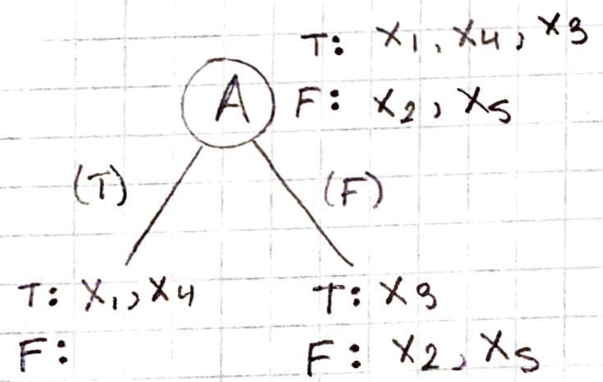
Problem 1

$$\text{remainder} = \sum_{i=1}^{v} \frac{p_i+n_i}{p+n} \, I\left(\frac{p_i}{p_i+n_i}, \frac{n_i}{p_i+n_i}\right)$$

$$I\left(\frac{p_i}{p_i+n_i}, \frac{n_i}{p_i+n_i}\right) = -\frac{p_i}{p_i+n_i}\log_2\left(\frac{p_i}{p_i+n_i}\right) - \frac{n_i}{p_i+n_i}\log_2\left(\frac{n_i}{p_i+n_i}\right)$$

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

- root A:



$$\text{remainder} = \frac{2}{5} I\left(\frac{2}{2}, 0\right) + \frac{3}{5} I\left(\frac{1}{3}, \frac{2}{3}\right)$$

$$= \cancel{0.4\left[-\log_2(1) - 0\right]} + 0.6\left[-\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{2}{3}\log_2\left(\frac{2}{3}\right)\right]$$

$$= 0 + 0.6\left[0.928\right]$$

$$\boxed{= 0.5568}$$

• root B

(B) T: X1 X3 X4
F: X2 X5

(T)
T: X1 X3 X4
F: X2

(F)
T:
F: X5

$$\text{remainder} = \frac{4}{5} I\left(\tfrac{3}{4}, \tfrac{1}{4}\right) + \cancel{\frac{1}{5} I(0,1)}_{\,0}$$

$$= 0.8 \left[-\frac{3}{4} \log_2\left(\tfrac{3}{4}\right) - \tfrac{1}{4} \log_2\left(\tfrac{1}{4}\right) \right]$$

$$\boxed{= 0.649}$$

• root C

(C) T: X1 X3 X4
F: X2 X5

(T)
T: X3 X4
F: X1

(F)
T: X5
F: X2

$$\text{remainder} = \frac{3}{5} I\left(\tfrac{2}{3}, \tfrac{1}{3}\right) + \frac{2}{5} I\left(\tfrac{1}{2}, \tfrac{1}{2}\right)$$

$$= 0.6 \left[-\frac{2}{3} \log_2\left(\tfrac{2}{3}\right) - \tfrac{1}{3} \log_2\left(\tfrac{1}{3}\right) \right]$$

$$+ 0.4 \left[\tfrac{1}{2} \log_2\left(\tfrac{1}{2}\right) - \tfrac{1}{2} \log_2\left(\tfrac{1}{2}\right) \right]$$

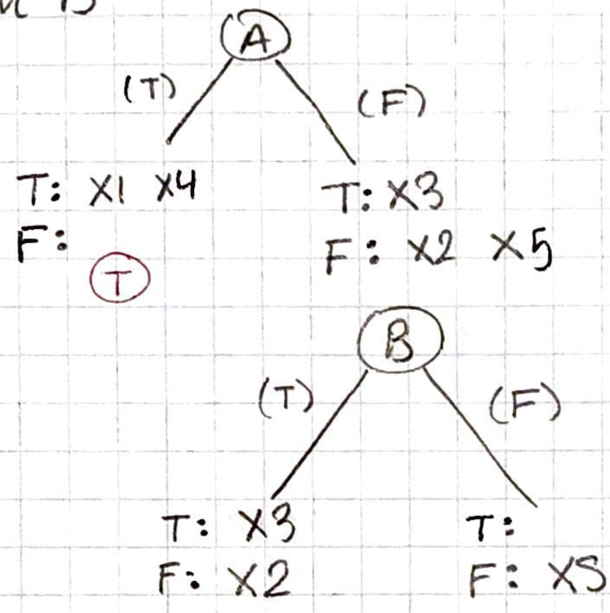
$$= 0.6 [0.928] + 0.4 [1]$$

$$\boxed{= 0.9568}$$

A : 0.5568
B : 0.649
C : 0.9598

Choose A as root, since it requires the least information

- check B

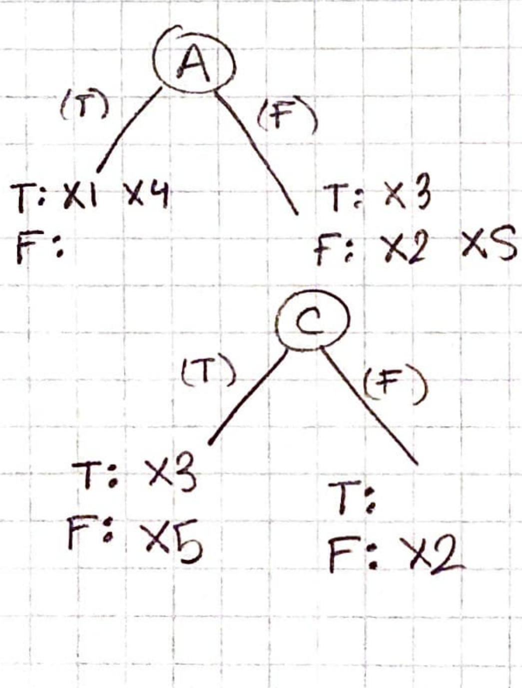


remainder : $\frac{2}{3} I(\frac{1}{2},\frac{1}{2}) + \underbrace{\cancel{\frac{1}{3} I(0,1)}}_{0}$

$$= \frac{2}{3}\left[-\frac{1}{2}\log_2(\frac{1}{2}) - \frac{1}{2}\log_2(\frac{1}{2})\right]$$

$$\boxed{= 0.667}$$

- check C

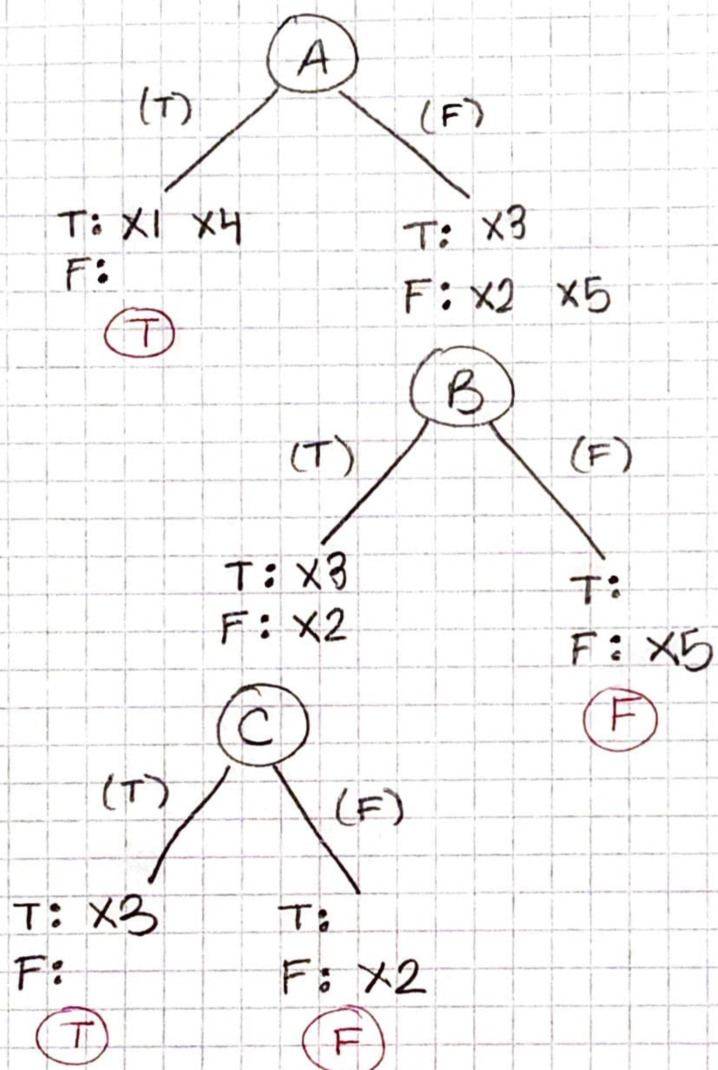


remainder: $\frac{2}{3} I(\frac{1}{2},\frac{1}{2}) + \cancel{\frac{1}{3} I(0,1)}$

$$= \frac{2}{3}\left[\frac{-1}{2}\log_2(\frac{1}{2}) - \frac{1}{2}\log_2(\frac{1}{2})\right]$$

$$\boxed{= 0.667}$$

Since B and C equivalent, choose B at random



This is optimal decision tree

Nicholas Archambault
CPSC 470
PS5

Question 1.1: 32

Question 1.2: 81

Question 1.3: Please see 'max_fitness_per_gen.png'

Question 2: Please see 'max_trial.txt'

Question 3.1:
Please see 'fitness_of_last_gen.png' to understand comparison visually. Individuals perform better on the Muir trail than on the Santa Fe trail. On Muir, they achieve a peak fitness of 81, while peak fitness achieved on Santa Fe is just 61.

Question 3.2:
Yes, individuals that perform well on one trail tend to perform well on the other because of general principles and traits that give individuals a high performance floor. Examples of such traits include

- if sensor is true, action triggered is "move forward"
- if sensor is false, action triggered is "rotate" clockwise or counterclockwise
- if sensor is false, do nothing and remain at the current state in order to avoid being caught in an infinite loop
- etc.

These types of built-in aspects allow for versatile, relatively stable performance across different trails where certain configurations might be different, but overall environment and the response it provokes remain similar.

Question 4.1/4.2/4.3:

- population size: 100
- max generations: 200
- crossover probability: 0.3
- mutation probability: 0.2
- number of elites to choose: 5
- mutation probability on single digit: 0.5

Selection incorporates elitism followed by rank selection. The fittest five individuals are chosen from the previous generation, though no additional consideration is given to the magnitude of their fitness. Parents are chosen according to ranks.

Crossover implements the three-point technique: each parent is split into four parts and the offspring of the two individuals is composed of the first and third part of one parent combined with the second and fourth part of the other. With no crossover happening, one of the parents is randomly selected as the offspring.

Mutation occurs with a probability of 0.5 on each digit. I chose (what I think is) a relatively high mutation probability in order to stimulate diversity and hopefully optimize fitness. Replacement governed by same elitism + rank selection strategy as above.

Please see 'geneticAlgorithm.py' to view the code.