Sujay Jakka

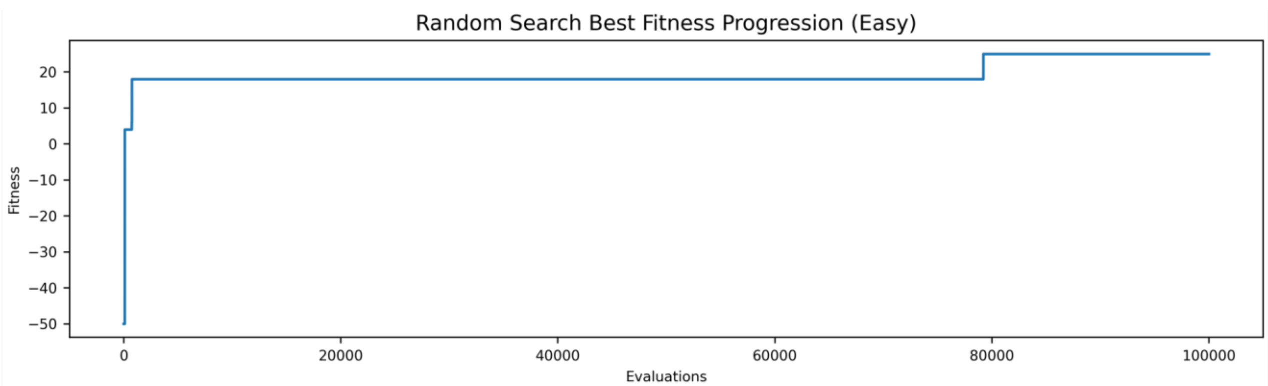
Svj0007@auburn.edu

COMP 5660 Fall 2024 Assignment 1a

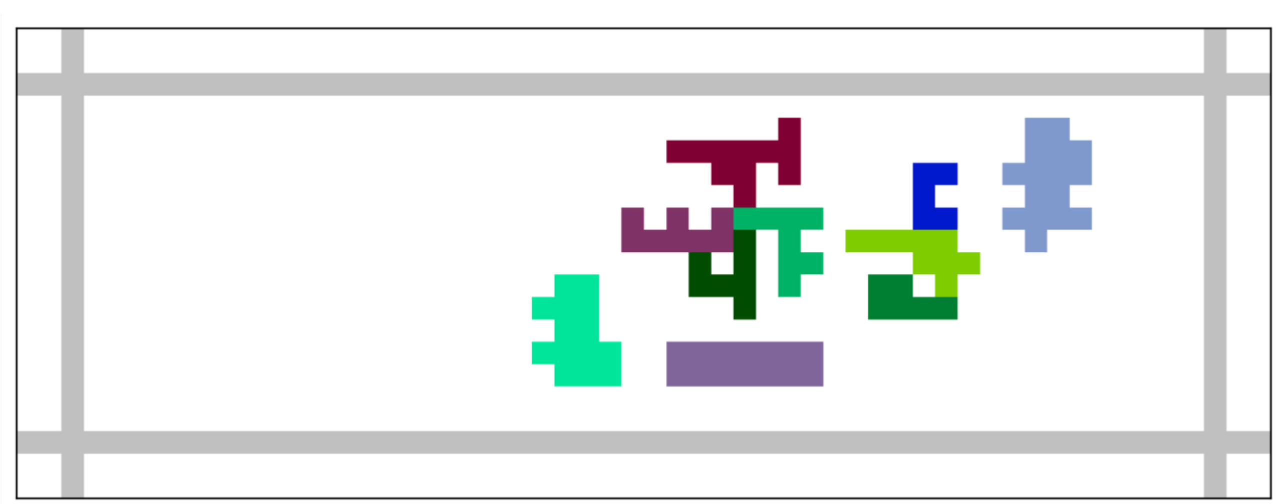
8 September 2024

Assignment 1a Report

I will first introduce the results for the easy problem. For the easy problem, the solution with the best fitness I have found using random search across 30 runs was 25. The images below are of the stairstep plot of the run that produced the best fitness value and the visualization of the best solution discovered across 30 runs respectively. As you can see from the stairstep plot we encountered our best solution around approximately 80,000 evaluations of random search.

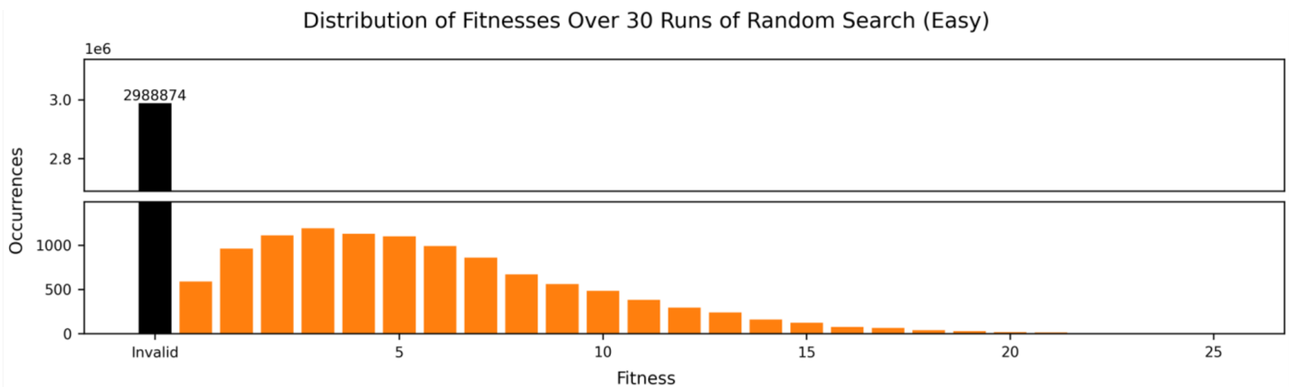


(Stairstep Plot of the Best Run)



(Visualization of the Best Solution across 30 Runs)

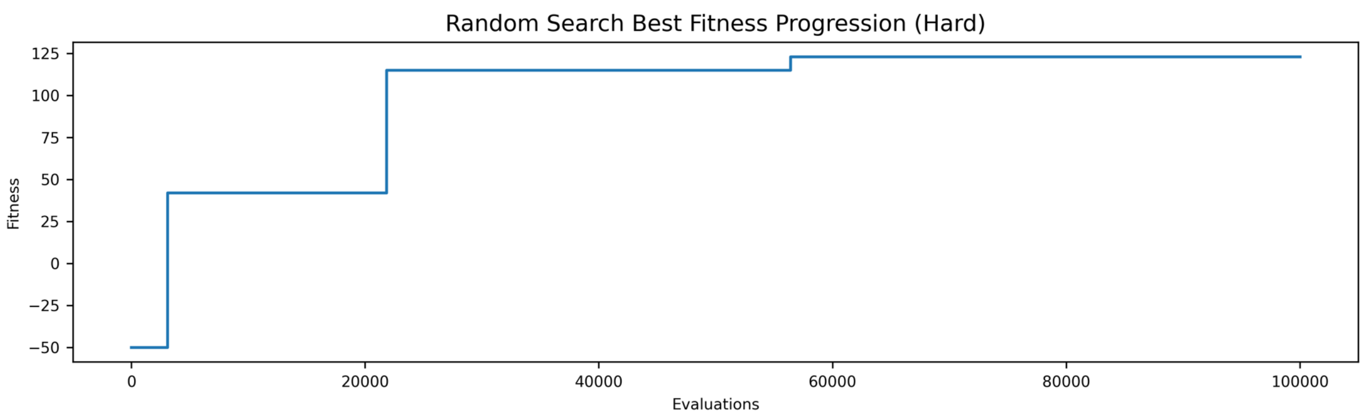
Using the histogram produced over 30 runs, I was able to calculate the proportion of valid solutions which was 0.0037087. From the histogram, I was also able to take note of the distribution of our valid solutions. The distribution of valid solutions seems to be unimodal and skewed right. Furthermore, it seems that most of the valid solutions seem to be concentrated around the 1 to 6 fitness range. The image below is the histogram of evaluations of randomly generated solutions across 30 runs.



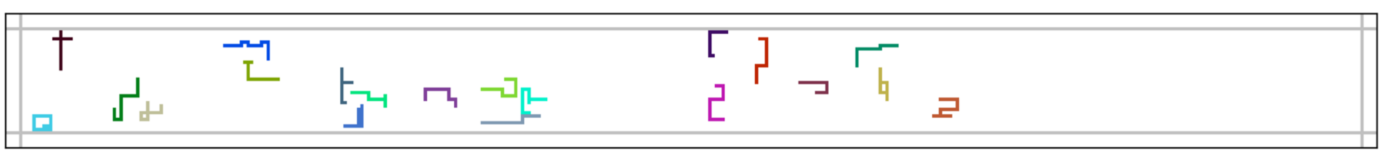
(Histogram of Evaluations across 30 Runs)

After conducting the experiment for the easy problem, I performed a statistical analysis between my sample of the best fitness value found from each run using random search compared to another sample of the best fitness value found from each run using a mystery algorithm. Both samples had a size of 30. The means of the random search sample and the mystery algorithm sample are 21.1 and 37.47 respectively. Furthermore, the standard deviation of the random search sample and mystery algorithm sample are approximately 1.75 and 1.01 respectively. I performed an independent two-sample t-test between the samples. I set my alpha value(significance level) to 0.05 and my degrees of freedom to 30, which is same value of the sample size. The p-value I received after running the t-test is 1.220e^-39 which is a lot smaller than my alpha value. This means that under the conditions of the null hypothesis(both samples have the same true mean) the probability is 1.22e^-37% of finding observations at least as extreme as our current observation. Because p is a lot less than alpha, the two algorithms(random search and mystery algorithm) have statistically significant differences in performance. Because it is highly unlikely that both sample distributions have the same population mean and that the mystery algorithm sample has a higher mean than the random search sample, we can conclude that the mystery algorithm performed better.

Now for the hard problem instance, my overall best fitness value I received across 30 runs is 123. Similar to the easy problem instance, the images below are of the stairstep plot of the run that produced the best fitness value and the visualization of the best solution discovered across 30 runs respectively. According to the stairstep plot, it seems that the random search algorithm encountered the best fitness value at approximately 60,000 evaluations in the run that produced the best fitness value.

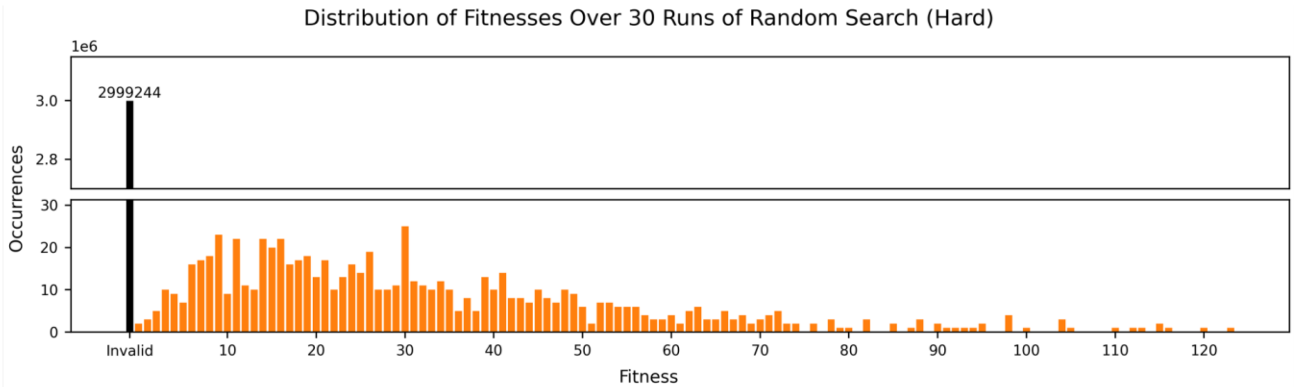


(Stairstep Plot of the Best Run)



(Visualization of the Best Solution across 30 Runs)

From the histogram produced across 30 runs, I was able to calculate the proportion of valid solutions which was 0.000252. The distribution of our fitness values across these runs seem to be multimodal.



(Histogram of Evaluations across 30 Runs)

Similar to the easy problem, I performed a statistical analysis between two samples. One sample consisted of the best fitness value for each run using random search while the other sample consisted of the best fitness value for each run using the mystery algorithm. Both samples had a sample size of 30. For the random search sample, my mean was 89.3 and my standard deviation was 20.521. Furthermore, for the mystery algorithm my mean was 334.93 and my standard deviation was 8.325. I performed an independent two-sample t-test. The goal was to determine if the two sample distributions have the same population mean(true mean). The null hypothesis is that both samples, the random search sample and mystery algorithm sample, are equally effective at solving this problem instance. I performed the t-test with an alpha set at 0.05 and the degrees of freedom set to 30. I received a p-value of 1.09e^-39 which is significantly smaller than my significance level of 0.05. This allows us to reject the null hypothesis and claim that the two algorithms have statistically significant differences in performance. Lastly, because it is highly unlikely that both sample distributions have the same population mean and that the mystery algorithm sample has a higher mean than the random search sample, we can conclude that the mystery algorithm performed better.