AI PS5

Saturday, March 27, 2021 4:07 PM

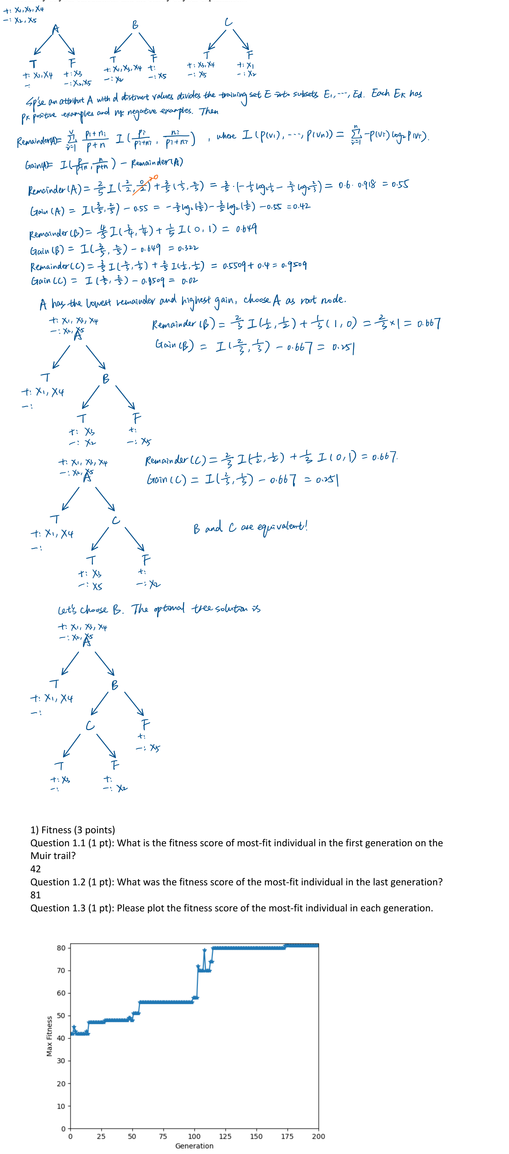
Problem #1 : Decision Trees (6 Points)

Given the dataset below that consists of five examples (X_1 through X_5) for three Boolean attributes (A, B, and C) and a Boolean outcome, you are to assemble (by hand) an optimal decision tree for this data.

	A	B	C	Outcome
X_1	T	T	F	T
X_2	F	T	F	F
X_3	F	F	T	T
X_4	T	T	T	T
X_5	F	F	T	F

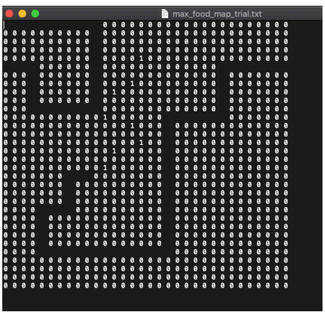
Your solution should show all calculations, starting with a comparison between using each of the three attributes (A, B, C) as the root of the tree. You should show the Remainder values for each comparison, and show the final decision tree that represents the optimal decision tree.

Please submit your solution to Gradescope. Your solution will be scored based on the accuracy of your calculations and the clarity of your explanations.

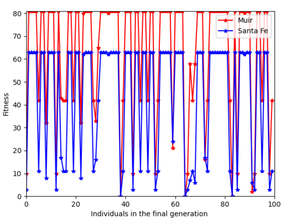


1) Fitness (3 points)
Question 1.1 (1 pt): What is the fitness score of most-fit individual in the first generation on the Muir trail?
42
Question 1.2 (1 pt): What was the fitness score of the most-fit individual in the last generation?
81
Question 1.3 (1 pt): Please plot the fitness score of the most-fit individual in each generation.

2) Please take a screenshot of the Path/trial of the most-fit individual in the final generation on the Muir trail. (2 points)



3) Trials comparison. (4 points)
Question 3.1 (1 pt): Please plot the individuals in the final generation and compare their performances on the Muir trail with their performances on the Santa Fe trail.



We first counted the number of food on each trail. There're equal amount of food (89) on both trails. From the plot we can conclude that for the final generation, performances on Muir trail is better than the performances on the Santa Fe trail.

Question 3.2 (3 pts): Do individuals that do well on one trail tend to do well on the other, and why?
Yes. This is obvious from the plot. This is because the genes encode how the ant finds the food. The food locations in the two maps share some common pattern, i.e. they are continuous. Hence, the gene that works well for some map will probably work well on the other map.

4) Your Implementation (15 points)
Question 4.1 (10 pts): Please provide a brief description of your genetic algorithm:

We use selection, cross over, and mutation to complete the genetic algorithm. In the beginning, the whole population is generated randomly. We randomly select a number from 1-4 to the first character of a state and select a number from 0-9 for the remaining two numbers. For each iteration, the individual genes in the population are input into the simulator to calculate the fitness scores, which are the number of foods that the ant eats in 200 steps. Then, all the individuals go over selection, cross over, and mutation. The new produced population replace the old population (generational selection).

The selection algorithm is implemented by rank selection, i.e. the fitness scores are ranked from the lowest score to the highest score. The ranks are normalized to serve as the probability to select the individual genes. We use two point cross over to combine every pair of the selected genes to a new pair of genes. In other words, given two positions, the initial part and the end part of the first gene and the middle part of the second gene is combined to form a new gene. The other parts of the two genes are formed to create the other gene. Finally, for mutation, we just randomly set a character to another character other than the current character (1-4 if the character is the first character of a state and 0-9 if the character is another character of the state).

Question 4.2 (5 pts): Please document the parameters you chose (please feel free to add more parameters in the empty cells):
Size of the population: 100
Number of generation: 200
Crossover probability: 0.1
Mutation probability: 0.02