Assignment #4

Instructor: Dr. Praveen Paruchuri

Guidelines

1. Construct a decision tree for the problem corresponding to ((roll numbers)%10+ 1). For the chosen decision tree, flip entries[1-indexed](make Yes as No and Vice Versa) for

   (a) (LastDigitofRollNum+SecondLastDigitofRollNum)%NoOfEntriesInDataset+1

   (b) (SecondLastDigitofRollNum)%NoOfEntriesInDataset+1

   If both of the above are equal flip:

   (a) (SecondLastDigitofRollNum)%NoOfEntriesInDataset+1

   (b) (SecondLastDigitofRollNum+1)%NoOfEntriesInDataset+1

   Example : For roll number is 20161075, construct decision tree for dataset 6, with label of row 1 as Yes(flipped) and label of row 8 as Yes(flipped).

2. Follow the textbook for decision tree algorithm [Greedy version]

3. Decision tree needs to be calculated manually without help of any library. On account of missing steps, you will be appropriately penalized.

4. Refer to submission format at the end of the PDF.

Problems

Problem 1: Basketball Champ   (100 points)

Manas has recently discovered his skill to play make 3-pointers in Basketball. After taking Machine, Data and Learning Course this semester, using feature engineering he has understood that only Wind speed, Angle and Spin are going to make sure if the shot is going to be made or not. Hence he has asked his TA friends to make sure his batch-mates can determine if his shot is going in or not and make an optimal decision tree for it. Hence using the data from below table determine the optimal decision tree.

Wind(mph)	Angle(degree)	Spin(rotations)	Shot
90	35	7	Y
75	40	12	Y
93	48	3	Y
85	38	12	N
77	42	10	N
97	50	7	N
105	47	6	Y
110	48	10	N
90	55	7	N
85	42	11	N

Table 1: Dataset *BB*

Problem 2: Ideal Friend (100 points)

Your TA is out searching for his ideal girlfriend. Based on his prior experience and browsing through pro-files, he has created the following table to find the ideal match. Create a decision tree to make his work easier.

Eye-Color	Height(cm)	Hair-length	Ideal
Black	170	Long	N
Black	179	Long	N
Black	169	Short	Y
Black	177	Short	Y
Brown	176	Long	Y
Brown	164	Short	N
Brown	176	Short	N
Brown	178	Short	N
Brown	164	Long	Y
Black	177	Long	N

Table 2: Dataset *GF*

Problem 3: Bull's eye (100 points)

After watching a movie, your TA is highly interested in determining if the sniper shot is going to make a kill or not. He is bad at physics but good at ML. Hence to confirm his hypothesis, he asks you to come up with a decision tree which fits the below data very well.

Horizontal Angle(degree)	Distance(m)	Wind Speed(mph)	Kill
1.5	450	220	N
4.5	520	-120	Y
3	490	120	Y
5.5	530	117	N
3.2	470	-170	N
5.2	505	-90	Y
1.85	465	120	Y
4.8	517	147	Y
1.7	430	-100	Y

Table 3: Dataset *SK*

Problem 4: Topple (100 points)

After watching a lot of Rohit Shetty movies, your TA knows he can no more rely on physics to see if a car is going to topple or not. Hence he asks for your help to create a decision tree and see if given car will topple or not based on the following data.

Horizontal Angle(degree)	Distance(m)	Wind Speed(mph)	Kill
1.5	450	220	N
4.5	520	-120	Y
3	490	120	Y
5.5	530	117	N
3.2	470	-170	N
5.2	505	-90	Y
1.85	465	120	Y
4.8	517	147	Y
1.7	430	-100	Y

Table 4: Dataset RS

Problem 5: Holiday (100 points)

You are fed up and want to take a break, and luckily you see a small continuous stretch of holidays. You have two options of spending the holidays – going on a trip or staying in college and relax. You would prefer to go on a trip, but it depends on the weather. You remember some previous instances when you chose to go on a trip based on the weather conditions. Design a decision tree which if given the current weather conditions would predict if you should go on the trip or not.

Forecast	Temperature	Humidity	Wind	Go on a trip
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes

Table 5: Dataset HD

Problem 6: Health (100 points)

You want to assess if you are living a healthy lifestyle or not but you can't make an objective decision about yourself on the topic. You observe habits of a few people around you and decide if that lifestyle is healthy or not. Based on your view of other people's lifestyle design a decision tree to see given your habits are you leading a healthy lifestyle or not.

Sleep Pattern	Junk Food Consumption	Exercise daily	Healthy
Irregular Sleep	High	Yes	No
Irregular Sleep	Normal	No	No
Irregular Sleep	Low	Yes	Yes
Irregular Sleep	Low	No	Yes
Good Sleep	High	Yes	Yes
Good Sleep	Normal	No	No
Good Sleep	Low	Yes	Yes
Good Sleep	High	No	No
Long Sleep	High	No	No
Long Sleep	Normal	Yes	Yes
Long Sleep	Low	Yes	Yes
Long Sleep	Normal	No	Yes

Table 6: Dataset SC

Problem 7: Assignment (100 points)

You have a assignment deadline for MDL and you want to decide if you should solve the assignment or not. As you have already been in college for nearly 2 years you have submitted multiple assignments and know which assignments you actually needed to submit. Based on your experience till now design a decision tree which will help you decide in future which assignments you should solve and which you can leave.

Weightage	Performance in the course till now	Workload	Do assignment
8%	Good	High	No
12%	Good	Medium	Yes
2%	Good	Low	Yes
6%	Good	High	No
8%	Average	Medium	Yes
6%	Average	Low	Yes
12%	Average	High	No
2%	Average	Medium	No
8%	Bad	Low	Yes
12%	Bad	High	Yes
2%	Bad	Medium	Yes
6%	Bad	Low	Yes

Table 7: Dataset AD

Problem 8: Corona Virus (100 points)

With the recent rise of positive corona cases, the World Health Organization (WHO) decided to put this issue on high priority and take every possible measure. So, they decided to install mini health-checkup stalls everywhere around the world where people can get a quick checkup done to see if they test positive or negative for corona. Unfortunately, they could not have a doctor for each stall. So, they decided to have a volunteer for each stall and for obvious reasons, they do not want the volunteers to mess up with the health of people. So to make the job of volunteers easy, WHO decided to design a decision tree where a person can be quickly tested based on the factors (a) **Country** i.e the place he lives in, (b) **Flu/Cough** which indicates whether he has Flu/Cough, (c) **SFT** i.e Sudden Fantasy of Travelling around the world. As WHO already spent a lot of money on treating corona patients, they are short of money and are desperate to get this work done for free. So, they requested the TAs of MD&L course to throw this as an assignment. Hence, construct a decision tree based on the data provided by WHO.

Country	Flu/Cough	SFT	Positive Case
China	Yes	Yes	Yes
China	Yes	No	Yes
China	No	Yes	Yes
Italy	Yes	No	No
Italy	Yes	Yes	Yes
Italy	No	Yes	Yes
India	Yes	Yes	Yes
India	Yes	No	No
India	No	No	No
USA	No	Yes	Yes
USA	Yes	No	No

Table 8: Dataset CV

Problem 9: Make Felicity great again!   (100 points)

After the recently concluded annual fest Felicity '20, everyone has super high hopes on next year's fest Felicity '21 and with the FC elections around the corner, Mr. Nerd desperately wants to become an FC. He has already started preparing for his campaign and speech and as his name suggests, he wants his campaign to be as nerdy as possible. To address the concern of increasing the participation for Felicity '21, he decided to show in his speech, a decision tree with some stats of Felicity '20 to the audience and talk about the measures he would take to increase the participation. Unfortunately, he did not take MD&L course and does not know how to construct decision trees and hence has requested you to construct one for him. Yes, as he is a nerd he could have learnt and done it by himself but unfortunately his girlfriend is mad at him and hence he wants to maximise the time he spends with her. So, why not help your batchmate?

| Profile	Introvert	Desire to travel	Attended Felicity
1st/2nd/3rd year	Yes	Yes	No
1st/2nd/3rd year	Yes	No	No
1st/2nd/3rd year	No	Yes	Yes
1st/2nd/3rd year	No	No	Yes
4th/5th year	Yes	Yes	No
4th/5th year	Yes	No	No
4th/5th year	No	Yes	No
4th/5th year	No	No	Yes
Outsider	Yes	Yes	No
Outsider	Yes	No	No
Outsider	No	Yes	No
Outsider	No	No	Yes

Table 9: Dataset FC

Problem 10: Who's doing this assignment? (100 points)

The TAs spent so much time setting this assignment and are now actually wondering how many students would put in a serious effort in solving this assignment. So, they decided to construct a decision tree themselves and check its accuracy after the deadline of the assignment. The TAs have considered three factors (a) **Single** i.e Relationship status, (b) **Performance in Mid exam**, (c) **CAG** i.e cares about grades. The TAs are super lazy and wanted to give this question too in the assignment. So, please help your TAs.

Single	Performance in Mid exam	CAG	Solved the Assignment
Yes	Good	Yes	Yes
Yes	Good	No	No
Yes	Average	Yes	Yes
Yes	Average	No	No
Yes	Bad	Yes	Yes
Yes	Bad	No	Yes
No	Good	Yes	No
No	Good	No	No
No	Average	Yes	Yes
No	Average	No	No
No	Bad	Yes	Yes
No	Bad	No	No

Table 10: Dataset ASSN
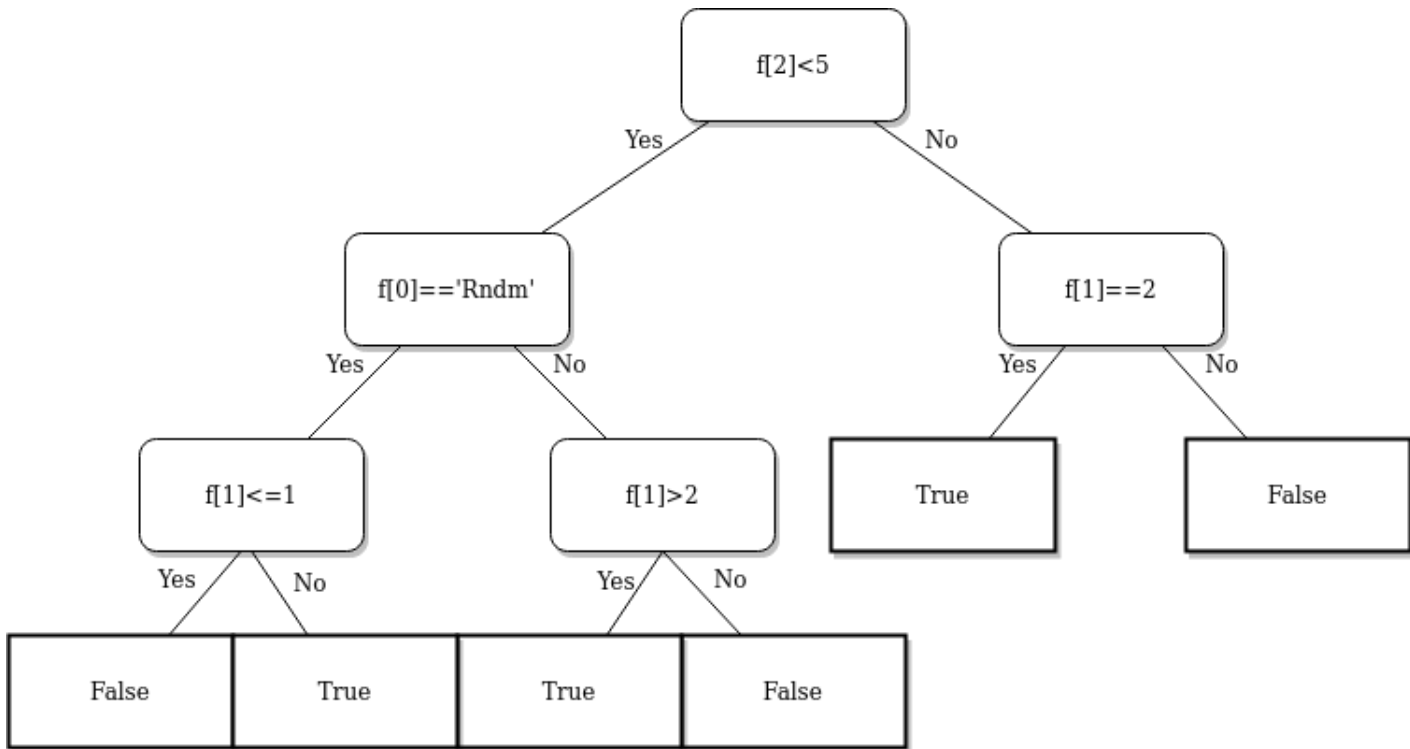
Submission Format

You will need to submit a zip of folder(RollNumber) containing two files:

- The evaluation part has an automated component. Make sure you follow the submission guidelines.

- A file(Rollnumber.txt) with the following syntax :

  - Line 1: DataseNumber(1-10)

  - Line 2 onwards : Put conditions for every path to leaf node separated by 'and'. Use $'str\_example'$ for strings[quotes]. Use == to check equality, comparison operators($<,>,\leq,\geq$) remain the same. Fields need to be mentioned as f[0],f[1],f[2],..., based on their position in table. At the end, separated by a comma, put the expected label(don't need quotes here). For example : $f[0] < 10$ and f[1]=='Sunny', Yes.

  - Example script for above tree :
    2
    $f[2] < 5$ and f[0]=='Rndm' and $f[1] <= 1$, False
    $f[2] < 5$ and f[0]=='Rndm' and $f[1] > 1$, True
    $f[2] < 5$ and f[0]!='Rndm' and $f[1] > 2$, True
    $f[2] < 5$ and f[0]!='Rndm' and $f[1] <= 2$, False
    $f[2] >= 5$ and $f[1] == 2$, True
    $f[2] >= 5$ and $f[1]! = 2$, False

- A brief report explaining how you reached your solution[step-wise decisions and entropy at each step]. (RollNumber.pdf)

- Make sure you report Entropy after each step.

- **Please show all steps involved.** On account of missing steps,you will be appropriately penalized