CS 451 ASSIGNMENT 2 REPORT

Orhun Güley S004428

Introduction

I implemented and ran genetic algorithm to maximize the prediction for a given input in CS 451 Assignment 2. Genetic algorithm is a type of evolutionary algorithm which is inspired by the natural selection. Genetic algorithms are commonly used in optimization problems, in a wide range of areas. Since I am majoring industrial engineering, I was familiar with the algorithm before the assignment was given. Generally, the main objective in Genetic Algorithms is to seek for better solutions in the entire space from multiple branches. It transfers the best solutions found to the next step and creates new solution candidates from them to be tested.

Algorithm

Initial Population: Initial population is consist of 200 weight vectors in my experiment. Since there are 12 weight values in each person in the population, that corresponds to a 200 x 12 matrix.

Fitness Function: Fitness function was given in the assignment information which can be seen below;

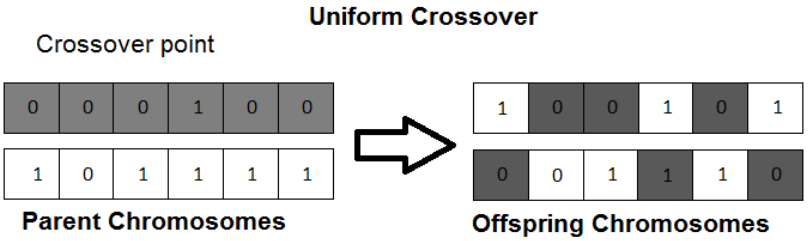
$$\sum_{i}^{\#Row} \sqrt{(y^i - y^i{}_{pred})^2}$$

Selection of Parents: Since the fitness function is basically corresponds to sum of squared error, the higher value implies a bad candidate in the population. As you can understand from the previous sentence, to calculate the selection probabilities of each person in the population, I had to change the fitness value by making a little trick. You can see the calculation below;

Probability of selection: $\frac{1}{\text{SSE}}$

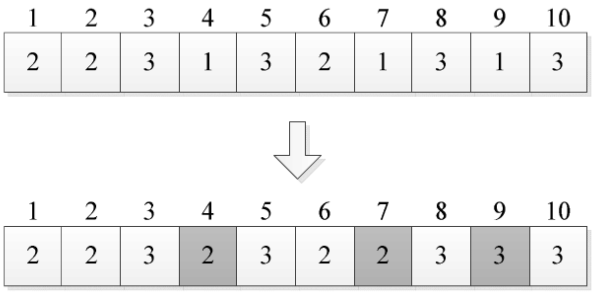
After the calculation of the probabilities for each individual in the population, I create a cumulative distribution array so that when I create a random number, I will be able to find out which individual is selected as a parent.

Crossover: There are different kinds of crossover types. In this project, I prefer uniform crossover. In uniform crossover, weights are randomly copied from the first or from the second parent. You can see the illustration below;



Mutation: For mutation operation, I use 2 different types of mutation.

1. Pick 2 weights randomly. Then with 0.5 probability, increase or decrease the weight by 0.1 ration. For example, if a weight is 1, with 0.5 probability, it will mutate as 1.1. Else, it will mutate as 0.9.
2. Pick 1 weight randomly and change it completely by replacing by a randomly created weight.



Results

Since my computer is capable of making large number of iterations, I set the population to 200 and X vector to a 100x3 matrix. I restrict the iterations with 100000. But there is a stopping condition. If the error of the fittest individual in the population has a lower sum of squared error than 0.05 the loop will be ended. I did bunch of experiments. In some of them, the algorithm was able to find an individual with a less error value than 0.05. However, there were some cases that the error value couldn't find an individual with a less error than 0.05. In almost every runs, the algorithm was able to find at least one individual who has less error value than 0.1. In my opinion, genetic algorithm is very successfully in that kind of optimization problems.

While playing with the parameters and trying to minimize the sum of squared errors, I found out that after some point, all the individuals became the same because of the nature of the genetic algorithm. In

that kind of situations, making crossover operations are pointless since every gene(weight values) are the same for the parents. In that kind of situations, I write an if statement that, if the parents have exactly the same fitness value, instead of making a crossover operation, make mutation operation to increase the variety. Actually, that speed up the process in good way.

I am going to share one run of my experiment results below:

Weights of the fittest in iteration 0 is :
[-0.38735912 0.28404594 0.43822221 -0.46684385 -0.84157052 -0.93680508
-0.84288538 -0.60419459 0.77178694 0.37718094 -0.04579028 -0.35182799]

Error of the fittest in iteration 1000 is : 1.2678009614989656

Weights of the fittest in iteration 1000 is :
[-0.24352821 0.82171002 -3.22117929 0.57915529 0.61424029 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 2000 is : 0.59761516447269

Weights of the fittest in iteration 2000 is :
[1.43153388 0.81349292 -3.22117929 0.63707082 0.61424029 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 3000 is : 0.500676161376158

Weights of the fittest in iteration 3000 is :
[1.43153388 0.81349292 -3.22117929 0.63070011 0.61424029 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 4000 is : 0.473199915590925

Weights of the fittest in iteration 4000 is :
[1.41721854 0.81253529 -3.22117929 0.63070011 0.61424029 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 5000 is : 0.473199915590925

Weights of the fittest in iteration 5000 is :
[1.41721854 0.81253529 -3.22117929 0.63070011 0.61424029 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 6000 is : 0.473199915590925

Weights of the fittest in iteration 6000 is :
[1.41721854 0.81253529 -3.22117929 0.63070011 0.61424029 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 7000 is : 0.32946188778274965

Weights of the fittest in iteration 7000 is :
[1.41721854 0.73128176 -3.22117929 0.63070011 0.61424029 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 8000 is : 0.18697022417648912

Weights of the fittest in iteration 8000 is :
[1.27549669 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 9000 is : 0.18697022417648912

Weights of the fittest in iteration 9000 is :
[1.27549669 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 10000 is : 0.18697022417648912

Weights of the fittest in iteration 10000 is :
[1.27549669 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 11000 is : 0.18697022417648912

Weights of the fittest in iteration 11000 is :
[1.27549669 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 12000 is : 0.18697022417648912

Weights of the fittest in iteration 12000 is :
[1.27549669 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 13000 is : 0.18697022417648912

Weights of the fittest in iteration 13000 is :
[1.27549669 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 14000 is : 0.18697022417648912

Weights of the fittest in iteration 14000 is :
[1.27549669 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 15000 is : 0.18697022417648912

Weights of the fittest in iteration 15000 is :
[1.27549669 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 16000 is : 0.18697022417648912

Weights of the fittest in iteration 16000 is :
[1.27549669 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 17000 is : 0.14616626137817534

Weights of the fittest in iteration 17000 is :
[1.28784244 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 18000 is : 0.14616626137817534

Weights of the fittest in iteration 18000 is :
[1.28784244 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 19000 is : 0.14616626137817534

Weights of the fittest in iteration 19000 is :
[1.28784244 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 20000 is : 0.14616626137817534

Weights of the fittest in iteration 20000 is :
[1.28784244 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 21000 is : 0.14616626137817534

Weights of the fittest in iteration 21000 is :
[1.28784244 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315

0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 22000 is : 0.14616626137817534

Weights of the fittest in iteration 22000 is :
[1.28784244 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 23000 is : 0.06982859676892317

Weights of the fittest in iteration 23000 is :
[1.31920835 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 24000 is : 0.06982859676892317

Weights of the fittest in iteration 24000 is :
[1.31920835 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 25000 is : 0.06982859676892317

Weights of the fittest in iteration 25000 is :
[1.31920835 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 26000 is : 0.06982859676892317

Weights of the fittest in iteration 26000 is :
[1.31920835 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 27000 is : 0.06982859676892317

Weights of the fittest in iteration 27000 is :
[1.31920835 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 28000 is : 0.06982859676892317

Weights of the fittest in iteration 28000 is :
[1.31920835 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 29000 is : 0.06982859676892317

Weights of the fittest in iteration 29000 is :
[1.31920835 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 30000 is : 0.06982859676892317

Weights of the fittest in iteration 30000 is :
[1.31920835 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 31000 is : 0.06982859676892317

Weights of the fittest in iteration 31000 is :
[1.31920835 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 32000 is : 0.06982859676892317

Weights of the fittest in iteration 32000 is :
[1.31920835 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 33000 is : 0.06982859676892317

Weights of the fittest in iteration 33000 is :
[1.31920835 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 34000 is : 0.06982859676892317

Weights of the fittest in iteration 34000 is :
[1.31920835 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 35000 is : 0.06982859676892317

Weights of the fittest in iteration 35000 is :
[1.31920835 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 36000 is : 0.06982859676892317

Weights of the fittest in iteration 36000 is :
[1.31920835 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 37000 is : 0.06982859676892317

Weights of the fittest in iteration 37000 is :
[1.31920835 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 38000 is : 0.06982859676892317

Weights of the fittest in iteration 38000 is :
[1.31920835 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 39000 is : 0.06982859676892317

Weights of the fittest in iteration 39000 is :
[1.31920835 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 40000 is : 0.06982859676892317

Weights of the fittest in iteration 40000 is :
[1.31920835 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 41000 is : 0.06982859676892317

Weights of the fittest in iteration 41000 is :
[1.31920835 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 42000 is : 0.06982859676892317

Weights of the fittest in iteration 42000 is :
[1.31920835 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 43000 is : 0.06982859676892317

Weights of the fittest in iteration 43000 is :
[1.31920835 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 44000 is : 0.06982859676892317

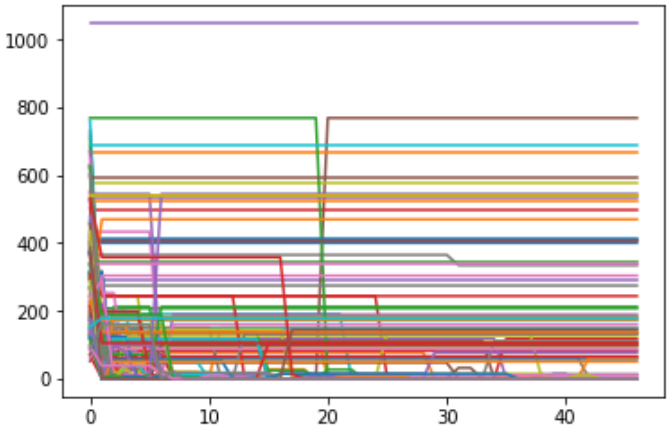
Weights of the fittest in iteration 44000 is :
[1.31920835 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 45000 is : 0.06982859676892317

Weights of the fittest in iteration 45000 is :
[1.31920835 0.73128176 -3.22117929 0.63070011 0.6148887 0.50855315
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]

Error of the fittest in iteration 45176 is : 0.02932523646892915

Weights of the fittest in iteration 45176 is :
[1.31920835 0.73128176 -1.34225265 0.63070011 0.6148887 0.55940846
0.86541087 0.2451829 -0.19076337 -0.20219911 -0.68963647 0.51703572]



As you can see above, at the case I tried, algorithm was able to find an individual which has a SSE less than 0.05 in 45176 iterations. I did not share the error matrix because the shape of the matrix will be 200x1 and it would occupy too much space in the document. As you can see from the graph, most of the individuals in the population get fitter in proportionally with iterations. The purple individual at the top was not able to get fitter. My comment on that one is that the SSE of the individual is so high that it wasn't selected as a parent.

Overall, assignment 2 was very informative and entertaining project to do. I had an experience with genetic algorithms in my senior design project last semester. The knowledge I gain in the senior design project help me in this assignment. I am very happy to have that kind of practical assignments in this course. It helps me to digest the topics covered in the lectures more.