Genetic algorithm (GA) is an evolutionary algorithm because it makes use of the concept "survival of the fitess!" You will understand this as you go further.

Now we know that for different feature set we will get different accuracy. let's consider following cases where we have total of 5 features and 1 represents we have considered that feature for ML model while 0 means we haven't considered that feature.

	Feature Subset	Accuracy
1.	11101	0.89
2.	11001	0.91
3.	01110	0.95
4.	11111	0.61
5.	00010	0.98
6.	01010	0.945

Initial population — randomly generated. — initial pop size = 6.

Now what GA says that instead of trying all 2^5 combination of feature subset to see which gives best accuracy and hence select the best feature subset, we consider only those feature subsets initially whose accuracy is good and try to develop new feature subset using those good feature subset which has good accuracy.

This logic works seemlessly effective because this is how nature works. When you mate two humans (male and female), qualities of them can be seen in their offspring. Similarly if the parents have good qualities then the offspring is a combination of good ~~and bad~~ qualities of both the parents.

Hence you can develop an analogy that the feature subset with good accuracy are parents and the child that is ~~both~~ born due to them is the new feature subset.
Mating of two parents is actually crossover of their genes, hence here we use the term "crossover of features" instead of mating.

Now we have all seen cases wherein child has some qualities which are not there in either of the parent. This is called mutation of genes. To try to copy same thing here in our algo and perform mutation on some of the children.

These children comprises of next generation and there is crossover b/w these members leading to next generation and so on. Crossover in each generation follows some rules.

Hence this algorithm is called genetic algorithm.

Now you need to put a threshold on the accuracy which defined good feature subset. Once putting a threshold is done, the way we select the feature subset that satisfies the threshold is called selection. There are many selection techniques like ranking, tournament and roulettle wheel. For eg:- in ranking method say we put a threshold of accuracy ≥ 0.90 then the feature subset is sorted in descending order and we keep selecting feature subset from top that satisfies the condition. Hence (00010) (01110), (01010) and (11001) are selected. This is called first generation.

There are several ways of doing crossover like single pt crossover, multi pt crossover, uniform crossover ..... you can develop you own crossover methods. For example single pt crossover about 2 between (00010) and (01110) is given by

00010 → 00110
01110 → 01010

Similarly you can do single pt crossover about any pt.

Likewise multiple crossover about 1 and 4 b/w (00010) and (01110) gives

0 | 0 0 1 | 0	→	0 1 1 1 0
0 | 1 1 1 | 0		0 0 0 1 0

Always remember that multiple pt crossover means about only 2 pts and not more than that, name is deceptive.

To control the crossovers i.e. no. of children in next generation we need to define a crossover rate. This is decided based on hit and trial method and is usually consider high like 0.8, or 0.9.

There are several mutation techniques like bit swap and bit flip. Mutation is also called variation.

Performing bit flip on child
(00110) → (11001)

Performing bit swap on child
(00110) using mask (11011)

(00110) → (11101)

To control the mutation i.e. no. of children that get mutated we need to define a mutation rate. This value is decided based on hit and trial method and is usually considered low like 0.1 or 0.2.

The stopping criteria of this algorithm is :- when we achieve a maximum accuracy threshold i.e. we started with 0.95, 0.945 and 0.98 and as we progress through generation it we get say 0.998 then we stop. Another can be that we get same accuracy over multiple generations means no better feature subset is available, or you can specify the no. of generations to stop.

Heuristic here is accuracy i.e. when GA is used for feature selection, we use accuracy as heuristic. Heuristic is also called fitness function/score.

- Hence operators of genetic algo are :- selection, crossover and mutation.

- Hence parameters of genetic algo are :- crossover rate, mutation rate and initial population size. All decided based on hit and trial approach.
- Other famous crossover techniques are partially matched crossover (PMX) and cyclic crossover.