



PES University, Bangalore

(Established under Karnataka Act No. 16 of 2013)

APRIL 2022: IN SEMESTER ASSESSMENT (ISA) B.TECH. IV SEMESTER

UE20MA251- LINEAR ALGEBRA

Project / Seminar

Session: Jan-May 2022

Branch : Computer Science and Engineering

Semester & Section : Semester IV Section B

Sl No.	Name of the Student	SRN	Marks Allotted (Out of 10)
1.	Anushka Siri Raghunandan	PES1UG20CS073	
2.	Vaishnavi V B	PES1UG20CS902	
3.	Devika S	PES1UG20CS123	
4.	Praneetha B	PES1UG20CS106	

Name of the Course Instructor : **Prof. Chaitra GP**

Signature of the Course Instructor

(with Date) : _____

TITLE: LINEAR ALGEBRA IN GENETICS

INTRODUCTION

The project explores the various applications of Linear Algebra in Genetics as well as Genetic Engineering. The following are the center of our discussion:

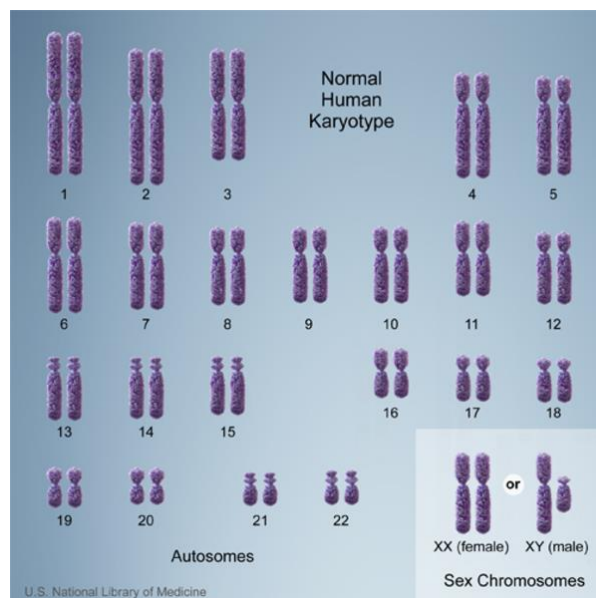
- Inheritance
- Genetic Algorithms
- DNA sequencing

LITERATURE REVIEW, THEORETICAL BACKGROUND:

Genetics is the branch of biology which deals with the study of inheritance, or the transmission of traits from one generation to the next. There are several modes of inheritances, but our main focus will be on:

- Autosomal inheritance
- X-linked inheritance

In autosomal inheritance, each heritable trait is assumed to be governed by a single gene on an autosomal chromosome.



There are two different forms or alleles of a gene(denoted by A and a). Each individual in a population carries a pair of alleles, which may be similar or

different. These pairs are called an individual's genotype, and there are three possible genotypes for a particular trait: AA, Aa, or aa.

It is the genotype that determines how the trait controlled by the gene is manifested in the individual. For example, in humans, eye coloration is controlled through autosomal inheritance. Genotypes AA and Aa have brown eyes, and genotype aa has blue eyes. In such a case, the A allele is dominant over the allele, or that the allele is recessive to the A allele, since genotype Aa has the same outward trait as genotype AA.

The genotype distribution of a particular trait in a population in the n-th generation can be represented by a genotype vector

$$x_n = \begin{bmatrix} a_n \\ b_n \\ c_n \end{bmatrix}$$

where a_n , b_n , and c_n denote the portion of the population with genotype AA, Aa, aa, respectively in the n-th generation. Since the genotype distribution changes over time, we can represent the succession of genotype distributions from one generation to the next in the form of a difference equation,

$$x_n = Mx_{n-1}$$

We will proceed to diagonalize M (that is, we will seek to find an invertible matrix P and a diagonal matrix D such that $M = PDP^{-1}$). Since,

$$M_n = PD_nP^{-1}$$

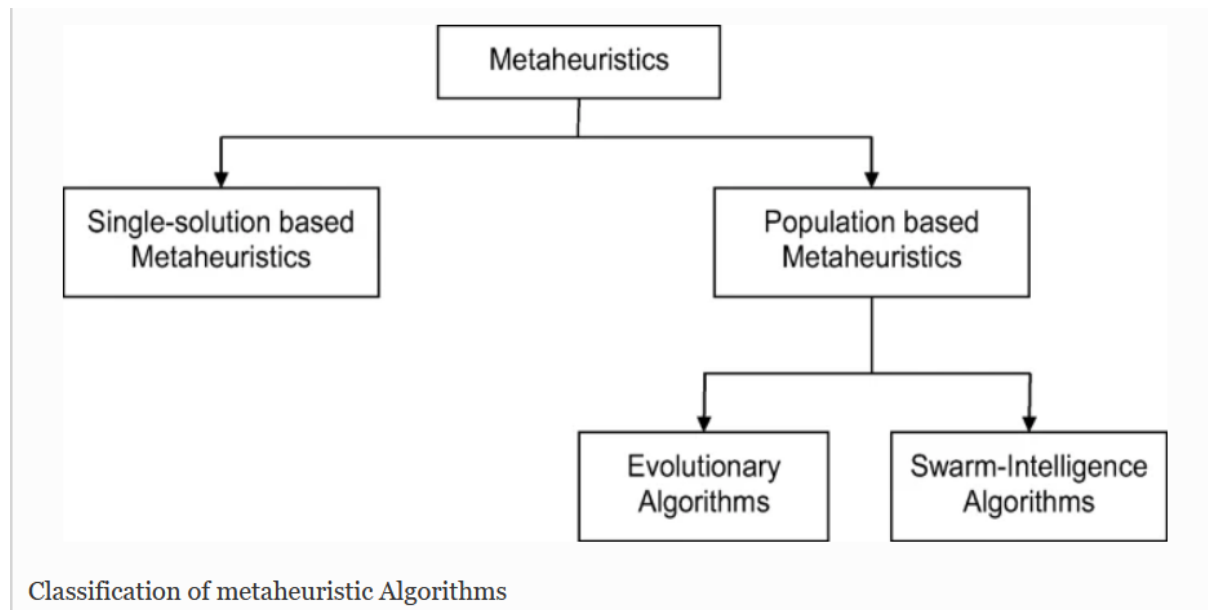
Once P and D have been specified, we will have,

$$x_n = PD_nP^{-1}x_0$$

In X-linked inheritance, patterns differ depending on the type of inheritance. Recessive X linked conditions are always passed through the maternal line with the condition appearing in males and being carried in females, but not usually expressed. Dominant X-linked conditions can be passed from either affected parent to their children. If the father has the affected copy of the gene he will only pass the condition to his daughters and not his sons. Mothers can pass X-linked dominant conditions to both their sons and daughters. Examples of X-linked conditions include Haemophilia and Duchenne muscular dystrophy.

The logic behind the calculation of the inheritance probability remains the same as autosomal inheritance.

In the recent years, metaheuristic algorithms are used to solve real-life complex problems arising from different fields such as economics, engineering, politics, management, and engineering. Intensification and diversification are the key elements of metaheuristic algorithm. The proper balance between these elements are required to solve the real-life problem in an effective manner. Most of metaheuristic algorithms are inspired from biological evolution process, swarm behavior, and physics' law.



Among the metaheuristic algorithms, Genetic algorithm (GA) is a well-known algorithm, which is inspired from biological evolution process. GA mimics the Darwinian theory of survival of fittest in nature. GA was proposed by J.H. Holland in 1992. The basic elements of GA are chromosome representation, fitness selection, and biological-inspired operators. Typically, the chromosomes take the binary string format. In chromosomes, each locus (specific position on chromosome) has two possible alleles (variant forms of genes) - 0 and 1. Chromosomes are considered as points in the solution space. These are processed using genetic operators by iteratively replacing its population. The fitness function is used to assign a value for all the chromosomes in the population. The biological-inspired operators are selection, mutation, and crossover. In selection, the chromosomes are selected on the basis of its fitness value for further processing. In crossover operator, a random locus is chosen and it changes the subsequences between chromosomes to create off-

springs. In mutation, some bits of the chromosomes will be randomly flipped on the basis of probability.

DNA sequencing determines the order of the four chemical building blocks - called "bases" - that make up the DNA molecule. The sequence tells scientists the kind of genetic information that is carried in a particular DNA segment. For example, scientists can use sequence information to determine which stretches of DNA contain genes and which stretches carry regulatory instructions, turning genes on or off. In addition, and importantly, sequence data can highlight changes in a gene that may cause disease.

FUTURE WORK:

Researchers now are able to compare large stretches of DNA - 1 million bases or more - from different individuals quickly and cheaply. Such comparisons can yield an enormous amount of information about the role of inheritance in susceptibility to disease and in response to environmental influences. In addition, the ability to sequence the genome more rapidly and cost-effectively creates vast potential for diagnostics and therapies.

GAs have been applied in different fields by modifying the basic structure of GA. The optimality of a solution obtained from GA can be made better by overcoming the current challenges. Some future possibilities for GA are as follows:

1. There should be some way to choose the appropriate degree of crossover and mutation operators. For example Self-Organizing GA adapt the crossover and mutation operators according to the given problem. It can save computation time that make it faster.
2. Future work can also be considered for reducing premature convergence problem. Some researchers are working in this direction. However, it is suggested that new methods of crossover and mutation techniques are required to tackle the premature convergence problem.

Genetic algorithms mimic the natural evolution process. There can be a possible scope for simulating the natural evolution process such as the responses of human immune system and the mutations in viruses.

3. In real-life problems, the mapping from genotype to phenotype is complex. In this situation, the problem has no obvious building blocks or building blocks are not adjacent groups of genes. Hence, there is a possibility to develop novel encoding schemes to different problems that does not exhibit same degree of difficulty.

BIBLIOGRAPHY:

1. <https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Fact-Sheet>
2. <https://link.springer.com/article/10.1007/s11042-020-10139-6#Sec44>
3. <https://www.genomicseducation.hee.nhs.uk/glossary/x-linked-inheritance/>
4. <https://sites.math.washington.edu/~king/coursedir/m308a01/Projects/m308a01-pdf/kirkham.pdf>