CS5401 **Tanner Wendland**
Assignment 2c 11/28/18
trwt33@mst.edu

Methodology

For this particular EA, the author used two different GP tree definitions. One used to represent the pacman controllers and one to represent the ghost controllers. Both of these trees are binary trees, and have the same function set.

The function set for both of them are addition(a, b), multiplication(a, b), subtraction(a, b), division(a, b), and rand(a, b) where if b is zero for division the result is one, and rand(a, b) chooses a floating point number uniformly in the segment (a, b).

The pacman GP tree has 5 terminals that are all floating point numbers. These sensor terminals are the L1 norm between pacman and the closest ghost, the L1 norm between pacman and the fruit, the L1 norm between pacman and the closed pill, the number walls directly adjacent pacman, and a constant floating point number. The constant floating point number is randomly chosen between 1 and 20.

The ghost GP tree has 3 terminals that are all floating point numbers. The sensor terminals are the L1 norm between the ghost and pacman, the L1 norm between a ghost and the nearest ghost that is not itself, and the L1 norm between the ghost and the fruit.

When pacman and the ghost are evaluating their trees to decide the direction to take, they will pick the direction that produces the minimum value. Because some controllers will produce multiple values that are the minimum, both pacman and the ghost will randomly choose directions if the minimum value is not unique.

The controllers are initialized by ramded half-and-half where each controllers has a 50% chance for being initialized by the full or grow method. In the implementation of both full and grow, the chance or getting a particular functions terminal when chosen to be a function is 20% since there are 5 functions in total in the function set. When chosen to be a terminal, the change of being any particular terminal for pacman is 20% since there are 5 terminals. For the ghost this is 33% since there are only three terminals for the ghost. In the grow method for both the pacman and the ghost controllers for any particular node the chance of being a terminal is 50%, and 50% chance it is a function.

Experimental Setup

There are in total three experiments. The parameter(s) that seem to have the largest affect on the fitness of pacman is the μ and λ parameters for pacman. Experiment one had $\mu = 100$ and $\lambda = 100$, experiment two had $\mu = 50$ and $\lambda = 50$. experiment three had $\mu = 25$ and $\lambda = 25$. The following parameters were kept constant throughout all three experiments. The parent selection operator is over-selection with $x\% = 20$ as this yielded results slightly better than fitness proportional selection. The survival selection operator used was truncation. The parsimony pressure coefficient for both the ghost and pacman are both 3, and the maximum size of a tree is 300 before a penalty took place. Evaluations for termination is set to 2000 since that is the number of evaluations set for each run in an experiment. The mutation chance is 10%, since a low mutation rate is recommended.

Results

Table 1 is the results from the first experiment, table 2 is the results form the second experiment, table 3 is the results from the third experiment. Table 4 is the results of running the f test to determine if the average best fitness over all 30 runs for each experiment has difference variances with the other experiments.

The each f test let our null hypothesis be "true ratio of variances is equal to one" and let the alternative hypothesis be "true ratio of variances is not equal to one". By the results on Table 4, we can see that experiments do not equal variances, experiments 2 and 3 do not have equal variances, and experiments 1 and 3 do not share the same variances. Thus by these results, we reject the null hypothesis to use the alternative. We shall use a two-tailed t test with $\alpha = 0.05$ to determine if there exist a statistical significance between the results from each experiment.

The results of the T test are in Table 5. Let our null hypothesis be "True different in mean is equal to zero", and our alternative hypothesis be "True difference in mean is not equal to zero". Using the table we can see that for experiment 1 vs. experiment 2, $p = 0.9182$ so $p > \alpha$ so we cannot reject the null hypothesis, there is no statistical significant difference between experiment 1 and experiment 2. For experiment 1 vs. experiment 3, we see $p = 2.504 * 10^9$ so $p < \alpha$ rejecting the null hypothesis in favor of the alternative, showing that there is a statistical significance between experiment 1 and experiment 3. For experiment 2 vs. experiment 3, we see $p = 4.744 * 10^{16}$ so $p < \alpha$ so we reject the null hypothesis in favor of the alternative hypothesis, showing that there is a statistical difference between experiment 2 and experiment 3.

Between experiments 1 and 3 and between experiments 2 and 3, the experiments with the highest mean must be the statistically better algorithm. From Table 6 we see that experiments 1 and 2 both have larger means so experiments 1 and 2 are both statistically better than experiment 3. There is no statistical significant difference between experiments 1 and 2.

Discussion

From the results of the statistical analysis, the population size of pacman paired with the generation step has no statistical impact on the results of the experiment. An interesting result of the experiments is how similar the pacman controllers are each time I conducted an experiment. Nearly every time I had ran the experiment, the optimal strategy found was to simply wait out the time and consume as must fruit as possible. This is possibly a side affect of how pacman chooses his next move, since pacman will take the direction that yield the smallest result of the GP tree, and thus the controller wants to minimize the distance between pacman and the fruit.

While the pacman controllers can be excellent, the ghost controllers seem to have some problems. The pacman population seems to be dominant over the ghost during evolution. We believe that this stems from the fact the fitness of the ghost is the negation of the pacman controller. This means that a poor pacman controllers that cannot get points means that the fitness of the ghost controller is the maximum. This leads to ghost controllers that are poor may have the maximum fitness because of the chance they were paired with poor pacman controllers. This can lead to the ghost breed poor solutions initially and then plague their population with bad solutions. In order to combat this, we had made the the ghost μ and λ smaller than their pacman counter parts so that the ghost controllers can play multiple games and use the average of the games played. Ideally each controller in a particular population would play a game with every controller in the opposing population for their fitness score.

Conclusion

We conclude by saying that while the μ and λ are kept constant for the ghost population, there is no statistical difference between $\lambda = 100, \mu = 100$ and $\lambda = 50, \mu = 50$ for the pacman population. Also that while the pacman controller can be very fit, the ghost population suffers from being dominated by the pacman controllers due to the possibility of breeding poor controllers.

Average Best Fitness	Evaluation
100	74.93333
200	90.63333
300	97.53333
400	105.96667
500	112.43333
600	118.46667
700	119.56667
800	125.80000
900	129.20000
1000	130.63333
1100	131.30000
1200	132.16667
1300	136.03333
1400	137.53333
1500	138.43333
1600	138.80000
1700	142.63333
1800	142.96667
1900	143.96667
2000	144.50000

Table 1: Experiment One

Best Average Fitness	Evaluation
50	59.23333
100	75.50000
150	81.66667
200	87.73333
250	95.30000
300	101.20000
350	104.30000
400	106.83333
450	110.90000
500	113.26667
550	120.13333
600	122.03333
650	122.66667
700	122.80000
750	124.13333
800	124.66667
850	125.90000
900	127.03333
950	130.90000
1000	131.70000
1050	134.13333
1100	134.63333
1150	135.86667
1200	136.70000
1250	136.70000
1300	138.03333
1350	138.70000
1400	138.70000
1450	140.36667
1500	141.13333
1550	142.30000
1600	142.60000
1650	142.60000
1700	143.30000
1750	144.13333
1800	145.23333
1850	146.26667
1900	146.26667
1950	146.83333
2000	147.46667

Table 2: Experiment Two

Best Average Fitness	Evaluation
50	41.43333
100	51.96667
150	61.28333
200	69.16667
250	71.30000
300	72.76667
350	75.76667
400	78.31667
450	79.91667
500	80.68333
550	81.75000
600	82.03333
650	82.45000
700	83.21667
750	83.21667
800	83.30000
850	83.30000
900	84.13333
950	84.40000
1000	84.80000
1050	85.10000
1100	85.13333
1150	85.48333
1200	86.91667
1250	87.78333
1300	88.26667
1350	88.26667
1400	88.28333
1450	88.76667
1500	89.03333
1550	89.46667
1600	89.46667
1650	89.46667
1700	89.93333
1750	91.35000
1800	91.88333
1850	92.08333
1900	92.16667
1950	92.16667
2000	92.38333

Table 3: Experiment Three

	F	num df	den df	p value	95% interval	ratio
Experiment 1 vs. Experiment 2	0.81007	19	39	0.6347	(0.386486, 1.893682)	0.810066
Experiment 1 vs. Experiment 3	3.2469	19	39	0.001816	(1.549087, 7.590127)	3.246851
Experiment 2 vs. Experiment 3	4.0081	39	39	$3.296*10^5$	(2.119898, 7.578252)	4.008132

Table 4: F Test between the average best fitness over all 30 runs for all experiments

	t	df	p value	95% interval
Experiment 1 vs. Experiment 2	-0.10329	41.914	0.9182	(-11.74116, 10.59783)
Experiment 1 vs. Experiment 3	9.0288	25.013	$2.504*10^9$	(32.58186, 51.83814)
Experiment 2 vs. Experiment 3	11.194	57.32	$4.744*10^{16}$	(35.12947, 50.43387)

Table 5: T Test between the average best fitness over all 30 runs for all experiments

	Mean	Standard Deviation
Experiment 1	124.675	19.46243
Experiment 2	125.2467	21.62404
Experiment 3	82.465	10.80105

Table 6: Means and Standard Deviations

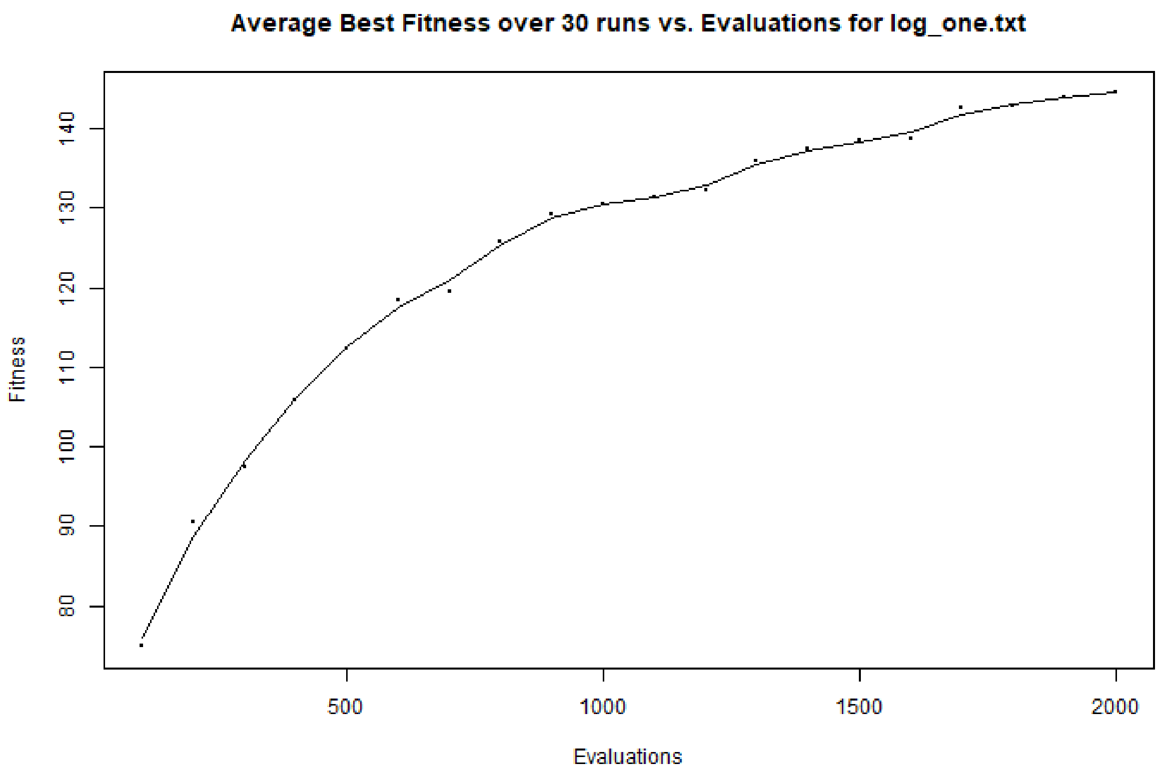

Figure 1: Experiment One

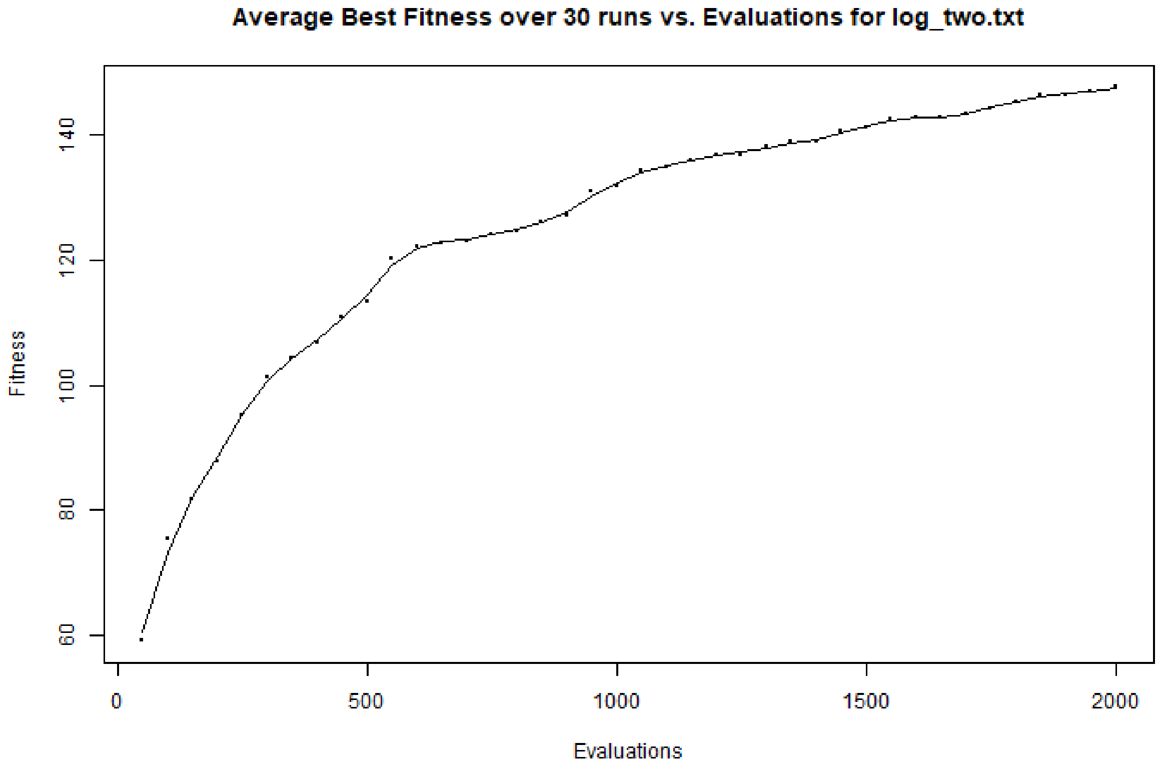

Figure 2: Experiment Two

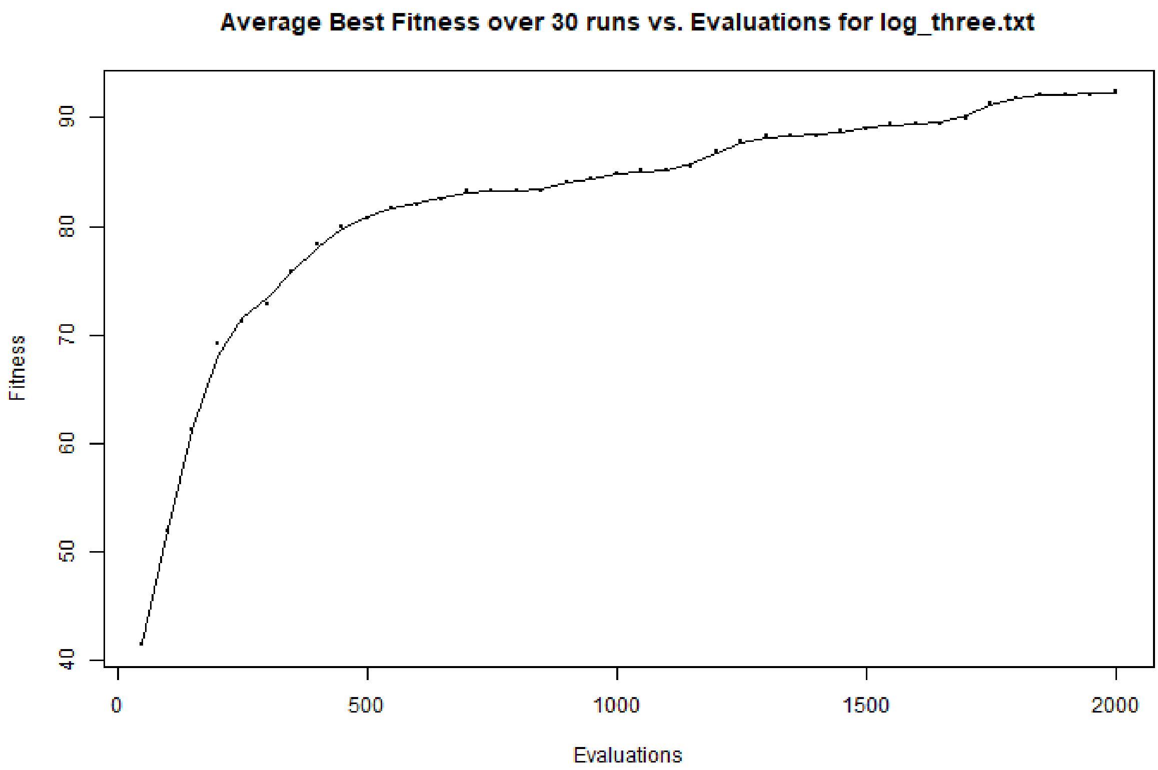

Figure 3: Experiment Three