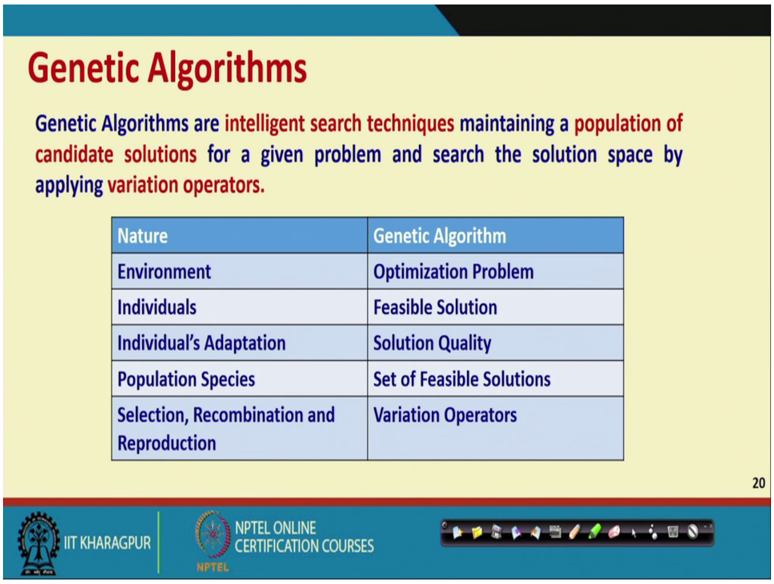
Selected Topics in Decision Modeling
Prof. Biswajit Mahanty
Department of Industrial and Systems Engineering
Indian Institute of Technology, Kharagpur

Lecture – 32
Genetic Algorithms

So, in our course; Selected Topics in Decision Modeling, now we are in our 32nd lecture that is on Genetic Algorithms, right. Now in our previous lecture, we have seen the Metaheuristics and one of the very important Metaheuristics are the evolutionary computing methods.

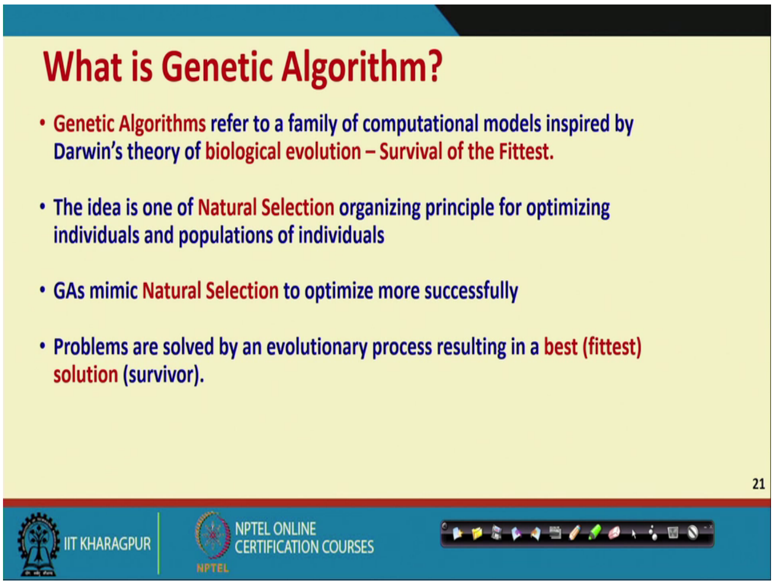
(Refer Slide Time: 00:40)



And they are all essentially coming from the genetic algorithms, right. So, the as the name suggests, the genetic algorithm really coming from the essential idea of genetics so, they are all nature inspired algorithms. So, here is a comparison, you can see that environment in nature in genetic algorithm is the optimization problem.

The feasible solutions are the individuals, the solution quality and individuals adaptation to the environment are comparable set of feasible solutions and the population species and the various operators the selection recombination and reproduction.

(Refer Slide Time: 01:47)



So, what really happening is that the genetic algorithms are intelligent search technique maintaining a population of candidate solutions for a given problem and search the solution space by applying variation operators.

See what really happens? The biological evolution process is one of survival of the fittest. So, what exactly it means? It means that assume that in a particular place you know the there are a large number of trees. So, assume that place is very near to a hilly region. So, what happens in a hilly region the sun's rays are coming, but sun's rays are not coming uniformly to all the trees, it is the tall trees, which gets the sun's rays, but this smaller trees are not getting the sun's rays. So, what will happen out there?

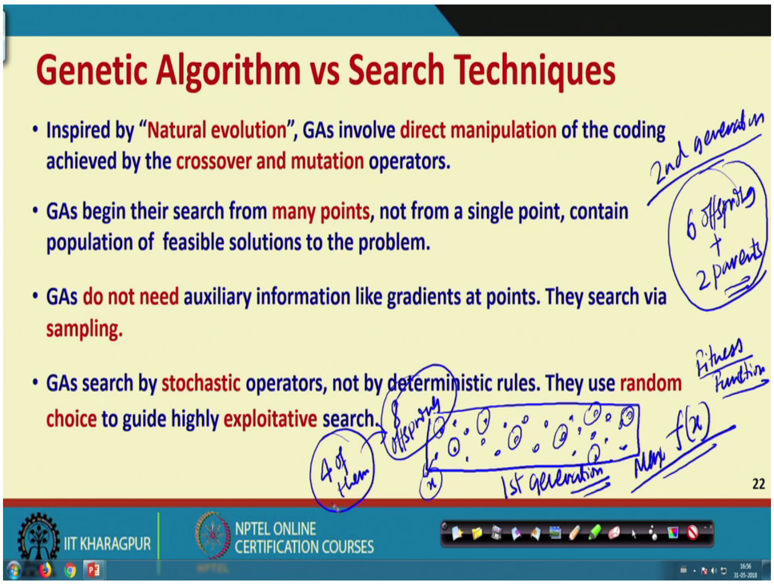
You see a over generations, you will find the that particular place in a hilly region, you only find tall trees and you will not find small trees; what is the reason for it? The reason is that it is not that there were not all kinds of trees we had all kinds of trees, but it is the tall trees which survives over generations and the small trees they die out because nature selection does not favor them.

The same thing about you know the animals you find in the hilly regions; you know who really are well equipped to face the winter you know the kind of animals, you find in the plane region if they really go without far to the hilly regions, they will not survive.

So, this is the process of natural selection, the tall trees grow near the mountains the animals with far; they are in the wintry regions the summer regions, we do not find animals with lot of far. So, these are the things that we actually observed in the natural situations in nature the question is that, how nature does this selection; what is the essential process of nature the essential process of nature is that you know that the parents who was selected in a natural process are by a process called survival of the fittest the fittest.

The fittest species, you know the those specific what you call parents are selected for reproduction, who are more fit, then the others alright and then the chromosomes, they join each other by processes like cross over and mutation to really produce offspring, is it alright? So, that is the essential biological process the genetic algorithms mimics it right. So, mimics it, but since it is an algorithm; obviously, we can apply these algorithm in the best possible manner. So, that we have the best possible solutions. So, this is the essential idea of the genetic algorithm.

(Refer Slide Time: 05:11)



So, as you can see, inspired by natural evolution, gas involved direct manipulation of the coding achieved by cross over and mutation it begins the search from many points not from a single point. So, therefore, it contains population of feasible solutions to the problem, is it alright? So, supposing I have a solutions space suppose this is our solution space, right and in this solution space, there are many points. Now question is out of all those different points that are possible solutions, the gas randomly starts with points which are really covering you know the entire search space entire search space.

So, you see; we define a population; what is a population supposing my population size is 8, then 1, 2, 3, 4, 5, 6, 7, 8; these are the 8 particles or may be points which we have selected to define as our first generation. Now what we do; you know, I have out of these 8, I select may be a few may be say, 4, I select based on which of these 8 has the highest fitness what is the fitness; fitness is the objective function in that case the function x fx. So, supposing there is a function fx which we are maximizing. So, these are like x each individual point is an x. So, if I know x, I know fx, right. So, that is our what is known as the fitness function.

So, based on this fitness function value, you know I choose out of these 8 may be 4 of them, then by the process of cross over and mutation which I will discuss you know from the 4 parents, we obtain the 8 offspring, right, sometimes we do not really have 8 offspring, we may have let us say 6 offspring and the best two parents we also add may be. So, you see we can have something like this either 8 offspring or 6 offspring plus 2 parents who are the best in terms of fitness value that becomes our second generation.

So, you see from my first generation of 8 points to second generation of another 8 points; what is the fitness of these 8 and what is the fitness of these 8? If I compare, I will find the second generation you know the points, we will have a higher fitness value, why it happened because we have chosen out of the 8, 4 for reproduction which are of higher fitness value and the off springs fitness are very near to the parents not exactly same because they are not same, but then they are very near to them because it is. So, you know I have reason to believe that the average fitness has really gone up is it alright. So, that is the essential idea about genetic algorithm as I move from population to population may be generation to generation. So, from second generation, when I go to third generation to fourth generation to fifth generation, you know, I have a set of points with higher fitness, right.

Again looking at it, while we define the population, we should see that the population diversity; that means, representing all types of points and selection pressure; that means, the points are having higher fitness value they are maintained, is it alright. So, that is the essential idea of genetic algorithm.

The advantage it is an answer based method, they do not need auxiliary information like gradients search via sampling. So, no advance calculations and therefore, calculation is easy the iterative process is also simple, is it alright, the search by stochastic operations not by deterministic rules. So, it is by a random choice to guide highly exploitative search, it is alright that is the essential idea of genetic algorithm.

Let us give another example, supposing you know there is a large city and you really want to find persons of a certain type, is it alright; persons of a certain type see you know there are some people who are highly knowledgeable on a given area, but they are very rare you know, you really do not know how to get there is it alright and do you know in a city there are. So, many of them may be in a city of several crores people you have may be 20 of them, you need the best one out of that who has knowledge on the particular field.

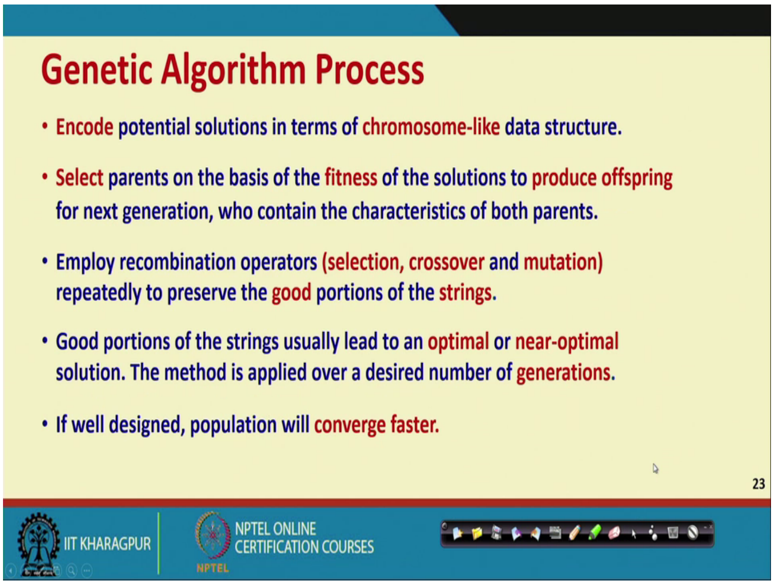
But then these 20 people; they are not together this twenty people are dispersed over the city in a random manner and how do you search them supposing you go to the city and you knock at every door that do you have this knowledge it is a foolish method, you just cannot really get that particular person by this method, is it all right?

But assume you hold a conference in that conference you randomly invite say 50 people, is it alright and these 50 people you tell that you go to several parts of the city and hold a another conference of 50 people, is it alright? So, you see from a 50 people a the next you know you are having another set of 50. So, 50 into 50; 2500 people are explored by second generation itself. So, that is the you know advantage of genetic algorithms, the search space is actually exploited in an exponential manner, is it alright, how it is done that that initial population of 50 that you choose randomly that is like holding the conference for the first generation.

But then out of these 50, you create through cross over and mutation if you know the next generation and the process is inherently such that it actually explores a much higher solutions space, is it alright. So, when you go from several generations you have actually exploited almost the entire search space. So, chances are if you are continuously evaluating the fitness function value you are going to reach at the optimum or at least near optimum the essential idea.

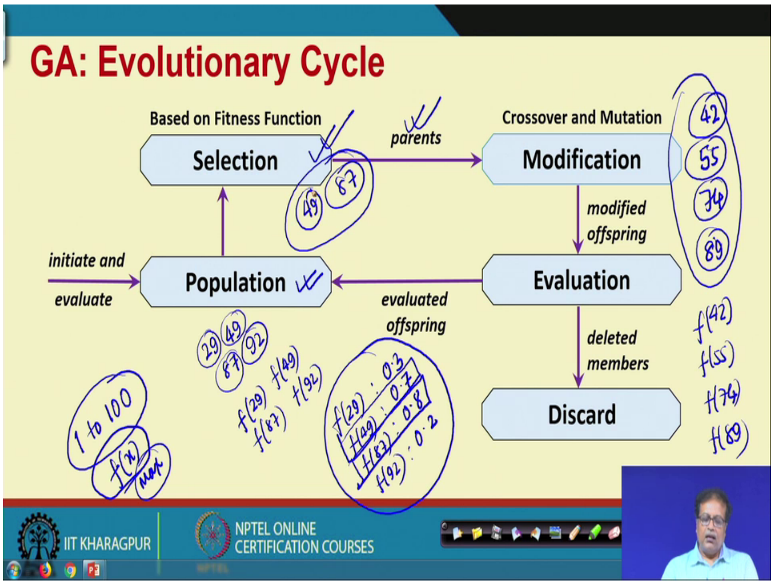
So, inspired by natural evolution, you know the gas are direction manipulation. So, already we have discussed this.

(Refer Slide Time: 13:41)



So, let us see our next slide that the idea of GA the process is such that we do an encoding, then select on the basis of fitness, then employ recombination operators that is selection cross over and mutation repeatedly to preserve the good portion of the string the string is the chromosome here the good portion of the strings or chromosomes usually lead to an optimal or near optimal solution the method is applied over a desired number of generations if well designed the population will converge faster.

(Refer Slide Time: 14:23)



So, let us look at the GA evolutionary process. So, you see, what really happen. We begin here that is we initiate and evaluate the population. So, initially we initiate the population, supposing you know, we are asked that find the number between 1 to 100, just take a very simple problem find the number between 1 to 2 you know which fits a functional value fx, right.

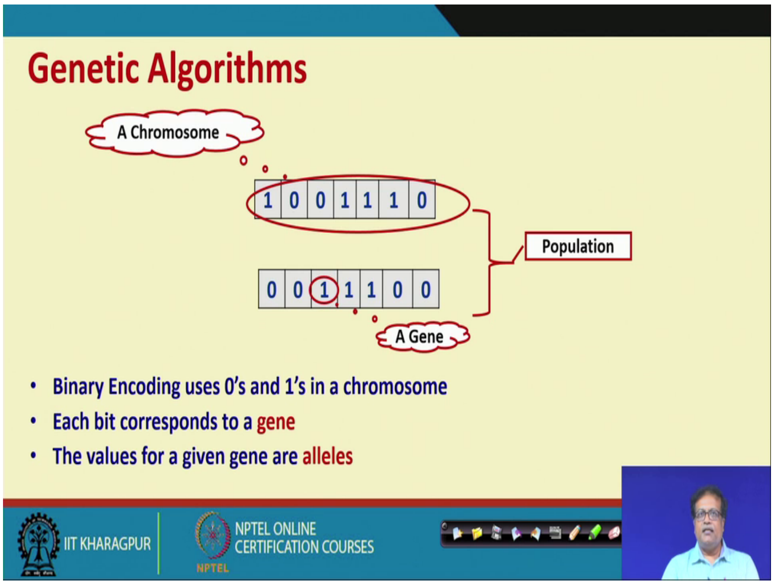
That means it returns a maximum value of this function fx. So, we randomly generate through random numbers may be some arbitrary number like 29, 49, 87, 92, arbitrarily, suppose we choose these 4 numbers and for each, I find out the functional value what are they; obviously, they will be f of 29, f of 49 and f of 92, supposing these are my functional values arbitrarily; I am writing the functional value returns these kind of numbers right suppose these are my functional values.

Now, look here this f 49 and f 87 are best possible parents. So, based on this fitness function, we select, right, we select out of these generation of 4, you know I select two of them which one. So, I select basically 49 and I select 87. So, this is where the selection is done they are my parents who are the parents the two which are selected, is it alright, then the process of cross over is really done. So, I will explain what is the process of cross over. So, through the process of cross over the 49 and 87 will become you know from these two, I can get 4 different numbers right suppose in those 4 different numbers are 41; say 42, 55, 78 and say 89. So, how these number are obtained you know, it will depend on how these are you know these chromosome structures are written.

So, these 4 numbers could be generated by the process of cross over and mutation of these two numbers. So, this is how the modification is done, then this is to be evaluated again that evaluated means again for each of them, I have to obtain the f 42, f 55, f 74 and f 89, right. So, after this evaluation, we might delete some members or I generate the next population. So, from this first population, I get the second population and then iterate till my fitness value is improved, right. So, that is the essential process of the genetic algorithm right.

So, we will discuss in more detail in our next class, but here I am just giving a overview.

(Refer Slide Time: 18:25)

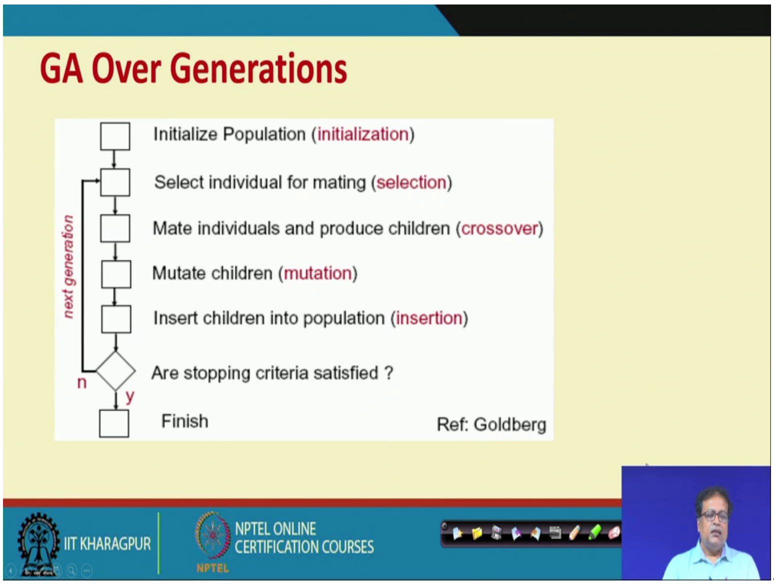


So, what is a chromosome? Now look here, a chromosome can be represented in many different ways one of the way is by binary encoding. So, what is binary encoding binary encoding is by writing the number by zeros and ones. Now it is not always that you know binary encoding simply means a number it could be something else also supposing in this example can be seen that supposing there are seven switches right the each switch could be 0 or 1.

Or assume you know in a knapsack problem where I have seven things, it is a 0-1 knapsack the knapsack basically is that there are seven things and a particular thing could be either there or not there each knapsack has a value right how to fill up the knapsack in the certain manner. So, that the value could be maximized and there would be some weight restriction also. So, that is how it is. So, now, here you see these are the seven things a particular chromosome could be 1 0 0 1 1 1 0; that means, first thing is there second and third not there fourth fifth and sixth is there and seventh is not there that is what the chromosome means in the context of that problem, I hope it is clear that a 0-1 knapsack, there are seven things a chromosome here means which ones are there which ones are not there

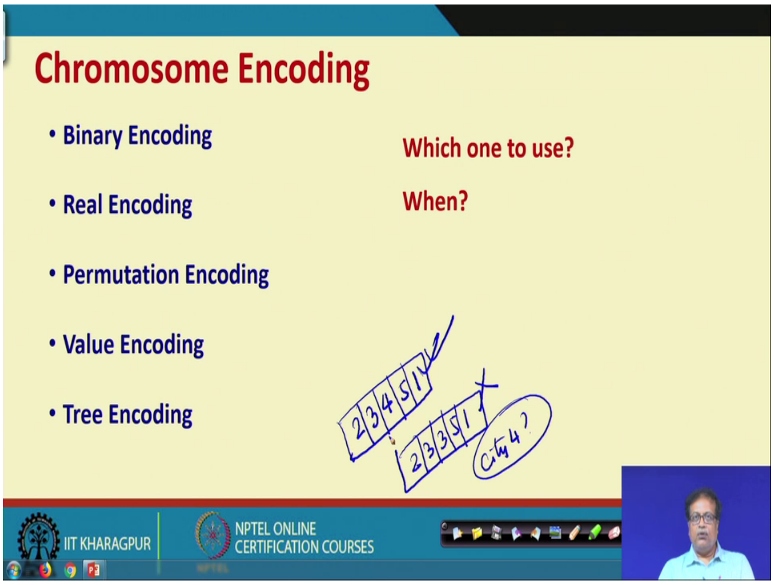
If it is a number, it can also represent a number, right. So, each bit corresponds to a gene and the value for a given genes are called alleles. So, what is an allele? A allele is the number that is the value, is it alright. So, that is the value is an allele and the chromosome is essentially you know the in case of binary encoding; that is the entire thing and population is a set of chromosomes right a set of chromosomes can be called a population. So, that is the essential idea of how these things are represented, right.

(Refer Slide Time: 21:13)



Now, having known this, now this is how the GA over the generations happens already we discussed that initialize selection cross over mutation insertion into the population and stop otherwise finish. So, that is how the Goldberg you know who has written a very important book on genetic algorithms gives the GA over generations, right.

(Refer Slide Time: 21:42)

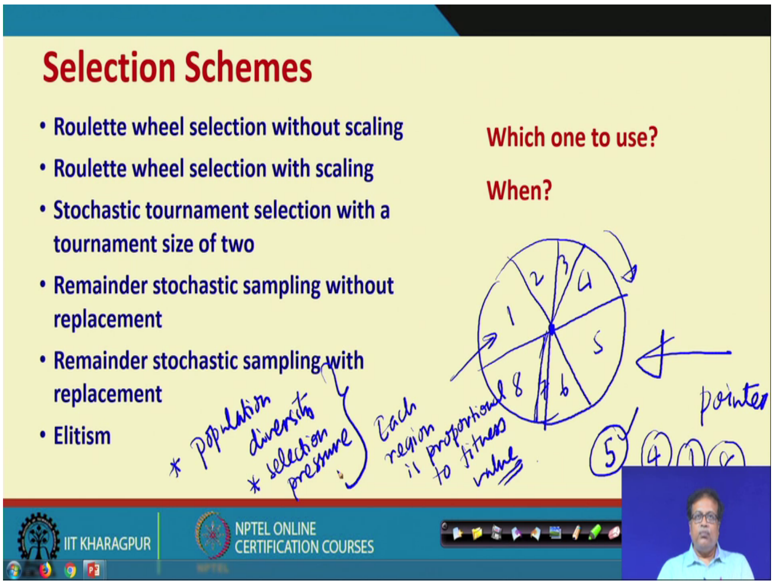


Now, encoding as I was saying, there are different kinds of encoding. What are they? It could be binary numbers, it could be real numbers, it could be a permutation encoding, see a example could be a travelling salesman problem supposing there are 5 cities. So, you have to travel to each city exactly once. So, here the encoding could be say 2, 3, 4, 5, 1; you see from 2 to 3 to 4 to 5 to 1; right that could be your encoding.

Now, you cannot really have 2 3 3 5 1, this is a mistake, why, it is a mistake because in case of a travelling salesman problem, you have to visit every city exactly once here where is city 4, right. So, this is not a good encoding whereas, this is a good permutation encoding right. So, in a permutation encoding every number should be exactly once which represents the order in which in a ts problem cities are visited then we could have value encoding where actual values could be used as genes in a chromosomes.

There are tree encoding also which we shall discuss later where the chromosome is a tree structure, they have got very unique uses alright the tree encoding. So, discussion I will have discussions later, right. So, that is how the different kind of encoding processes that are available for chromosome.

(Refer Slide Time: 23:31)



Now, question is; how do you select? The selection is also a very interesting process it could be roulette wheel selection with or without scaling then a stochastic tournament selection with a tournament size of two a remainder stochastic sampling with replacement remainder stochastic sampling with and without replacement and elitism is, alright.

So, you see what is really done the as I said that the initial population has got each member has got a fitness value a fitness function value. So, based on supposing I I make a roulette wheel, supposing, I make a roulette wheel in this manner, I make a wheel and supposing there are 8 members in a population. So, I divide this population into 8 regions. So, in this case; 1, 2, 3, 4, 5, 6, there could be seven and 8 also.

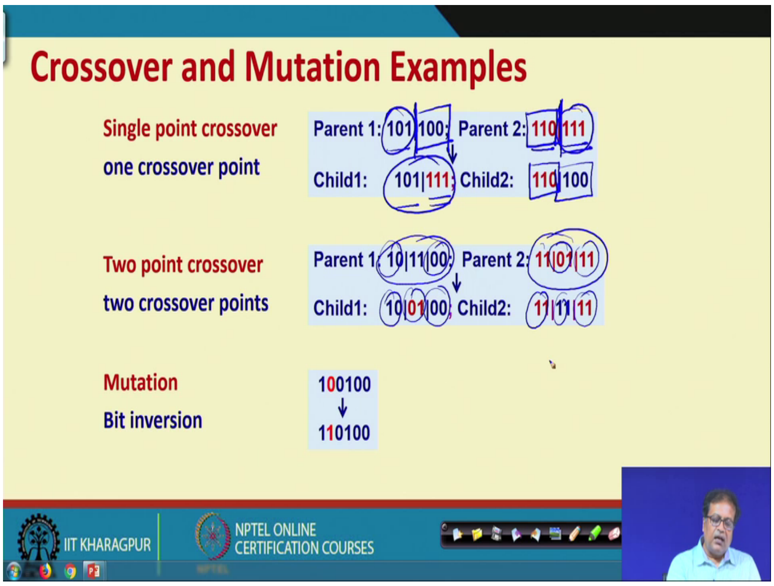
So, you see these are my 8 regions, as you can see those 8 regions are not equal the regions are suppose region 1, 2, 3, 4, 5, 6, 7, 8, each region is proportional to fitness value alright, it is actually proportional to the fitness value. So, chromosome one has got higher fitness. So, it is given higher region 3 or 7 has less fitness. So, they have been given less amount of region right. So, that is the roulette wheel.

Now, you rotate the roulette wheel and you know it just rotates and when it stops may be there is a pointer here that is a pointer. So, pointer will select one of them. So, suppose the pointer selects 5. So, 5 is chosen that is 50 is selected, right. So, so that is that is the essential idea of roulette wheel selection method. So, suppose you turn every time you turn may be you get 4, next time you get 1, next time and you get 8 next time. So, you see the chances of getting selected directly depends on the fitness. So, if you have higher fitness there is more chance of getting selected right.

Scaling is something where you actually modify the fitness values based on certain criteria is, alright, sometimes, we do select two them and then make a tournament the better 1, we select right remainder stochastic is basically divide the fitness value by average fitness and any value where the integer portion of numbers automatically are selected the roulette wheel is constituted based on the remainders, is it alright? An elitism is a very specific ones where the parent directly goes to the next generation not all parents some of the parents are directly selected in the next generation.

Now, every selection scheme has got pros and cons see, specifically, it is known that for genetic algorithm, it is important that I have a balance between population diversity with; that means that whether I represent points from the entire region and selection pressure; that means higher fitness value. So, balance has to be obtained between the population diversity and selection pressure that is very important for the genetic algorithms, right.

(Refer Slide Time: 28:09)



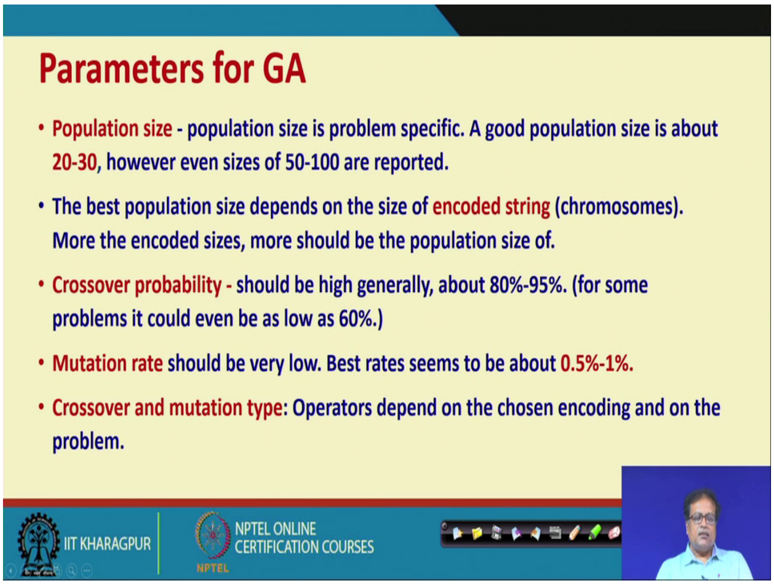
So, having said that let us move over and look at the cross over and mutation processes.

See cross over essentially look at, suppose, I have two parents 101 110, 001 111. So, there is a cross over point. So, these line is basically is a cross over point. So, these line is like a cross over point. So, what happens the, at the cross over points these hundred one is here and these hundred ten sorry these hundred eleven is here alright. So, these portion and these portion together they become a child and these portion and these portion because these should be first that is the first three and these are the last 3. So, these becomes another child right. So, that is how the crossover can be done these are called single point cross over, right.

And there is a two point cross over you see this is one gene sorry chromosome this is another chromosome. So, there are two cross over points. So, one is 10, then these 01 and then these 01 and then this 00, this is 1 and 11 these eleven these 11 and these eleven. So, this becomes another, right. So, this is how the cross over is made and from two parents we get two childs, is it alright. So, that is how we actually obtain the future generation from a set of parents.

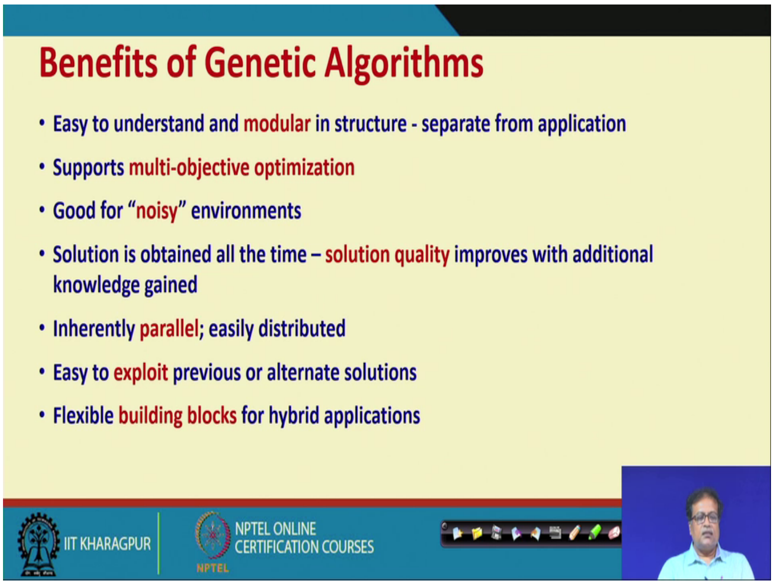
Sometimes a mutation is also done may be a bit is inverted a zero is converted to one as of to make mutation, is it alright.

(Refer Slide Time: 30:16)



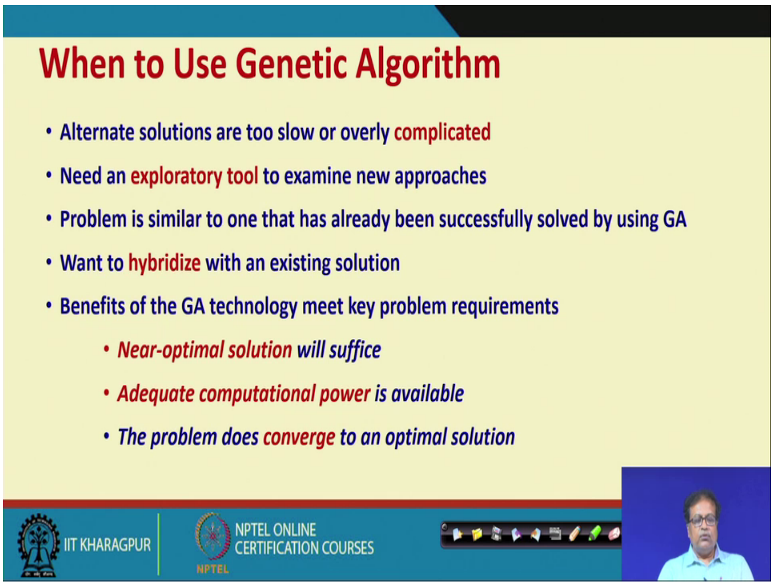
So, these are cross over and mutation examples, there are different parameters which we will also discuss in due time there is a size of the population how big should be taken 20, 30 or 50,000, its problem specific the cross over probability whether to do cross over or not a usually a high, probability is taken the low probability is taken for mutation rate and cross over and mutation type, there are large number of cross over the mutation types, which are again problem dependent, is it alright?

(Refer Slide Time: 30:53)



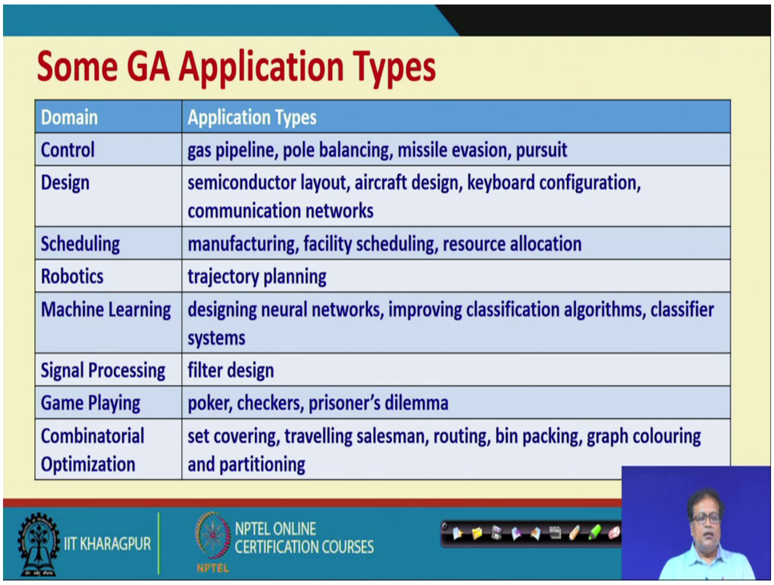
So, they are all used here are some benefits the benefits, you know, they are modular in structure good for noisy environment solutions are obtained all the time distributed parallel. So, parallel computing can be made use of easy to exploit flexible building blocks can be used in multi objective optimizations. So, those are the benefits, right.

(Refer Slide Time: 31:20)



And when to use a genetic algorithm when alternate solutions are slow need and exploratory tool problem is similar to one that has already been solved when we can hybridize with an existing salutations and where near optimal solution adequate computational power is available and problem converges that is when the benefits are maximum.

(Refer Slide Time: 31:49)



Finally here are some GA applications, right, they could be used in control design scheduling robotics machine learning signal processing game playing combinatorial optimization and many many other situations, see as I said the optimization problems are nowadays more and more dependent on nature inspired algorithms where GA is probably the pioneering one and many others like PSO SCO and many other algorithms they are essentially derived from the basic genetic algorithm process.

So, if you understand the basic genetic algorithm process, you can understand other variations also. Now, a since we are increasingly going to have large optimization problems, it is essential that we solve them and quickly solve them. So, that we get a solution if not the optimal at least the near optimal. So, it is always better than not solving at all, is it alright?

So, we shall more details about the GA process in our next lecture, right.

So, thanks for patiently hearing.