized search and optimization technique guided by the principle of natural genetic systems.
- Inspired by the biological evolution process
- Uses concepts of "Natural Selection", "Genetic Inheritance" and "Survival of the Fittest" (Darwin 1859)
- Originally developed by **John Holland** (1975) and gained popularity during late 80's.

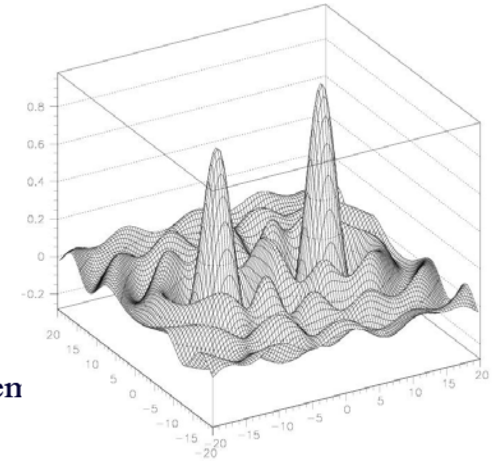
Genetic Algorithms

Why GA?

- Most real life problems can not be solved in polynomial amount of time using any deterministic algorithm.
- Sometimes near optimal solutions that can be generated quickly are more desirable than optimal solutions which require huge amount of time.
- When the problem can be modeled as an optimization one.
- Efficiently searches for global optima when multiple optima exist.
- Parallel implementation is easy.



Specialized algorithms – best performance for special problems
Genetic algorithms – good performance over a wide range of problen



GA Features

- Evolutionary Search and Optimization Technique
- Principles of Evolution (survival of the fittest and inheritance).
- Work with encoding of the parameter set.
- Searches from a population of points.
- Uses probabilistic transition rules.

GA vs. Nature

• **A solution (phenotype)**	**Individual**
• Representation of a solution (genotype)	Chromosome
• **Components of the representation**	**Genes**
• Solution’s quality (fitness function)	Individual’s degree of ability to adopt with surrounding
• **Stochastic operators**	**Selection, Crossover (reproduction), Mutation**

Simple GA

Produce an initial population of individuals

Evaluate the fitness of all individuals

While termination condition not met **do**

  select fitter individuals for reproduction

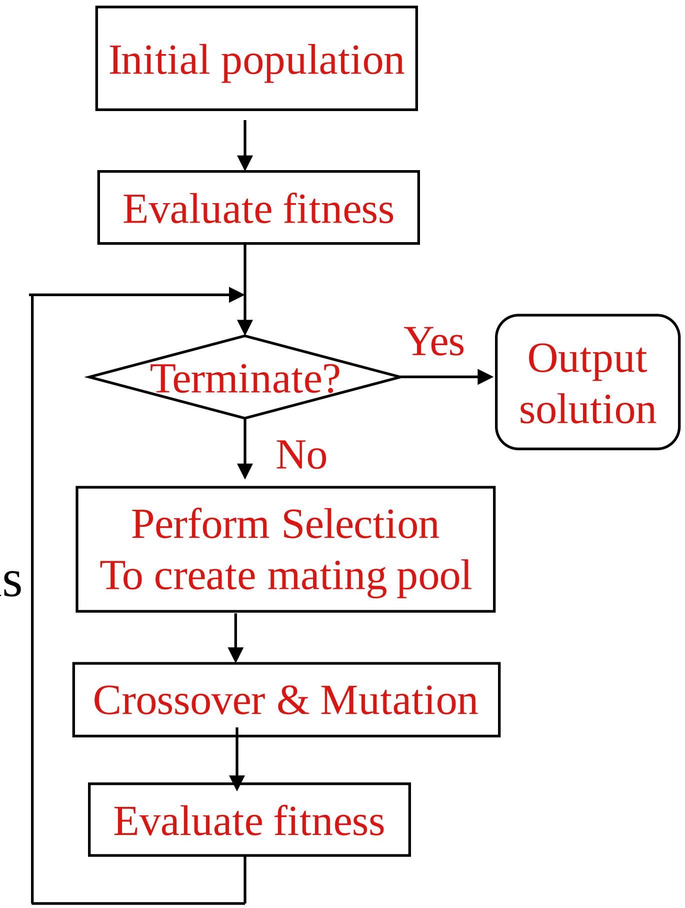
  Recombine (crossover) between individuals

  Mutate individuals

  Evaluate the fitness of the modified individuals

  Generate a new population

End while

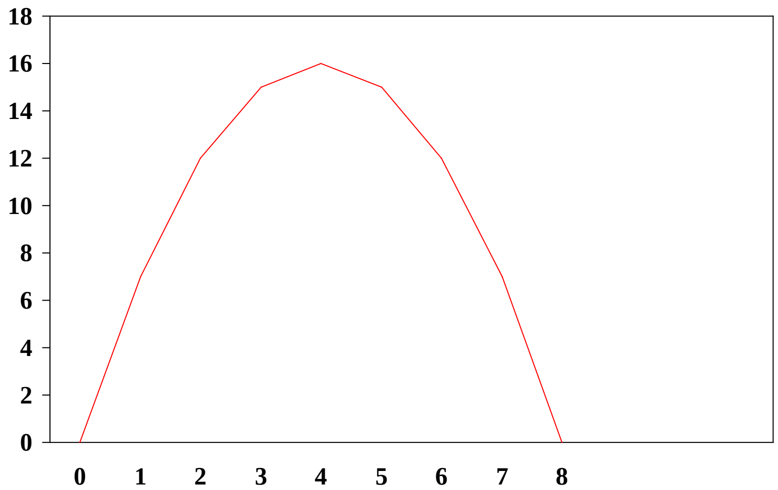


Encoding and Population

- **Chromosome encodes a solution in the search space**
  - Usually as strings of 0's and 1's
  - If *l* is the string length, number of different chromosomes (or strings) is 2^l
- **Population**
  - A set of chromosomes in a generation
  - Population size is usually constant
  - Common practice is to choose the initial population randomly.

Encoding and Population - Example

Optimize $f(x) = x(8 - x)$, $x=[0,8]$



Binary String of 8 bits

0-255 $\longleftrightarrow$ 0-8

1	**0**	**0**	**1**	**1**	**0**	**1**	**0**

= 154

$$x = 0 + (8/255) * 154 = 4.8313$$

Fitness Evaluation

- Fitness/objective function is associated with each chromosome
- This indicates the degree of goodness of the encoded solution
- The only problem specific information (also known as the payoff information) that GA uses
- If minimization problem is to be solved then

  fitness = 1/objective.

Fitness Evaluation - Example

Function *f(x) = x(8-x)*

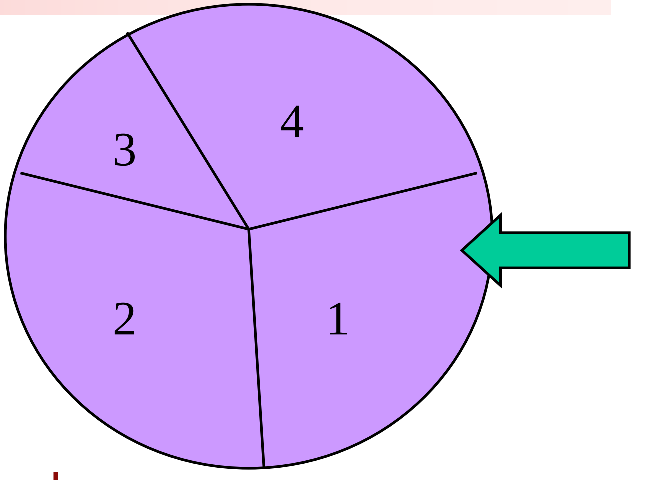
Population (size = 4)	**Corresponding *x***	**Fitness/ Objective Fn.**
1 0 01 1 0 1 0	4.8313	15.3089
0 1 1 0 0 1 1 1	3.2313	15.4091
0 0 0 1 0 1 0 1	0.6588	4.8363
1 0 1 1 1 1 0 0	5.8980	12.3975

Selection

- More copies to good strings
- Fewer copies to bad string
- Proportional selection scheme
  - Number of copies taken to be directly proportional to its fitness
  - Mimics the natural selection procedure to some extent.
- **Roulette wheel selection** and **Tournament selection** are two frequently used selection procedures.

Roulette Wheel Selection - Example

Chromosome #	Fitness
1	15.3089
2	15.4091
3	4.8363
4	12.3975



Individual i will have a probability $\frac{f(i)}{\sum_i f(i)}$ to be chosen

- Spin 1 Chromosome 2 is selected
- Spin 2 Chromosome 1 is selected
- Spin 3 Chromosome 2 is selected
- Spin 4 Chromosome 4 is selected

Mating Pool

0 1 1 0 0 1 1 1
1 0 0 1 1 0 1 0
0 1 1 0 0 1 1 1
1 0 1 1 1 1 0 0

Roulette Wheel Selection - Implementation

- Sort the solutions in descending order of fitness.
- Compute the cumulative sum.

0

Chrom 2

0 + 15.4091 = **15.3089**

Chrom 1

0 + 15.4091 + 15. 3089 = **30.7180**

Chrom 4

0 + 15.3089 + 15.4091 + 12.3975 = **43.1155**

Chrom 3

0 + 15.3089 + 15.4091 + 12.3975 + 4.8363 = **47.9518**

Generate random number between 0 & 47.9518

Tournament - Selection

- Repeat until mating pool is full
  - Select a set (size < population size) of chromosomes randomly.
  - Copy the best chromosome among them into the mating pool.
- Usually tournament size is 2 (**binary tournament**).
- The chromosome with lowest fitness value can never be copied into the mating pool.

Crossover

- Exchange of genetic information
- It takes place between randomly selected parent chromosomes
- **Single point crossover** and **Uniform crossover** are the most commonly used schemes.
- Probabilistic operation

Single Point Crossover – Example

1	**0**	**0**	**1**	**1**	**0**	**1**	**0**
0	**1**	**1**	**1**	**1**	**1**	**0**	**1**

Here l (string length) = 8. Let k (crossover point) = 5

Offspring formed:

1	**0**	**0**	**1**	**1**	**1**	**0**	**1**
0	**1**	**1**	**1**	**1**	**0**	**1**	**0**

Uniform Crossover – Example

Parents:

1	0	0	1	1	0	1	0

0	1	1	1	1	1	0	1

Mask:

1	1	0	1	0	1	1	0

Offspring formed:

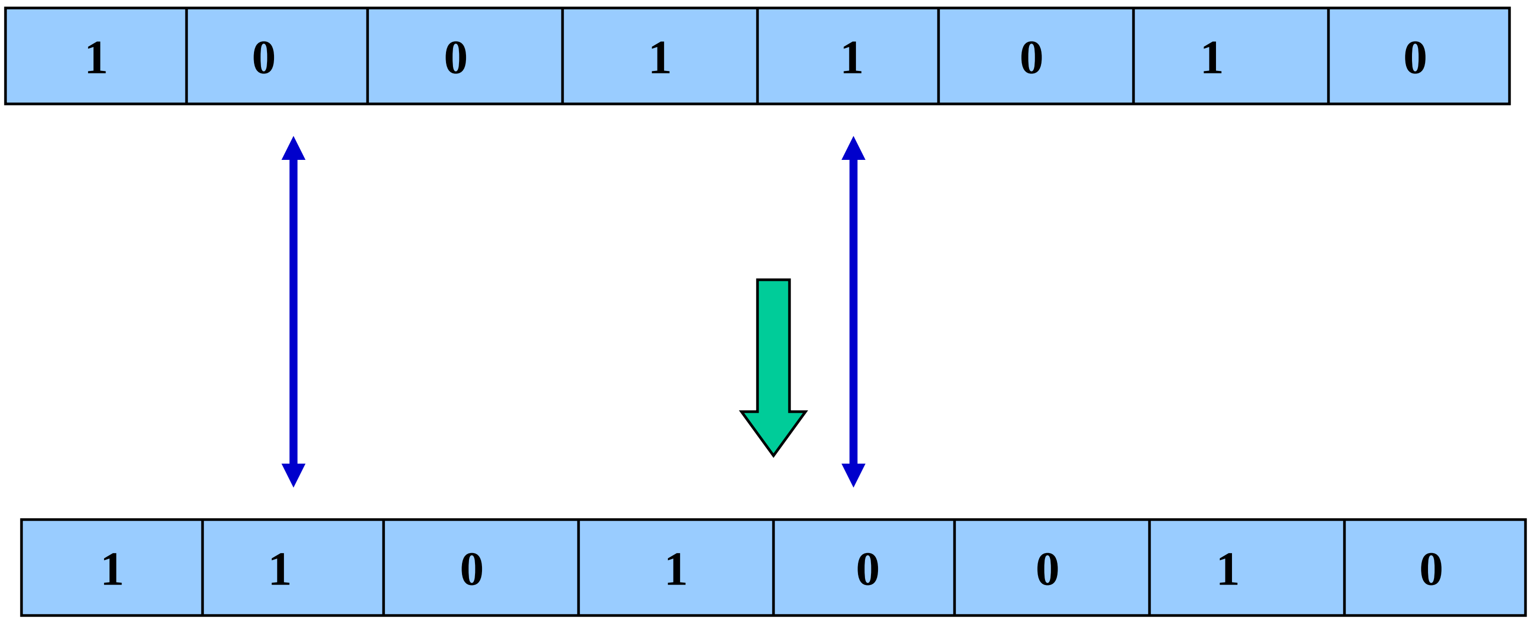
0	1	0	1	1	1	0	0

1	0	1	1	1	0	1	1

Mutation

- **Random alteration** in the genetic structure
- Introduces **genetic diversity** into the population.
- Exploration of new search areas
- Mutating a binary gene involves simple **negation of the bit**
- Mutating a real coded gene defined in a variety of ways
- Probabilistic operation

Mutation – Example



Parameters

- Population size – usually fixed
- String length – usually fixed
- Probabilities of crossover, μ_c, and mutation, μ_m

  μ_c is kept high and μ_m is kept low.
- Termination criteria
- Parameters are often manually tuned
- Sometimes may be adaptive.

Parameters – Example

For the example being considered,

- $P = 4$, $l = 8$.

- But for most realistic cases P is usually chosen in the range 50-100.

- $\mu_c = [0.6\text{-}0.9]$,

- $\mu_m = [0.01\text{-}0.1]$.

- l usually depends on the required precision

Termination Criterion

The cycle of selection, crossover and mutation are repeated for number of times until one of these occurs

- Average fitness value of a population more or less constant over several generations,
- Desired objective function value is attained by at least one string in the population,
- **Number of generations (or iterations) is greater than some threshold ----- most commonly used.**

Elitist Model of GAs

The best string seen up to the current generation is preserved in a location either inside or outside the population.

Variation of Fitness over Generation