Report
MDL Project

Team 84
Harshit Sharma (2019101083)
Aaditya Sharma (2019113009)

1. Summary of Genetic Algorithm

Genetic Algorithm is based on few steps. First one is setting up Initial Population, second is selection of the population on the basis of fitness, third one is crossover and last one is mutation. This keeps on repeating itself till we find our proper and suitable population set. In the genetic algorithm used in this project , we just take input of a parent vector from a file, and for the initial population we just render it 20 times in an initial population array. Then the fitness of every 20 parent is checked and stored. The parents are sorted according to the fitness, the fittest parent is added first. If the current parent array has some ancestors, then their fitness is compared with fitness of their parent and best 20 parents are kept for further mating. This is called **Selection** or **Mating Pool**. Now that we have the mating pool , we take adjacent vectors and perform crossover on them. A function such as the child has characteristics or features of both of the parents. This refeers to **Crossover** where the child has features or genes of both the parents. With generations and births/lives etc. changes come in the man. We are not the same as we were like a century or two ago, we change accroding to surroundings and other factors, these changes that come with generations are called **Mutations**. As the child is born, a mutate probability is assigned to the child, which characterises its mutation type and amount. This selection,crossover and mutation keeps on going till we find our favourable generation. Our main.py works for 10 generation when the program is implemented once.

3. Fitness Function

The fitness function is very important for the selection or generating the mating pool for cross over. It tells us how fit is a parent for crossover and is a good factor of comparison for fittest parents. We just made a fitness function as a linear function of train error and validation error of a parent vector.
We have used train_factor and validation_factor as the coefficients of train error and validation error respectively. We tried a number of combinations of them. At first we kept both of them at 1 and 1 but it got stuck at minimum at a given error. Then we tried 0.3 and 0.5 , 0.7 and 0.8 etc. But finally fixed train_factor as 0.7 and validation_factor as 1.
Now as fitness is defined as 0.7*train_error + 1*validation_error for a given parent vector. So, lesser this value more would be the fitness of the parent.

4. Crossover Function

Crossover is a genetic recombination to combine genetic information of two parents to make new offspring. The offspring has genetic information of both the parents.In our genetic algo crossover is implemented as follows :
In the 20 parents sorted according to their fitness, parents are selected adjacently to make 20 new offsprings. First 6 genes of first parent and last 5 genes of second parent are taken , this crossover gives a new child which have genetic information of both the parents. Thus 20 new offsprings are made using the parents.

5. Mutations

As a child is created by the crossover of two parents its given a mutation probability which is a random integer from 0 to 10. Now if a child has mutation probability less than 3, it would not undergo mutation. If the mutation probability is less than 7 and greater than equal to 3, the child would undergo first type mutation. If the mutation probability is greater than equal to 7, the child would under second type mutation.
First type of mutation :
A random gene is selected from first 6 and a random gene is selected from last 5 and they are swapped.

Second Type of mutation :
A random gene is selected and its multiplied by a random float between 0.3 and 1.6.

6. HyperParameters

Pool size also determines the diversity of the genetic algorithm and helps in analyzing the good working of the algorithm. Initially when we coded the algorithm , we kept the pool size or size of the mating pool as 10 but later changed it to 20. So, at a time 20 fittest parents cross overed to make 20 offsprings. At the crossover, we took first 6 genes of first parent and last 5 genes of the second parent for making the new offspring, it was totally random. We also tried some generations with vice-versa situation but nothing significant changes were observed , so the plan was dropped. In th selection process, the fitness of ancestors and parents were checked and the 20 most fittest among them were considered as mating pool. 10 Generations are formed using the program once for a given parent vector.

7. Statistical Information

We viewed that some vectors we used gave really high error initially but even after some iterations , the errors did not reduced.
On some good vectors which have 7.6e12 error it got almost 120-125 generations to bring down the error to 5.5e12.
Sometimes after 100-110 generations on a vector , the error is stuck in the minimum and we need to change the vector by checking the error .

8. Heuristics that did not work

We tried to make new vector using swapping the genes of the some generations trained vector but there was no major changes error used to either decrease or increase by just a small factor. There were no major changes.
We also tried to use 0 as the parent vector but it used to get stuck in minimum , so it was also not a good idea either.
Also some changes with mutation were required,
Our initial mutation were swapping two random genes and swapping a stream of genes in a child vector, but it was observed that even after some generations it also gave repititive errors.
Also the mutation prob was changed such that 70% chances of mutation are possible for a certain child made from crossover of two parents.

9. Train and Validation error achieved

We started from an error set of [1.1758916277716e+16, 1.1910931502171786e+16] after mutation and regular training the error set achieved is [1911202442105.4507, 1773594703135.9011] for [7.45912001e-12 ,-2.39320228e-07 , 1.54009038e-12 , 1.09381114e-06 ,-2.49609150e-14 , 8.74441901e-17 ,-4.79219068e-18 ,-1.85823970e-13 ,-5.44897788e-13, 0.00000000e+00, -9.50795943e-11] parent vector.

10. Tricks

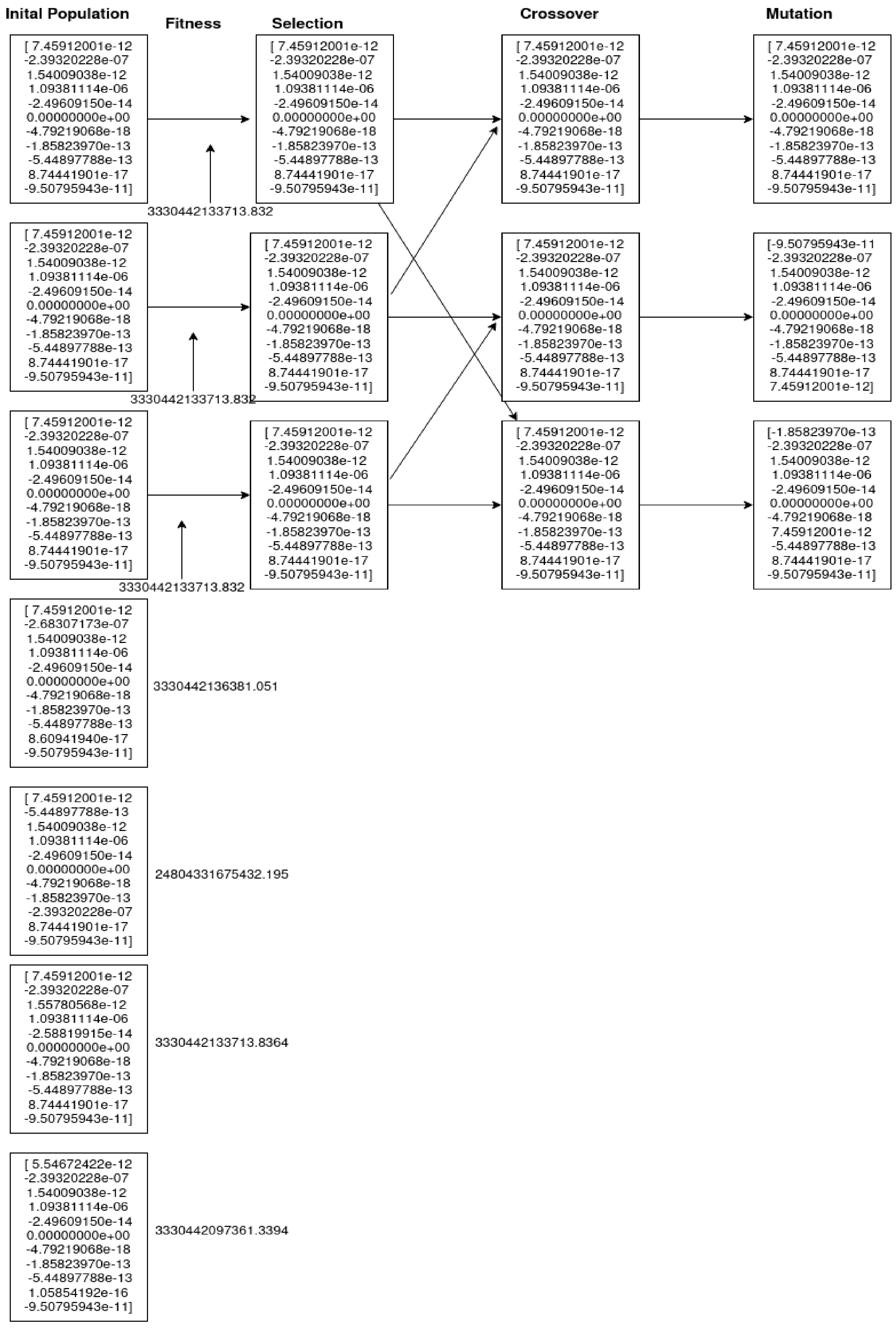
After the first interim submission, we need something to generate new vectors and even check how our mutations could be made better and more useful. So, we created a program called mutate.py. It took our best vector as an input and mutate it according to our mutation code , POPULATION_SIZE times and its error was measured. It was a good way to check different types of mutations and how much effect they are and also a better way to generate nice parent vectors.

Diagrams

1. Diagram 1 is of 1st Generation

2. Diagram 2 is of 2nd Generation

3. Diagram 3 is of 3rd Generation



Inital Population

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```
3330442133713.832

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```
3330442133713.832

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```
3330442133713.832

```
[ 7.45912001e-12
-2.68307173e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.60941940e-17
-9.50795943e-11]
```
3330442136381.051

```
[ 7.45912001e-12
-5.44897788e-13
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -2.39320228e-07
 8.74441901e-17
-9.50795943e-11]
```
24804331675432.195

```
[ 7.45912001e-12
-2.39320228e-07
 1.55780568e-12
 1.09381114e-06
 -2.58819915e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```
3330442133713.8364

```
[ 5.54672422e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 1.05854192e-16
-9.50795943e-11]
```
3330442097361.3394

Selection

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

Crossover

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

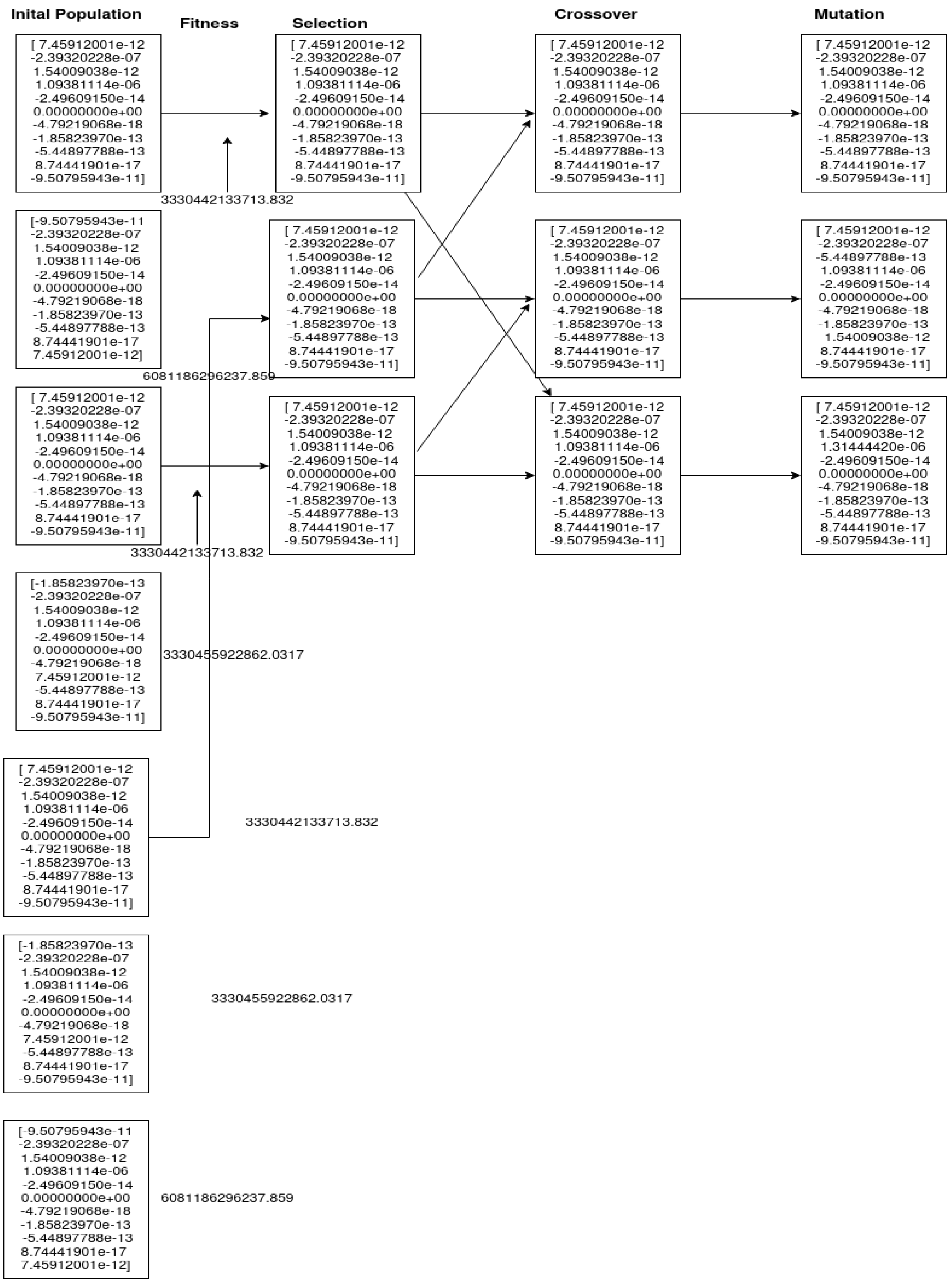
```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

Mutation

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

```
[-9.50795943e-11
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
 7.45912001e-12]
```

```
[-1.85823970e-13
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 7.45912001e-12
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```



Inital Population

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

```
[-9.50795943e-11
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
 7.45912001e-12]
```

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

```
[-1.85823970e-13
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
-4.79219068e-18
 7.45912001e-12
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

```
[-1.85823970e-13
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
-4.79219068e-18
 7.45912001e-12
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

```
[-9.50795943e-11
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
 7.45912001e-12]
```

Selection

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

Crossover

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

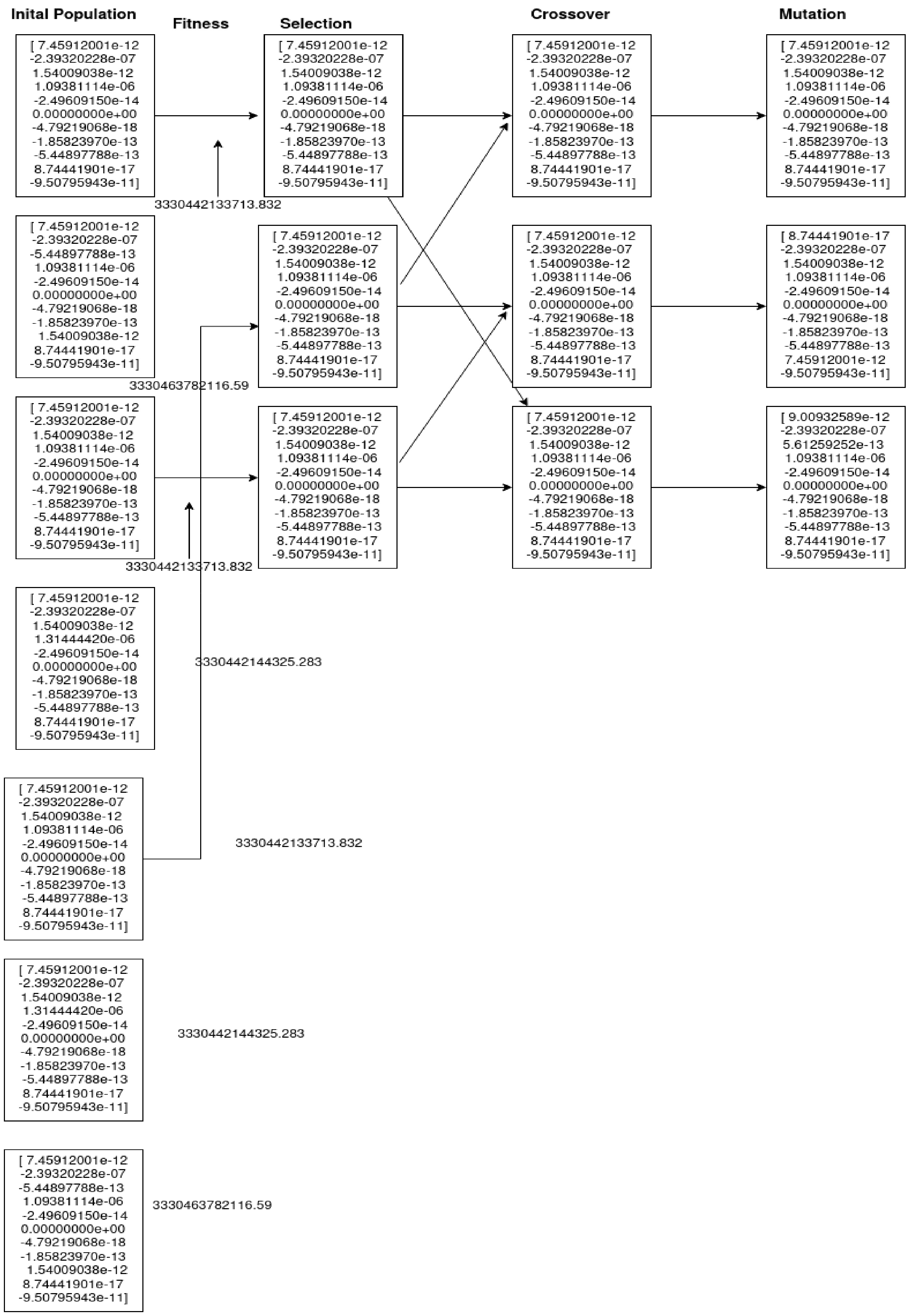
```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

Mutation

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

```
[ 7.45912001e-12
-2.39320228e-07
-5.44897788e-13
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
  1.54009038e-12
 8.74441901e-17
-9.50795943e-11]
```

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.31444420e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```



Inital Population

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```
3330442133713.832

```
[ 7.45912001e-12
-2.39320228e-07
-5.44897788e-13
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
  1.54009038e-12
 8.74441901e-17
-9.50795943e-11]
```
3330463782116.59

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```
3330442133713.832

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.31444420e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```
3330442144325.283

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```
3330442133713.832

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.31444420e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```
3330442144325.283

```
[ 7.45912001e-12
-2.39320228e-07
-5.44897788e-13
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
  1.54009038e-12
 8.74441901e-17
-9.50795943e-11]
```
3330463782116.59

Selection

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

Crossover

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

Mutation

```
[ 7.45912001e-12
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```

```
[ 8.74441901e-17
-2.39320228e-07
 1.54009038e-12
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 7.45912001e-12
-9.50795943e-11]
```

```
[ 9.00932589e-12
-2.39320228e-07
 5.61259252e-13
 1.09381114e-06
 -2.49609150e-14
 0.00000000e+00
 -4.79219068e-18
 -1.85823970e-13
 -5.44897788e-13
 8.74441901e-17
-9.50795943e-11]
```