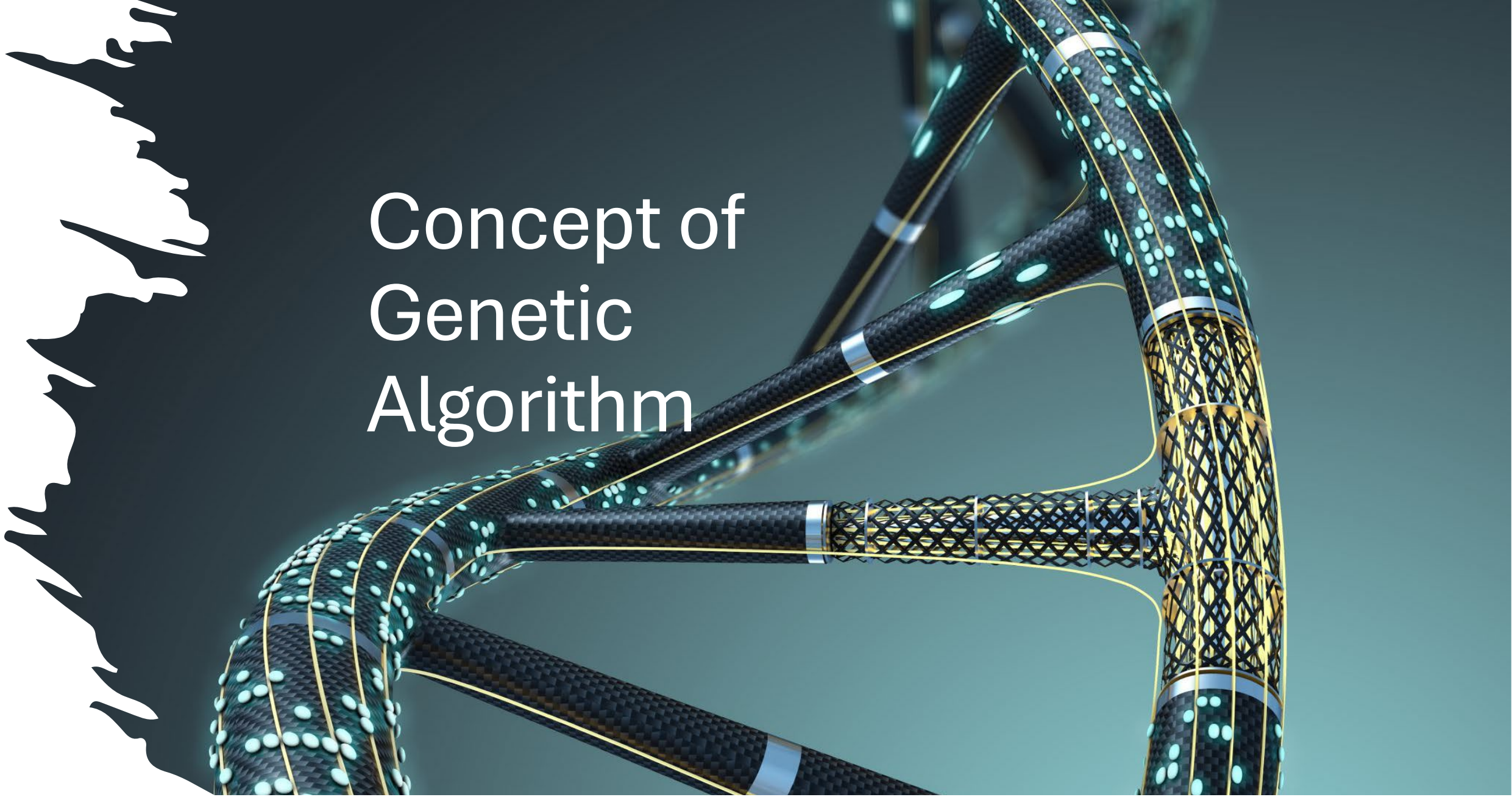
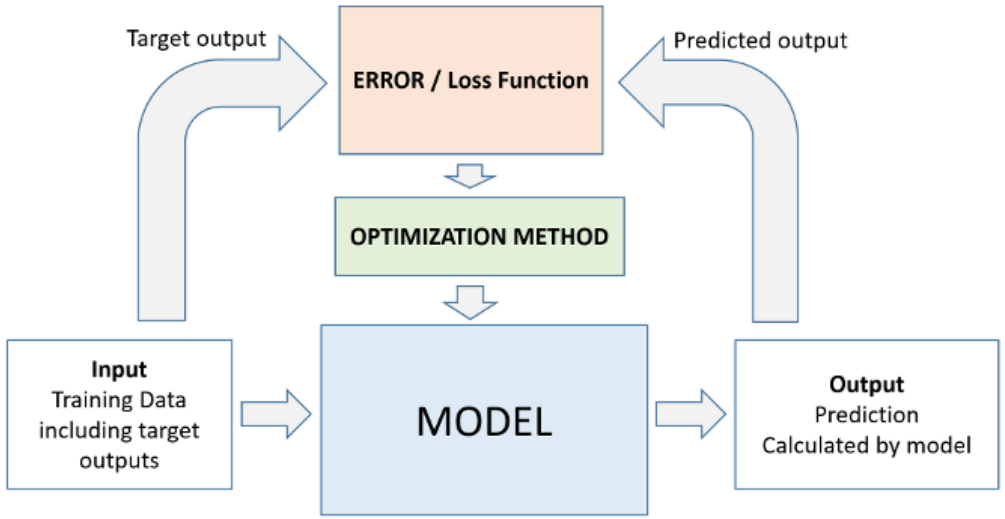
Concept of Genetic Algorithm



Optimization Techniques in Machine Learning

- Optimization is the process of selecting the best solution out of the various feasible solutions that are available.

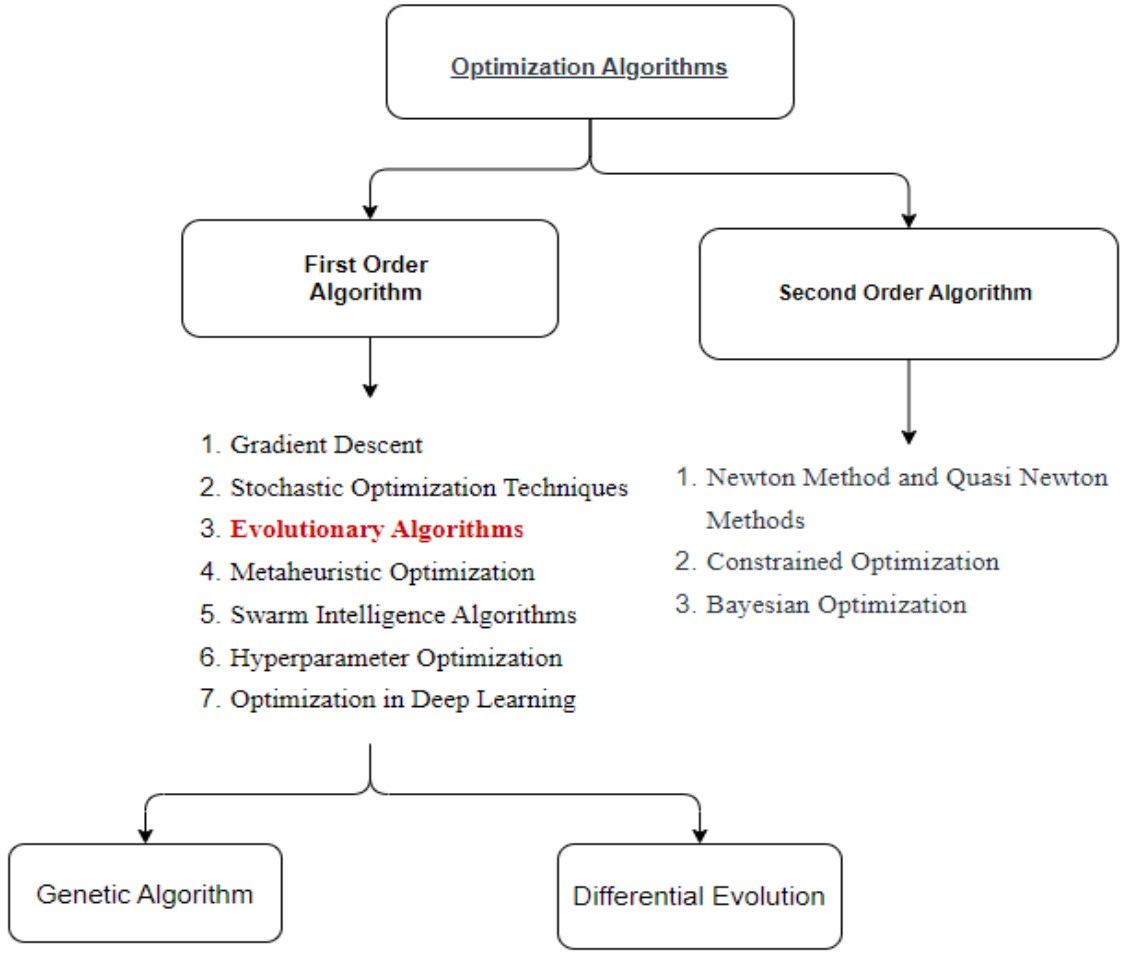


Why Optimization?

- Model Training
- Improving Accuracy
- Reducing Computational Cost
- Enhancing Generalization
- Hyperparameter Tuning
- Finding Global solution

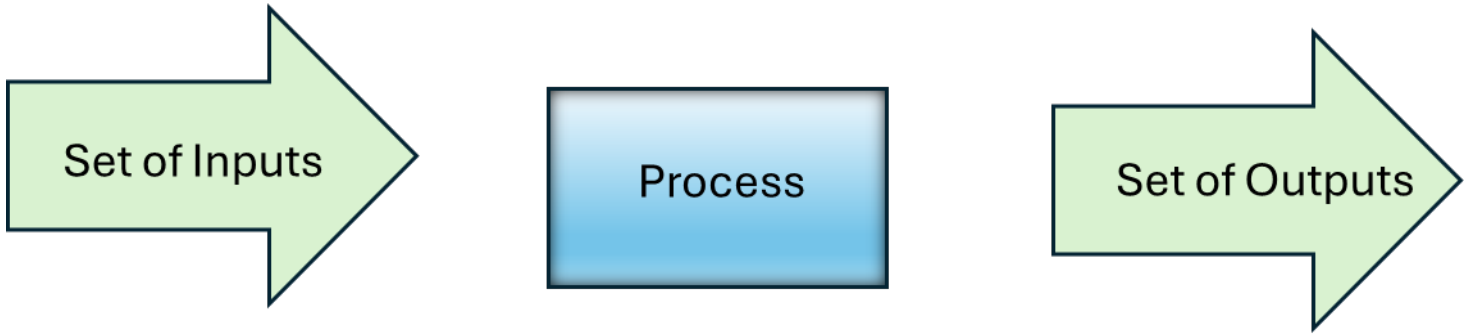
Optimization Techniques in Machine Learning

- Types of Optimization:



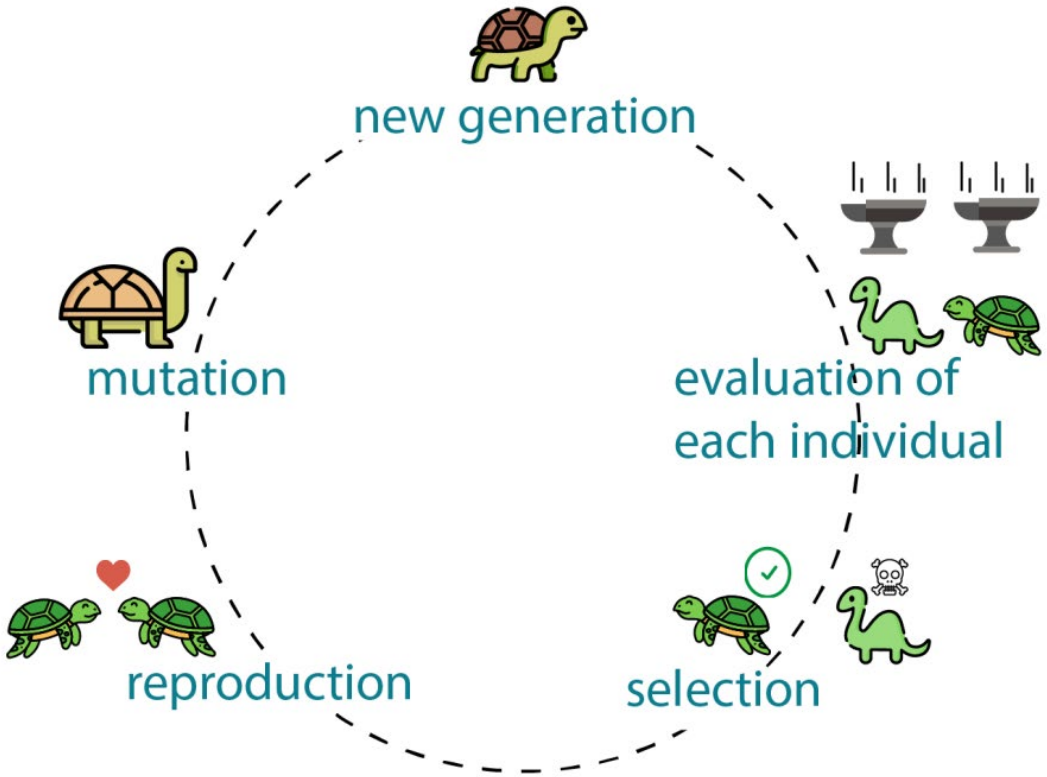
Genetic Algorithms – An Overview

- Genetic Algorithm (GA) is a search-based optimization technique based on the principles of Genetics and Natural Selection.
- It is frequently used to find optimal or near-optimal solutions to difficult problems which otherwise would take a lifetime to solve.
- Optimization is the process of making something better.



What are Genetic Algorithms?

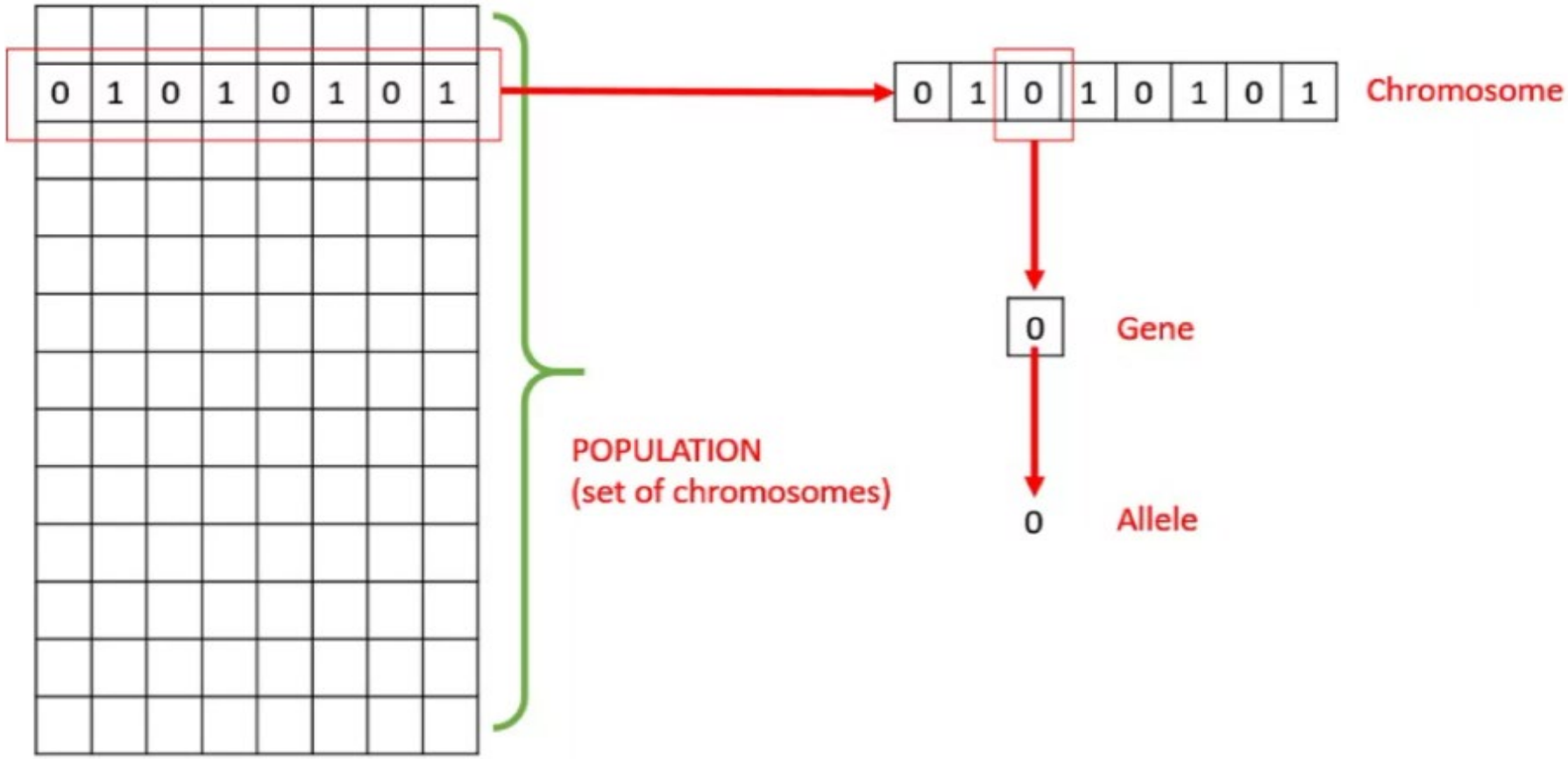
- Genetic algorithms are categorized as global heuristics.
- Genetic algorithms are a particular class of evolutionary algorithms that uses a techniques inspired by evolutionary biology such as inheritance, mutation, selection, and cross over (recombination)



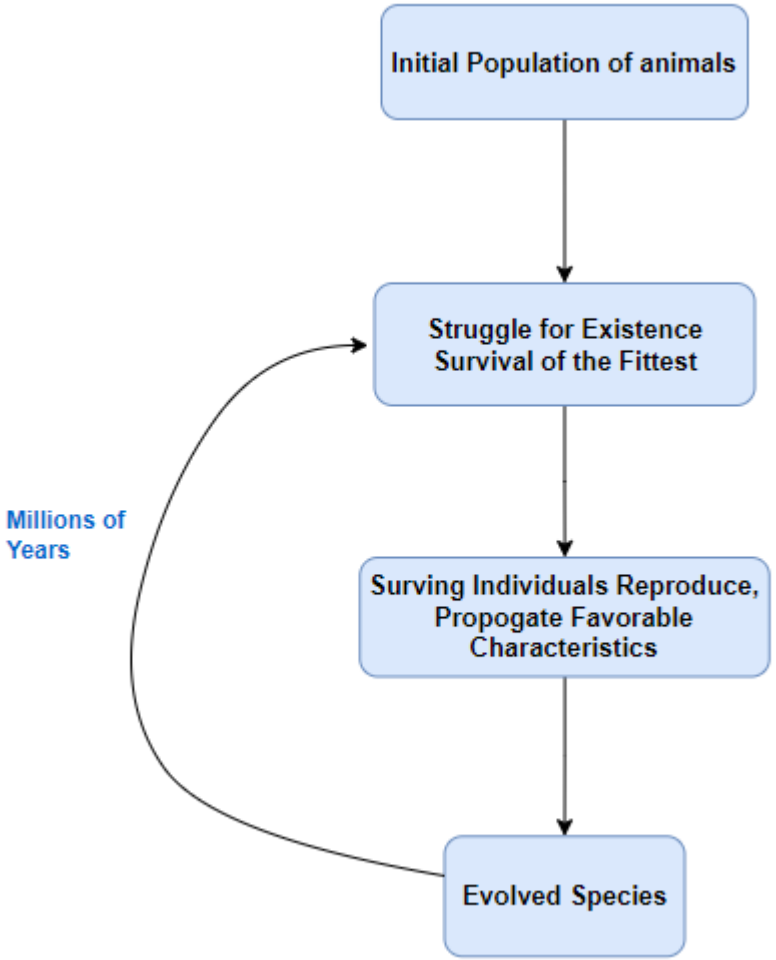
Basic Terminology

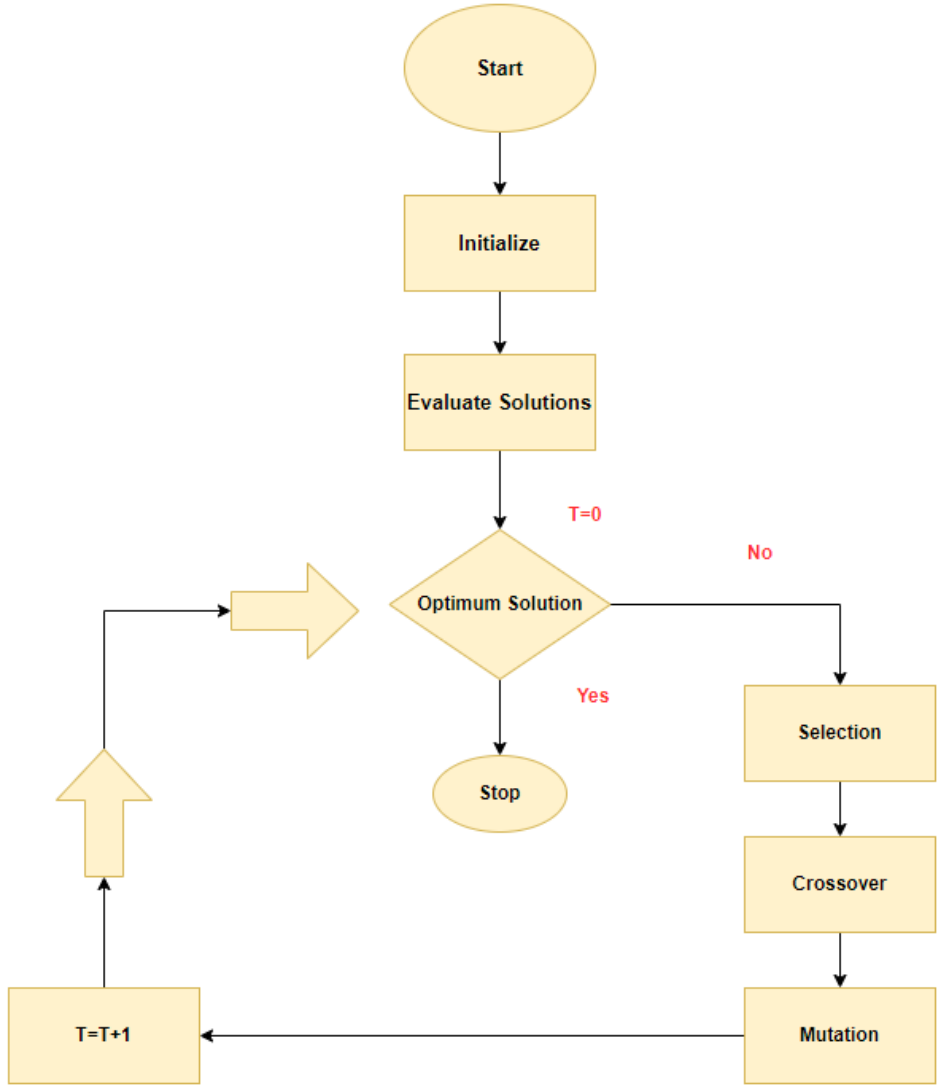
- Population – It is a subset of all the possible (encoded) solution to the given problem.
- Chromosomes – One such solution to the given problem.
- Gene – One element position of a chromosome.
- Allele – Value a gene takes for a particular chromosome.

Basic Terminology



Basic Concepts – How it works?





Basic Concepts – How it works?

GA converts design space into genetic space

- Works with coding variables
- Traditional optimization techniques are deterministic in nature, but GA uses randomized operators
- 3 important aspects
  1. Definition of objective function.
  2. Definition and implementation of genetic representation.
  3. Definition and implementation of genetic operators.

Key Components

- Genotype – The population in the computation space. In the computation space, the solution are represented in a way which can be easily understood and manipulated using computation.
- Phenotype - The population in the actual real world solution space in which solutions are represented in a way they are represented in real world solutions.
- Decoding – Process of transforming a solution from genotype to phenotype
- Encoding - Process of transforming from phenotype to genotype.

Key Components

- Fitness Function – A function which takes the solution as input and produces the suitability of the solution as the output.
- Genetic Operators – These alters the genetic composition of the offsprings.

- **Initialization**
- **Selection**
- **Crossover**
- **Mutation**
- **Fitness Evaluation**
- **Termination**

Simple Genetic Algorithm

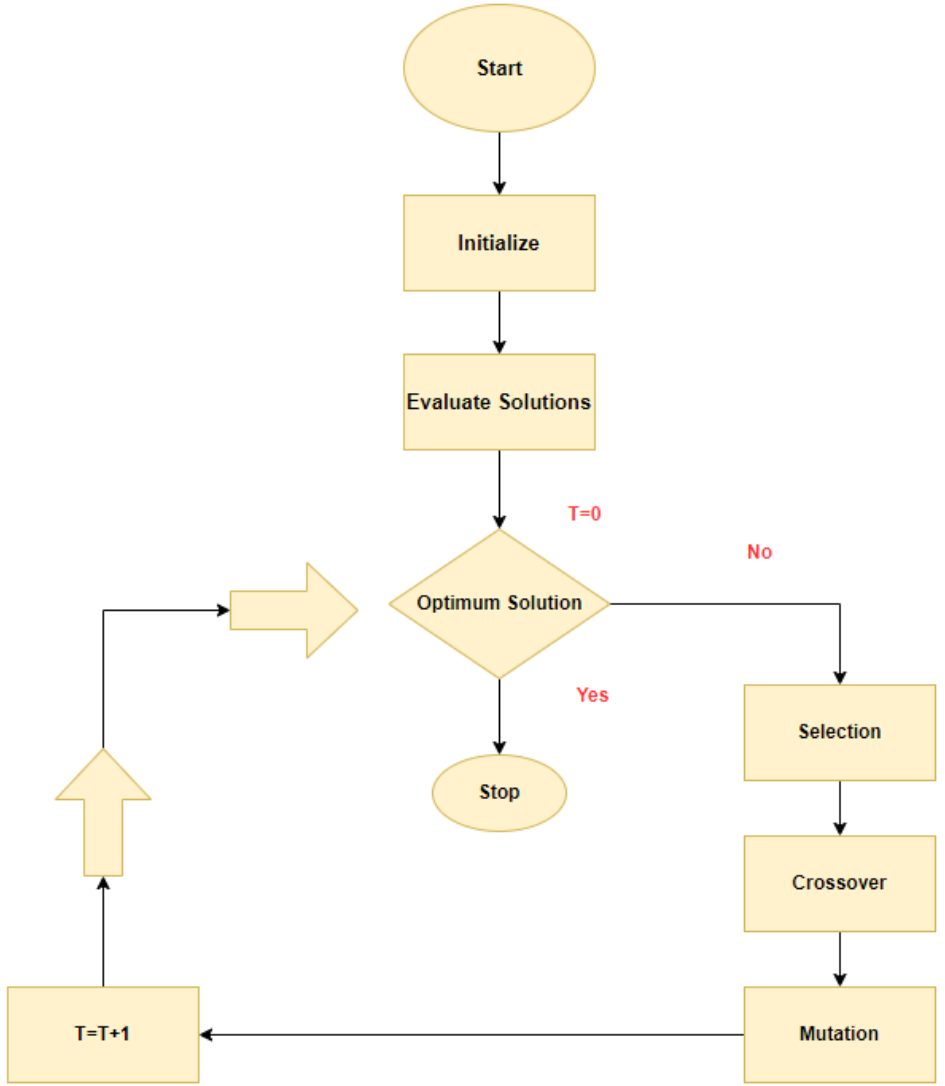
```
function sga ()
        {
        Initialize population;
        Calculate fitness function;

        While(fitness value != termination criteria)
                {
                    Selection;

                    Crossover;

                    Mutation;

                Calculate fitness function;
                }
        }
```



Example: Fitness Function in TSP

- Fitness is defined as the inverse of the total distance traveled, where shorter distances result in higher fitness. A fitness function value quantifies the optimality of a solution. The value is used to rank a particular solution against all the other solutions

- The goal is to find the route that minimizes the total distance while visiting each city exactly once.

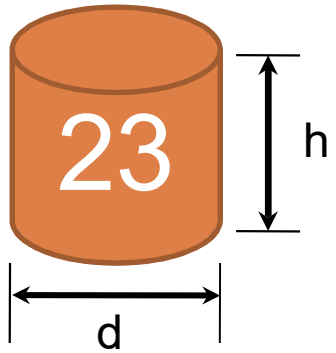
- A fitness value is assigned to each solution depending on how close it is actually to the optimal solution of the problem

Fitness Function

$$\begin{aligned} \text{Minimize} \quad & f(d,h) = c((\pi d^2/2) + \pi dh), \\ \text{Subject to} \quad & g_1(d,h) \equiv (\pi d^2 h/4) \geq 300, \\ \text{Variable bounds} \quad & d_{\min} \leq d \leq d_{\max}, \\ & h_{\min} \leq h \leq h_{\max}. \end{aligned}$$

Considering c = 0.0654

$$\begin{aligned} F(s) &= 0.0654(\pi(8)^2/2 + \pi(8)(10)), \\ &= 23, \end{aligned}$$



1 Selection Strategy

- The process that determines which solutions are to be preserved and allowed to reproduce and which ones deserve to die out.

"Select the best, discard the rest"

- Techniques: Roulette Wheel, Tournament Selection, Proportionate selection, Rank Selection, Steady state selection, *etc*
- Ensuring better solutions get passed on

Functions of Selection operator

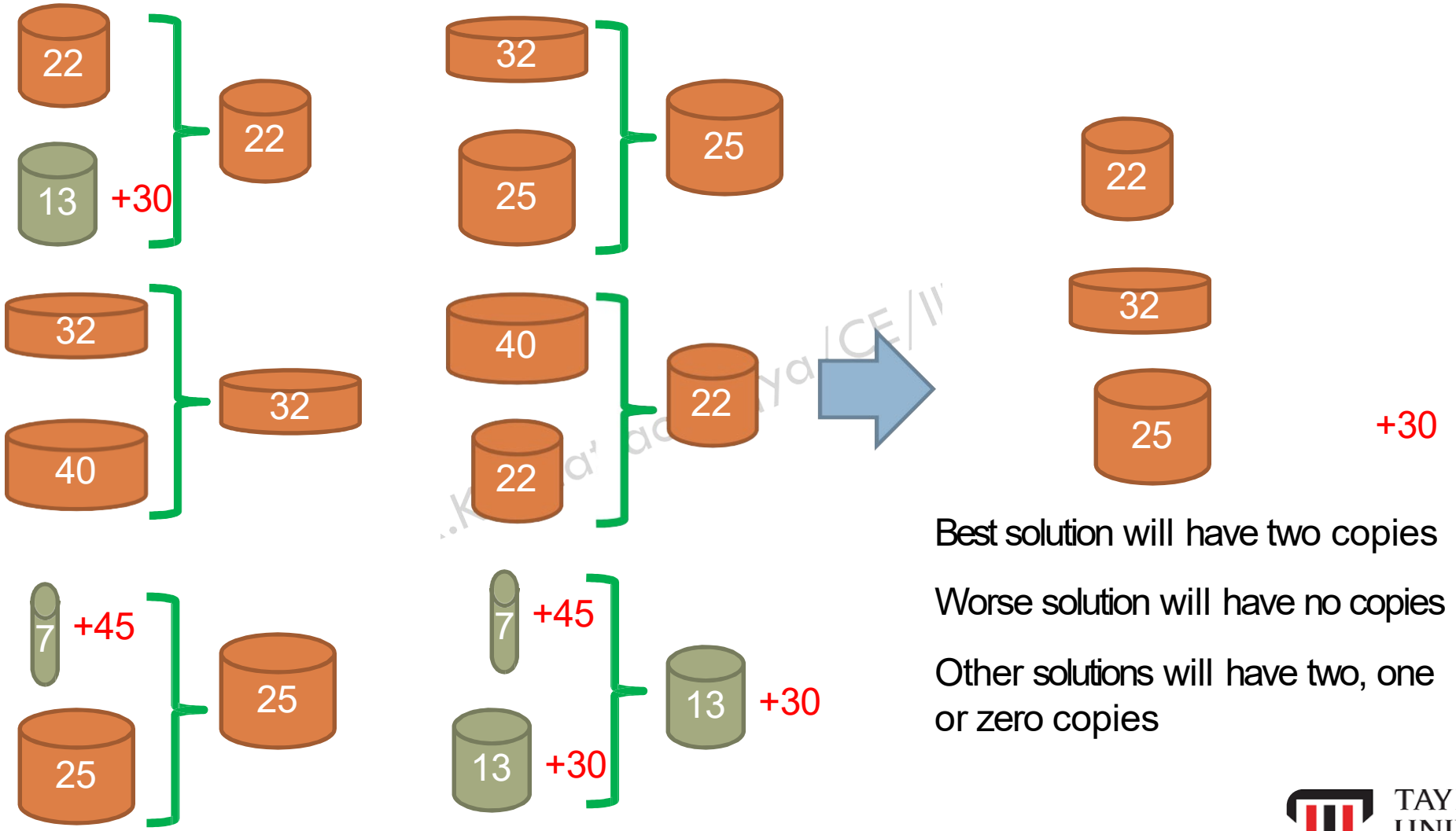
- Identify the good solutions in a population
- Make multiple copies of the good solutions
- Eliminate bad solutions from the population so that multiple copies of good solutions can be placed in the population

- **Now how to identify the good solutions?**

Tournament selection

- In tournament selection several tournaments are played among a few individuals. The individuals are chosen at random from the population.
- The winner of each tournament is selected for next generation.
- Selection pressure can be adjusted by changing the tournament size.
- Weak individuals have a smaller chance to be selected if tournament size is large.

Tournament selection

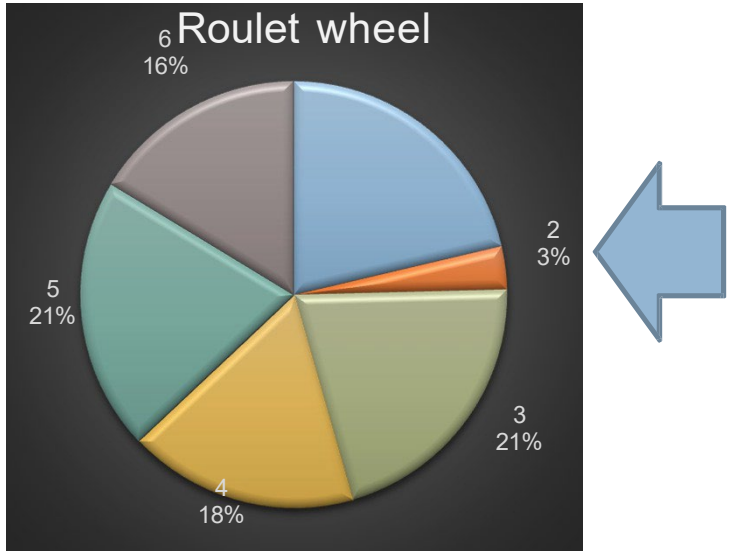


Roulette wheel and proportionate selection

Parents are selected according to their fitness values

Chrom #	Fitness	% of RW	EC	AC
1	50	26.88	1.61	2
2	6	3.47	0.19	0
3	36	20.81	1.16	1
4	30	17.34	0.97	1
5	36	20.81	1.16	1
6	28	16.18	0.90	1
	186	100.00	6	6

The better chromosomes have more chances to be selected

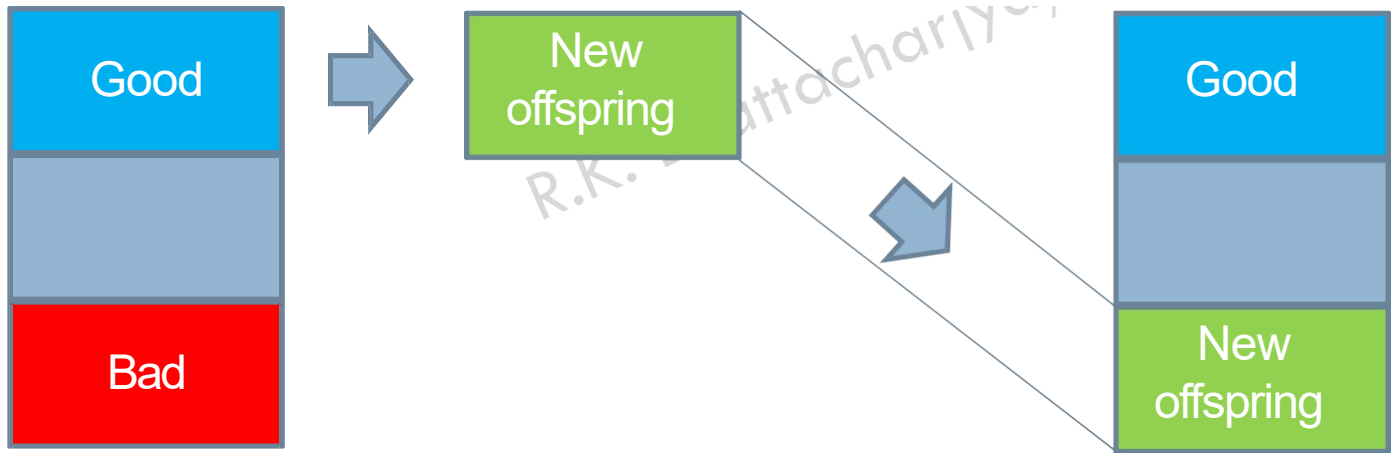


Steady state selection

In this method, a few good chromosomes are used for creating new offspring in every iteration.

Then some bad chromosomes are removed and the new offspring is placed in their places

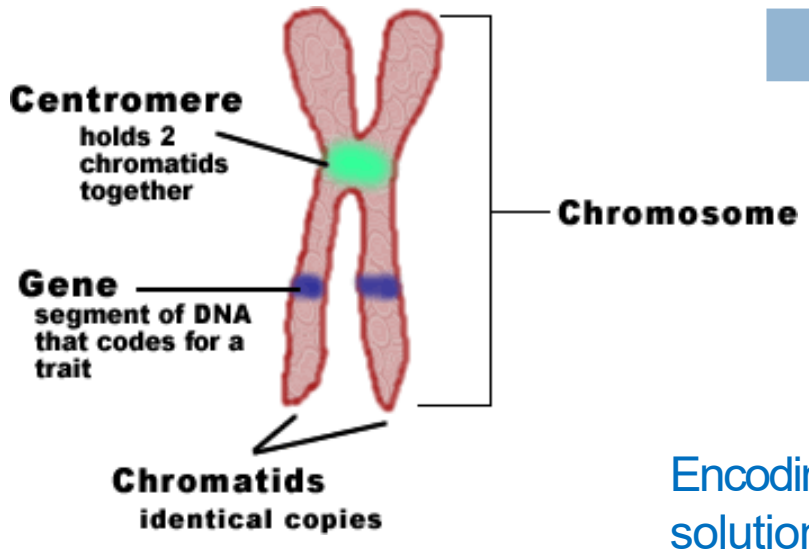
The rest of population migrates to the next generation without going through the selection process.



2. Crossover

The crossover operator is used to create new solutions from the existing solutions available in the mating pool after applying selection operator.

This operator exchanges the gene information between the solutions in the mating pool.



Encoding of solution is necessary so that our solutions look like a chromosome

Source: http://www.biologycorner.com/bio1/celldivision-chromosomes.html

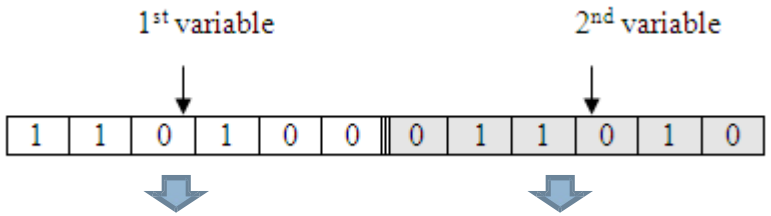
Encoding

- The process of representing a solution in the form of a string that conveys the necessary information.

- Just as in a chromosome, each gene controls a particular characteristic of the individual, similarly, each bit in the string represents a characteristic of the solution.

Encoding Methods

- Most common method of encoding is **binary coded**. Chromosomes are strings of 1 and 0 and each position in the chromosome represents a particular characteristic of the problem



Decoded value 52 26

Mapping between decimal and binary value

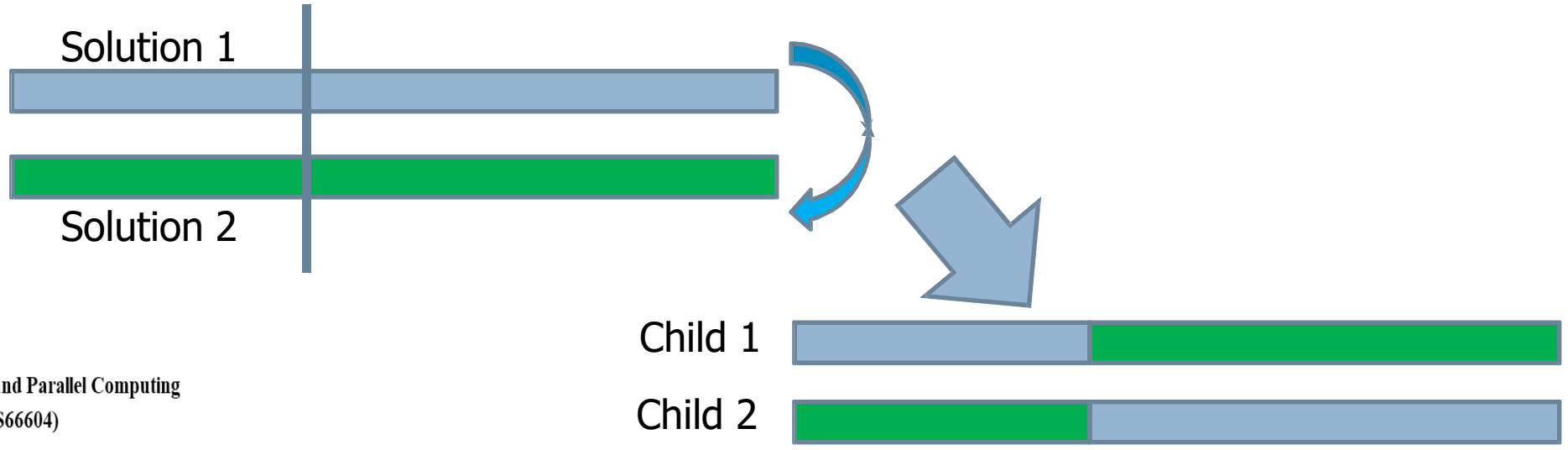
$$x_i = x_i^{\min} + \frac{x_i^{\max} - x_i^{\min}}{2^{l_i} - 1} \mathrm{DV}(s_i)$$

Crossover operator

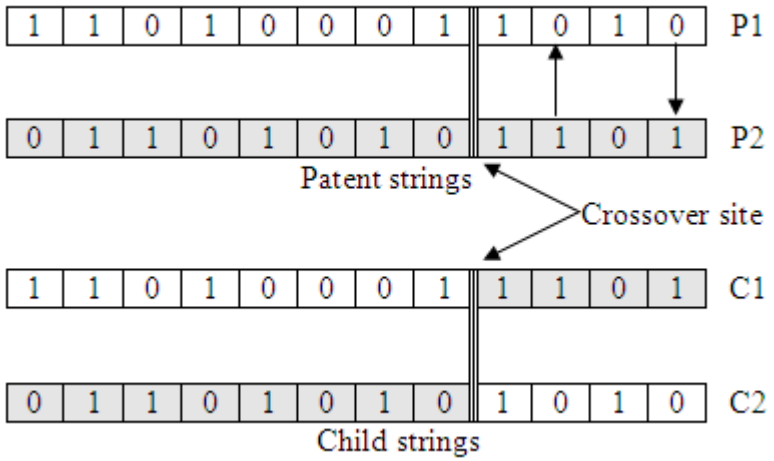
The most popular crossover selects any two solutions strings randomly from the mating pool and some portion of the strings is exchanged between the strings.

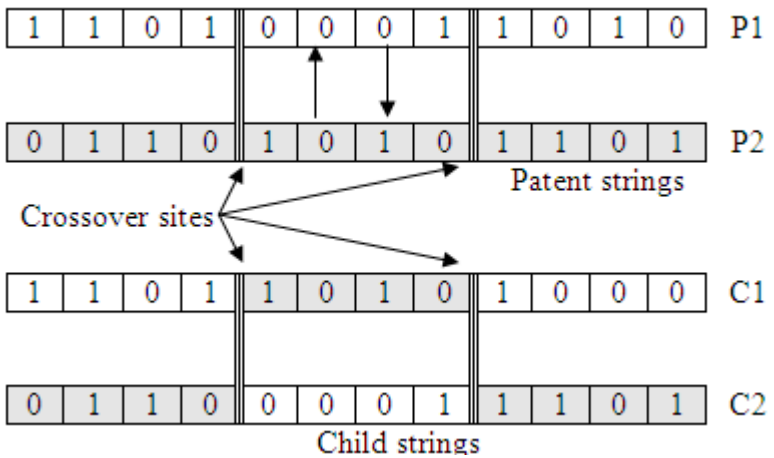
The selection point is selected randomly.

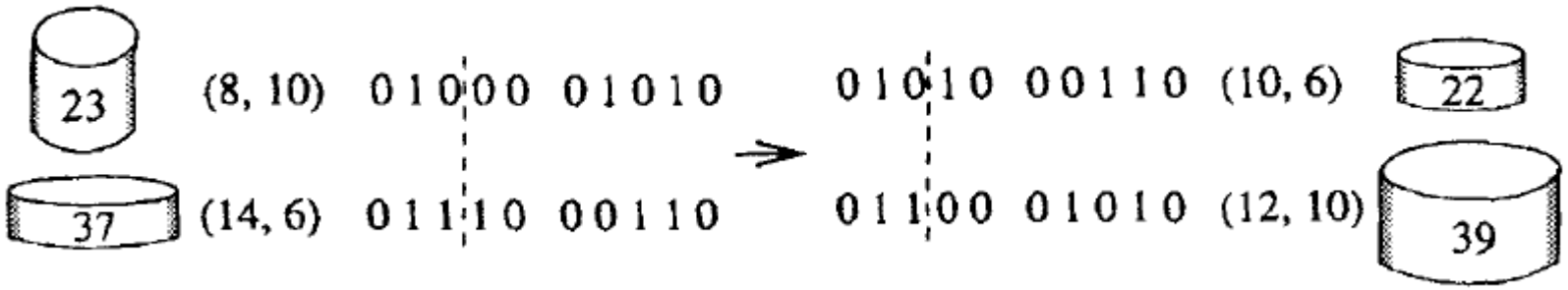
A probability of crossover is also introduced in order to give freedom to an individual solution string to determine whether the solution would go for crossover or not.



Binary Crossover







Source: Deb 1999

3. Mutation operator

- Mutation is the occasional introduction of new features into the solution strings of the population pool to maintain diversity in the population.
- Though crossover has the main responsibility to search for the optimal solution, mutation is also used for this purpose.

Before mutation

After mutation

Termination Condition

Example: Real-World ML Application of GA

1. **Feature Selection in Medical Diagnosis**:

- Problem: Selecting the most relevant medical tests and biomarkers to diagnose a disease accurately while minimizing the number of tests.

2. Optimizing Stock Portfolio Allocation:

- Problem: Finding the optimal combination of assets to maximize returns while minimizing risk, considering constraints like budget and asset diversity.

3. Game Strategy Optimization:

- Problem: Evolving strategies in computer games or simulations to find the most effective approach to win or achieve a specific goal.

Any Question?