Rohit Ravishankar
rr9105@rit.edu

Home Work 4

CSCI - 630 Foundation of Intelligent Systems

Collaborators: None

1.

Examine the totally fictional data set here. Presume that this data was collected at Midnight Oil - each line represents one drink that was ordered, and whether it was hot, sweet, and/or caffeinated.

- Based on this data set, what is your best estimate of P(Caff)?
- What is your best estimate of P(Hot | Sweet)?
- Based on this data, do any two of these variables appear to be independent? If so, which two? If not, why not?

Ans.

1

(1)

a) $P(\text{caffe}) = \frac{15}{30}$

$= 0.5$

(b) $P(\text{Hot} \mid \text{Sweet}) = \frac{P(\text{Hot} \cap \text{Sweet})}{P(\text{sweet})}$

$P(\text{sweet}) = \frac{13}{30}$

$P(\text{Hot and Sweet}) = \frac{2}{30}$

$P(\text{Hot} \mid \text{Sweet}) = \frac{2}{\cancel{30}} \times \frac{\cancel{30}}{13} = \frac{2}{13} = 0.154$

(c)

$\rightarrow P(\text{Hot} \mid \text{Sweet}) = \frac{P(\text{Hot} \cap \text{Sweet})}{P(\text{Sweet})} = \frac{2}{13} = 0.154$

$P(\text{Hot}) = \frac{13}{30}$

$P(\text{Hot}) \not\approx P(\text{Hot} \mid \text{Sweet})$ $\therefore$ They are not independent

$\rightarrow P(\text{Hot} \mid \text{Caffe}) = \frac{P(\text{Hot} \cap \text{Caffe})}{P(\text{Caffe})} = \frac{4}{15}$

$P(\text{Hot}) = \frac{13}{30}$

$P(\text{Hot}) \not\approx P(\text{Hot} \mid \text{Caffe})$ $\therefore$ They are not independent

$\rightarrow P(\text{Sweet} \mid \text{Caffe}) = \frac{P(\text{Sweet} \cap \text{Caffe})}{P(\text{caffe})} = \frac{8}{15}$

$P(\text{Sweet}) = 13/30$

$P(\text{Sweet} \mid \text{Caffe}) \not\approx P(\text{Sweet})$ $\therefore$ They are not independent

$\therefore$ NONE ARE INDEPENDENT

2.

Consider the following set of boolean variables about a randomly chosen day:

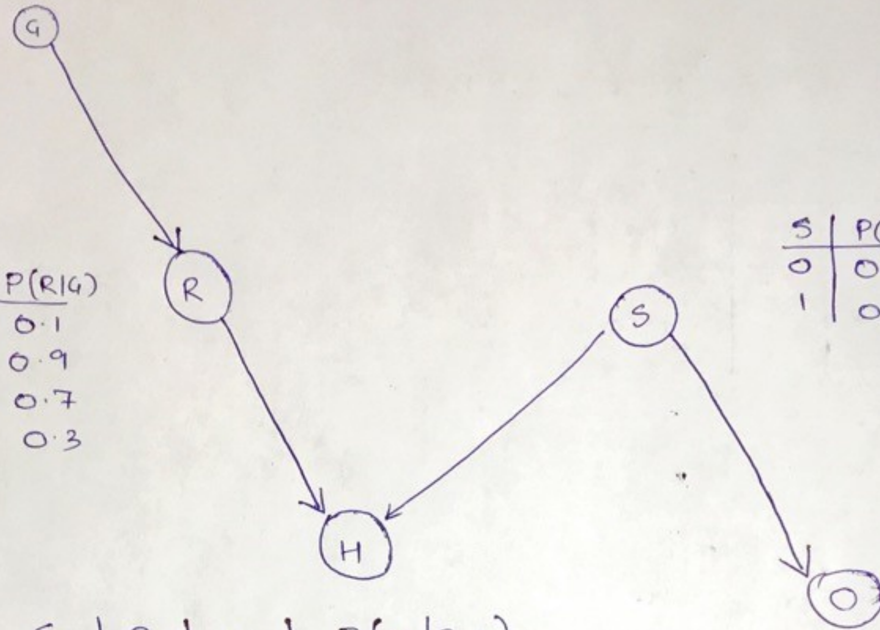
- G : I was up late grading the night before
- R : I get up feeling well-rested
- S : it is a pleasant sunny day
- H : I am feeling happy
- O : people are sitting outside to have coffee

a. Draw a Bayes net that reasonably represents the relationships between those variables.

b. Annotate the Bayes net with all necessary probabilities so that any possible question about the variables can be answered For the values, choose any reasonable number other than 0 or 1 (and don't just use the same value over and over!).

c. Compute the following probabilities based on your network. You may reuse any computations as needed.

- P(R)
- P(G^S)
- P(¬H)
- P(R^¬H)
- P(S|O)



G	P(G)
0	0.9
1	0.1

G	R	P(R|G)
0	0	0.1
0	1	0.9
1	0	0.7
1	1	0.3

S	P(S)
0	0.8
1	0.2

S	R	H	P(H|R,S)
0	0	0	0.25
0	0	1	0.75
1	1	0	0.35
1	1	1	0.65
0	0	0	0.55
0	0	1	0.45
1	1	0	0.15
1	1	1	0.85

S	O	P(O|S)
0	0	0.6
0	1	0.4
1	0	0.2
1	1	0.8

⊛ $P(R)$

$$P(R \wedge G) = P(R|G) \cdot P(G)$$

R	G	$P(R \wedge G)$
0	0	0.09
1	0	0.81
0	1	0.07
1	1	0.03

$P(R)$

R	$P(R)$
0	0.16
1	0.84

⊛ $P(G \wedge S) = P(G) \cdot P(S)$

G	S	$P(G \wedge S)$
0	0	0.72
0	1	0.18
1	0	0.08
1	1	0.02

(3)

(*) P(H)

$$P(H) = P(H|R=0,S=0) \cdot P(R=0) \cdot P(S=0) + P(H|R=0,S=1) \cdot P(R=0) \cdot P(S=1) + P(H|R=1,S=0) \cdot P(R=1) \cdot P(S=0) + P(H|R=1,S=1) \cdot P(R=1) \cdot P(S=1)$$

$$= (0.75)(0.16) \cdot (0.8) + (0.65)(0.16)(0.2) + (0.45)(0.84)(0.8) + (0.85)(0.84) \cdot (0.2)$$

$$= 0.562$$

$$P(\neg H) = 1 - P(H) = 0.438$$

H	P(H)
0	0.438
1	0.562

(*) $P(R \wedge \neg H)$

R	P(R)
0	0.16
1	0.84

H	P(H)
0	0.438
1	0.562

$$P(R \wedge \neg H) = P(R) \cdot P(\neg H)$$

R	H	¬H	$P(R \wedge \neg H)$
0	0	1	0.08992
0	1	0	0.07008
1	0	1	0.47208
1	1	0	0.36792

④

(*) P(S|O)

- $P(S \wedge O) = P(O|S) \cdot P(S)$

S	O	P(S ∧ O)
0	0	0.48
0	1	0.32
1	0	0.04
1	1	0.16

- P(O)

O	P(O)
0	0.52
1	0.48

- $P(S|O) = \dfrac{P(S \wedge O)}{P(O)}$

O	S	P(S|O)
0	0	0.923
0	1	0.077
1	0	0.67
1	1	0.33

3.

For each of the following machine learning scenarios, several possible input variables are listed. For each variable, briefly justify whether you think it would be useful, useful but impractical to obtain (for training and/or testing), or not useful. Also, suggest how accurate you think such a system could be, given the input variables you have chosen and a feasible amount of training data, and why.

1. Predicting a student's grade on the final exam in this course: grade on the midterm; grade in Algorithms; amount of time studying; number of characters on the student's cheat sheet; amount of coffee consumed in the week before the exam.
2. Predicting whether it will rain in Rochester tomorrow: whether it rained in Rochester today; whether it rained in Cleveland yesterday; the locations in the US where it rained yesterday; the wind speed in Rochester yesterday; the day of the week; the month of the year
3. Predicting whether a song will become a hit: previous sales figures of the artist; length of the song; company releasing the song; lyrics of the song
4. Predicting whether you will like a particular restaurant: the opinions of the last hundred people who ate there; the Yelp review of the restaurant; the type of food; the number of insects in the kitchen

3/

(5)

Predicting a student's final exam grade in this course

1. Grade on midterm - useful

Grade on algorithms - useful

Amount of time studying - not useful

Number of characters on student's cheat sheet - not useful

Amount of coffee consumed in the week before exam - not useful

→ The student's grade on the midterm would help judge the percentile of the student in the class and hence be used to predict where the student would stand in class on the final exam. → Algorithmic thinking forms the bases for many courses in CS and the grade on algorithms course would determine the problem solving skills of the student

→ Amount of time spent studying is not useful as a person studying less can possibly score more and vice versa.

→ Amount of coffee consumed in the week before the exam is not useful as a student not drinking coffee may also do better

→ Number of characters on cheat sheet does not help predict the final score as the cheat sheet has no bearing on how well one does on an exam

2. Predicting whether it will rain in Rochester tomorrow

whether it rained in Rochester today - useful

whether it rained in cleveland yesterday - not useful

~~whether~~ The locations in US where it rained yesterday - useful but impractical

the wind speed in Rochester yesterday - useful

The day of the week - not useful

The month of the year - useful

→ The day of the week has no bearing on the weather & hence isn't useful

→ The month of the year because it helps determine if it is the monsoon season

→ Wind speed is a useful attribute because wind can bring or take moisture away / bring or take clouds away based on where it passes (water body / land)

→ The rain in cleveland yesterday cannot be used to accurately determine rain in Rochester tomorrow and hence it is not useful

→ Whether it rained in Rochester today is useful because there is possibly moisture in the air & if it is the rainy season the probability of rain tomorrow might increase

→ Locations of rain in US could mark the direction of movement of winds but with so many locations it is impractical

3. Predicting whether a song will become a hit

Previous sales figures of the artist - useful

length of the song - useful

company releasing the song - useful

lyrics of the song - useful

→ Previous sales figures of the artist determine how popular an artist is. If the previous sales are high, a new song is likely to sell well too

→ If a song is 1 hour long, it is likely that a large section of crowd may not listen to it and hence, it is a useful variable

→ The lyrics of a song help the song connect with the audience and hence, lyrics of the song is a useful attribute / variable.

→ company releasing the song is important as some companies have more funds and hence maybe able to spend more on marketing the song and hence, possibly making the song a hit.

4. Predicting whether you will like a particular restaurant

The opinions of the last hundred people who ate there - useful but impractical

the Yelp review of the restaurant - useful

The type of food - useful

The number of insects in the kitchen - not useful

→ The opinions of the last 100 people maybe useful but impractical because it is small subset and hence maybe biased.

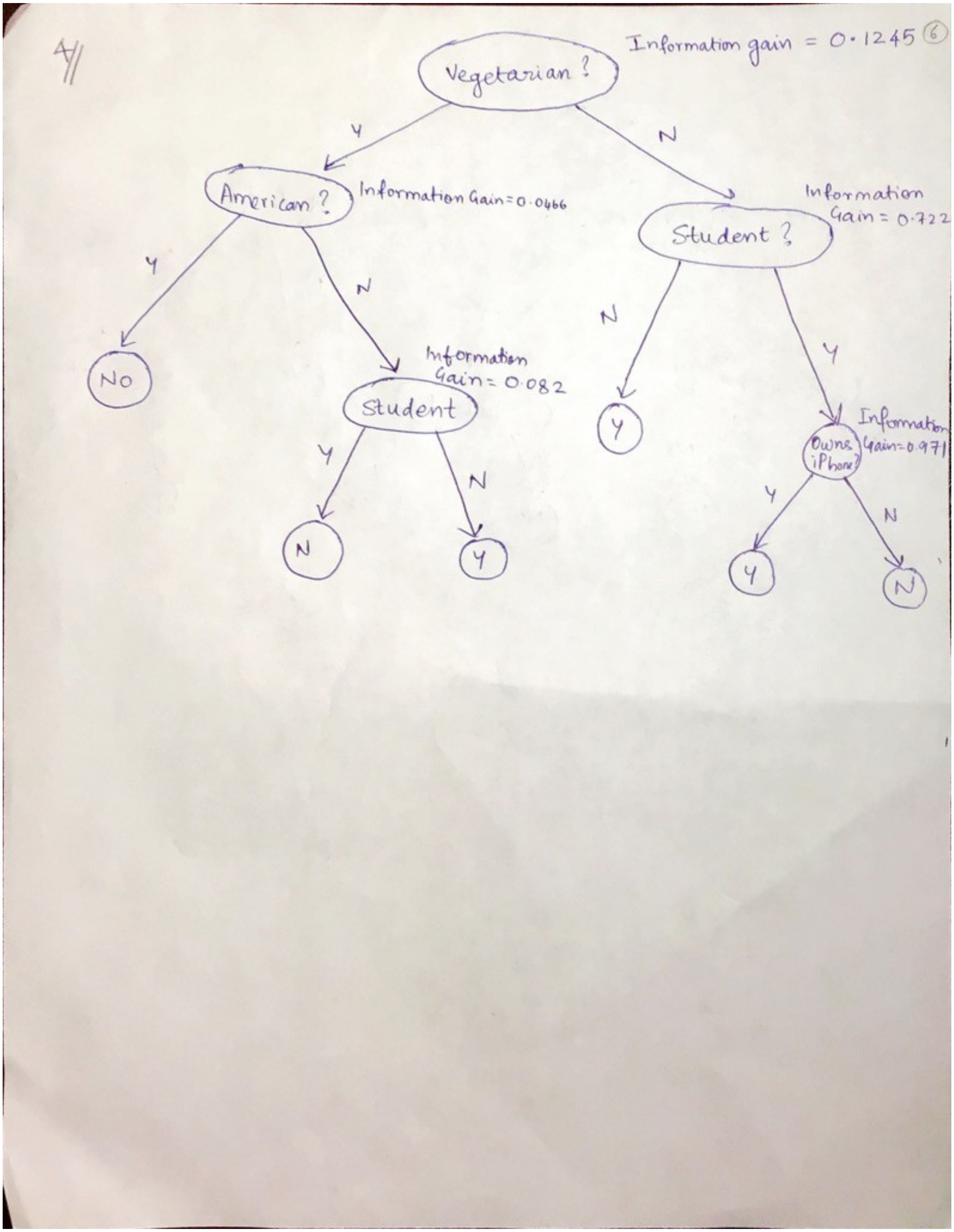
→ The Yelp review of the restaurant is useful because it provides a large subset of data to base your judgement

→ The type of food served in a restaurant maybe useful because it will help determine whether you will like that cuisine or not.

→ The number of insects in the kitchen is not an attribute you would know as a customer and it does not help in training or testing

4.

For the following data set, build a decision tree (using information gain) that best answers the question of whether a person drinks coffee based on the four given criteria. You should continue asking questions as long as the question gives some information gain up to a maximum tree depth of three; if you find more than one equally useful criteria at any point, pick any one. If you run out of useful questions to ask at any particular point, state that as well. You are free to use some code, or just do it by hand - you do not need to show all your entropy calculations (or code to do them), but do show them for the root node, and give at least the gain for each other node in your tree.


4/1
Information gain = 0.1245 ⑥
Vegetarian ?
Y
N
American ?
Information Gain = 0.0466
Information Gain = 0.722
Student ?
Y
N
No
Information Gain = 0.082
Student
N
Y
Y
Information Gain = 0.971
Owns iPhone?
Y
N
N
Y
Y
N

4/1 ⑦

Class Entropy for "drinks coffee"

Yes = 12
No = 8

$$E = \left(-\frac{12}{20}\log_2 \frac{12}{20}\right) + \left(-\frac{8}{20}\log_2 \frac{8}{20}\right)$$

$$= \underline{0.971}$$

① → Entropy for Student

Student	Drinks Coffee Y	N
Y	5	6
N	7	2

$$E_Y = -\frac{5}{11}\log_2 \frac{5}{11} - \frac{6}{11}\log_2 \frac{6}{11} = 0.994$$

$$E_N = -\frac{7}{9}\log_2 \frac{7}{9} - \frac{2}{9}\log_2 \frac{2}{9} = 0.764$$

→ Information Gain for student

$$IG_S = 0.971 - \frac{11}{20} \times 0.994 - \frac{9}{20} \times 0.764$$

$$= 0.0805$$

② → Entropy for American

American	Drinks Coffee Y	N
Y	6	4
N	6	4

$$E_Y = -\frac{6}{10}\log_2 \frac{6}{10} - \frac{4}{10}\log_2 \frac{4}{10} = 0.971$$

$$E_N = -\frac{6}{10}\log_2 \frac{6}{10} - \frac{4}{10}\log_2 \frac{4}{10} = 0.971$$

→ Information Gain for American

$$IG_A = 0.971 - \frac{10}{20} \times 0.971 - \frac{10}{20} \times 0.971$$

$$= 0 //$$

(3) → Entropy for Vegetarian

	Drinks Coffee	
Vegetarian	Y	N
Y	4	6
N	8	2

$$E_Y = -\frac{4}{10} \log_2 \frac{4}{10} - \frac{6}{10} \log_2 \frac{6}{10} = 0.971$$

$$E_N = -\frac{8}{10} \log_2 \frac{8}{10} - \frac{2}{10} \log_2 \frac{2}{10} = 0.722$$

→ Information Gain for Vegetarian

$$IG_V = 0.971 - \frac{10}{20} \times 0.971 - \frac{10}{20} \times 0.722$$

$$= 0.1245$$

(4) → Entropy for Owns iPhone

	Drinks Coffee	
Owns iPhone	Y	N
Y	6	3
N	6	5

$$E_Y = -\frac{6}{9} \log_2 6/9 \; * \; -3/9 \log_2 3/9 = 0.918$$

$$E_N = -6/_{11} \log_2 6/11 \; * \; -5/11 \log_2 5/11 = 0.994$$

→ Information Gain for Owns iPhone (8)

$$IG_P = 0.971 - \frac{9}{20} \times 0.918 - \frac{11}{20} \times 0.994$$

$$= 0.0112$$

$IG_S = 0.0805$ $IG_A = 0$ $IG_V = 0.1245$ $IG_P = 0.0112$

Iteration 2 - For Vegetarian? Yes

- Class Entropy for drinks coffee

Yes = 4
No = 6

$$E = -\frac{4}{10}\log_2\frac{4}{10} - \frac{6}{10}\log_2\frac{6}{10}$$

$$= 0.971$$

(1) Entropy/IG for Student

	Drinks coffee	
Student	Y	N
Y	2	4
N	2	2

$$E_Y = -\frac{2}{6}\log_2\frac{2}{6} - \frac{4}{6}\log_2\frac{2}{6} = 0.918$$

$$E_N = -\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4} = 1$$

$$IG_S = 0.971 - \frac{6}{10} \times 0.918 - \frac{4}{10} \times 1 = \underline{0.0212}$$

② Information Gain for American

American	Drinks Coffee Y	N
Y	1	3
N	3	3

$E_Y = -\frac{1}{4}\log\frac{1}{4} - \frac{3}{4}\log\frac{3}{4} = 0.811$

$E_N = -\frac{3}{6}\log\frac{3}{6} - \frac{3}{6}\log\frac{3}{6} = 1$

$IG_A = 0.971 - \frac{4}{10} \times 0.811 - \frac{6}{10} \times 1.0 = \underline{\underline{0.0466}}$

③ Information Gain for Owns iPhone

Owns iPhone	Drinks Coffee Y	N
Y	2	3
N	2	3

$E_Y = -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5} = 0.971$

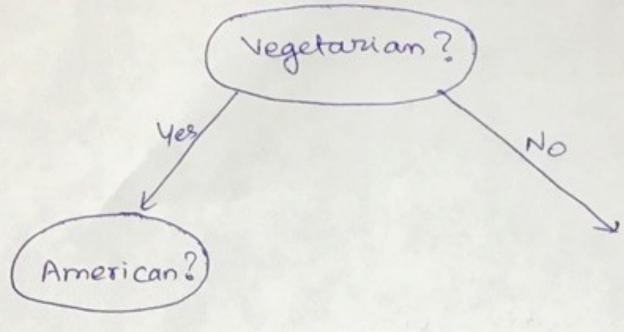
$E_N = -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5} = 0.971$

$IG_P = 0.971 - \frac{5}{10} \times 0.971 - \frac{5}{10} \times 0.971$

$= 0$ //

$IG_S = 0.0212 \quad IG_A = 0.0466 \quad IG_P = 0$

(9)



Iteration 3 - For American? Yes

Class Entropy for "drinks coffee"

$$E = -\tfrac{1}{4}\log\tfrac{1}{4} - \tfrac{3}{4}\log\tfrac{3}{4}$$

$$= 0.971$$

(1) **Information Gain for Student**

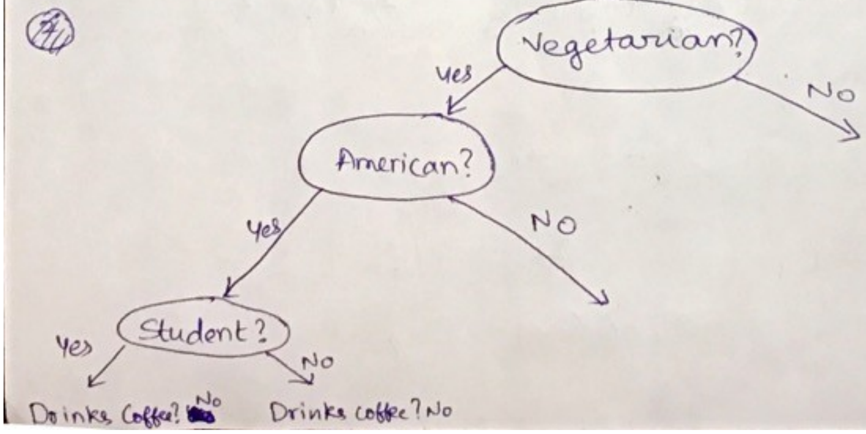
Student	Drinks Coffee Y	Drinks Coffee N
Y	1	2
N	0	1

$$E_Y = -\tfrac{1}{3}\log_2\tfrac{1}{3} - \tfrac{2}{3}\log_2\tfrac{2}{3} = 0.918$$

$$E_N = \tfrac{0}{1}\log_2(\tfrac{0}{1}) - \tfrac{1}{0}\log_2\tfrac{1}{0} = \text{undef}$$

We have reached a pure node

(4)



Iteration 3 - For American? No

$E = -\frac{3}{6}\log_2 3/6 - 3/6 \log_2 3/6$

$= 1$

① Information Gain for Student

Student	Drinks coffee Y	N
Y	1	2
N	2	1

$E_Y = -1/3 \log_2 1/3 - 2/3 \log_2 2/3 = 0.918$

$E_N = 0.918$

$IG_S = 1 - \frac{3}{6} \times 0.918 - \frac{3}{6} \times 0.918$

$= 0.082$ //

② Information Gain for Owns iPhone

Owns iPhone	Drinks coffee Y	N
Y	1	1
N	2	2

$E_Y = -1/2 \log_2 1/2 - 1/2 \log_2 1/2 = 1$ ~~$= 0.918$~~

$E_N = 1$

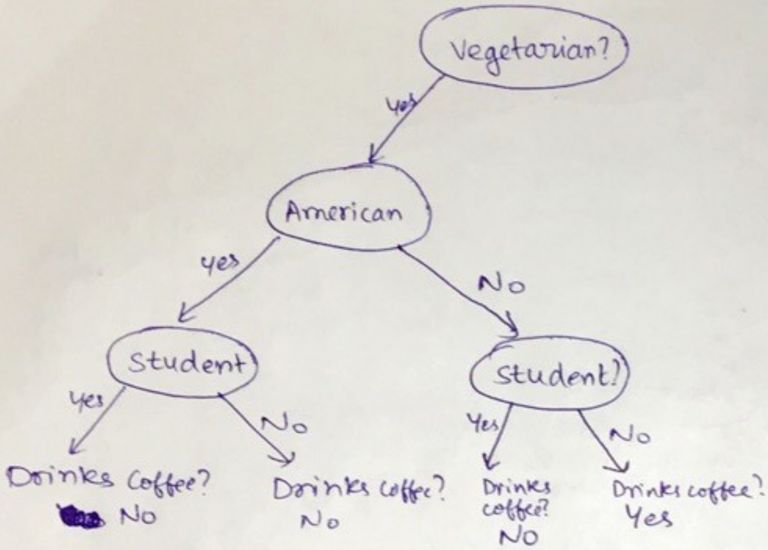
~~$IG_P = 1 - \frac{3}{6} 0.918 - \frac{4}{6}$~~

$IG_P = 1 - 2/6 - 4/6$

$= 0$ //

$IG_S = 0.082$ $IG_P = 0$

10



Iteration 2 - Vegetarian ? No

Class Entropy for Drinks coffee

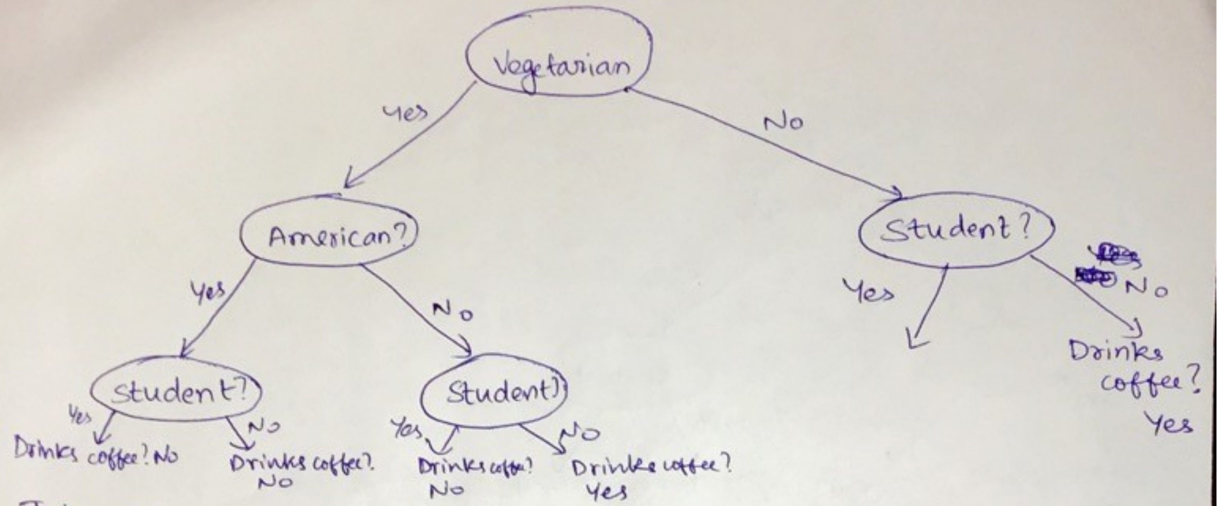
$$E = -\frac{8}{10} \log_2 \frac{8}{10} - \frac{2}{10} \log_2 \frac{2}{10} = 0.722$$

① Information gain for Student

Student	Drinks coffee Y	N
Y	3	2
N	5	0

$$E_Y = -\frac{3}{5} \log_2 3/5 - \frac{2}{5} \log_2 2/5$$

We have reached a pure node for Vegetarian ? No and Student ? No



Iteration 3 – For Student? ~~No~~ Yes

Class entropy for "drinks coffee"

$$E = -\frac{3}{5}\log_2 3/5 - \frac{2}{5}\log_2 2/5$$

$$= 0.971$$

① Information Gain for Owns iPhone

Owns iPhone	Drinks coffee Y	Drinks coffee N
Y	3	0
N	0	2

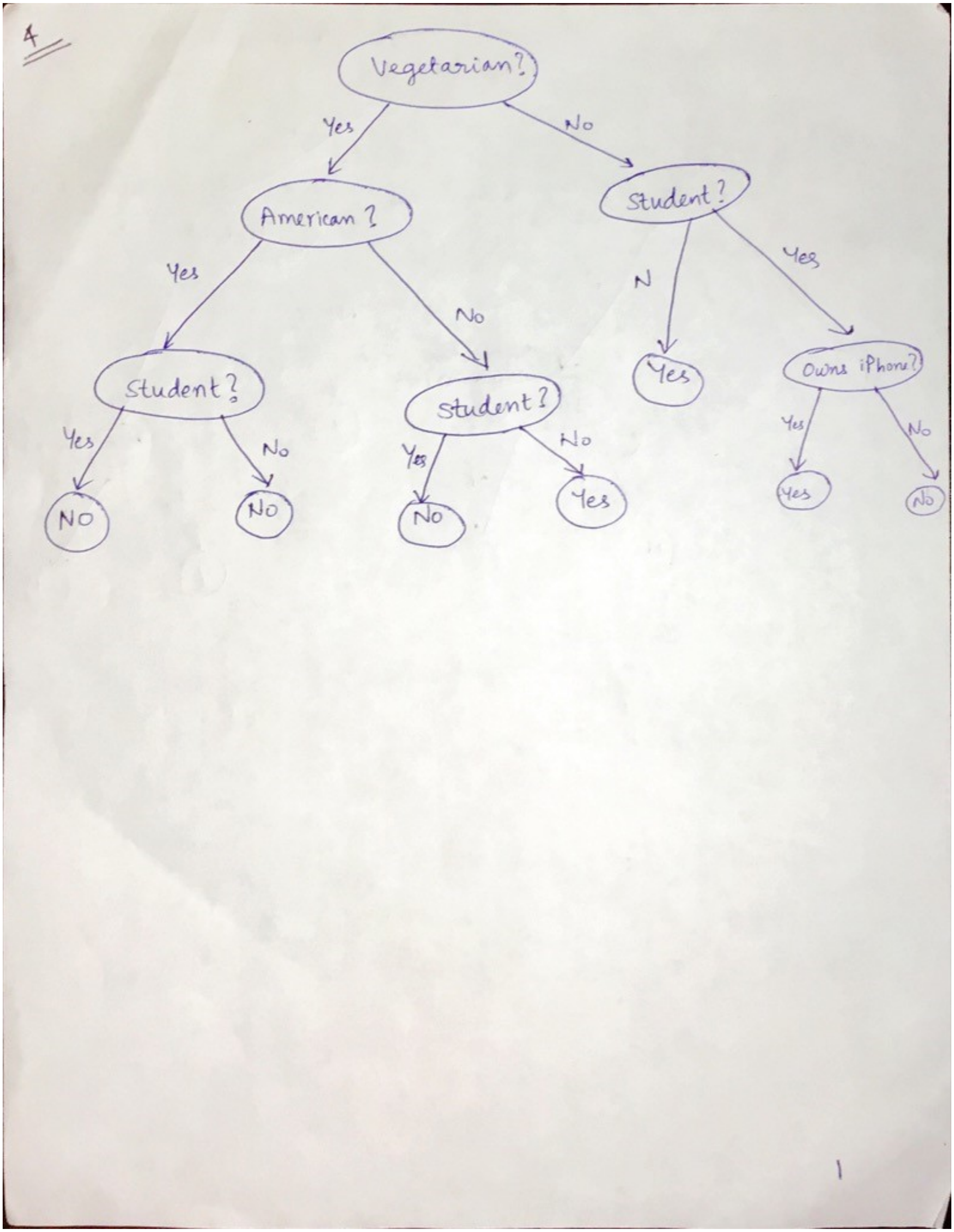
We have reached a pure node
vegetarian? No Student? Yes Owns iPhone

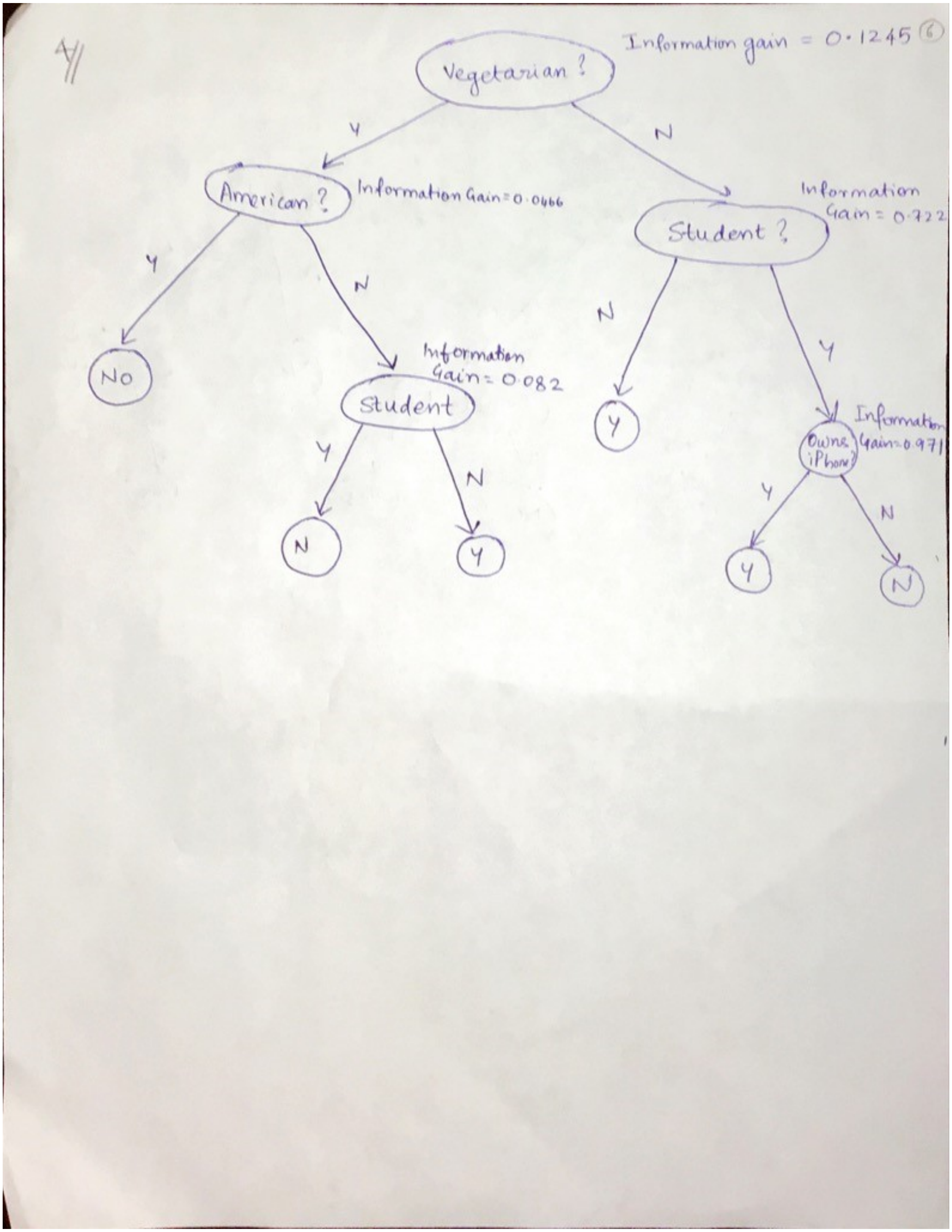
② Information Gain for American

American	Drinks coffee Y	Drinks coffee N
Y	1	1
N	2	1

$E_Y = 1$ $\quad E_N = 0.918$

$IG_A = 0.971 - 2/5 \times 1 - 3/5 \times 0.918 = 0.202$


4
Vegetarian?
Yes
No
American ?
Student?
Yes
No
N
Yes
Student?
Student?
Yes
Owns iPhone?
Yes
No
Yes
No
Yes
No
Yes
No
NO
No
No
Yes
Yes
No
1



4/
Information gain = 0.1245 6
Vegetarian ?
Y
N
American ?
Information Gain = 0.0466
Information Gain = 0.722
Student ?
Y
N
No
Information Gain = 0.082
Student
N
Y
Y
Information Gain = 0.971
Owns iPhone?
Y
N
N
Y
Y
N
Y
N

5.

Consider using Adaboost with decision stumps (using information gain) for the previous problem.

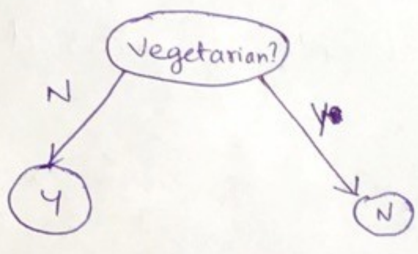
- What weight would each example (the **w**s in the Adaboost pseudocode) have initially?
- What would Adaboost select as the first stump? (Hint: this should be very easy if you doing the homework problems in order...) How would this stump make predictions (that is, what prediction would it make, based on the value of what attribute?
- After the first iteration, what would the example weights be for an example that the first stump got right, and for an example that the first stump got wrong?

5//

a) Each example would have a weight(w) of $1/20 = 0.05$

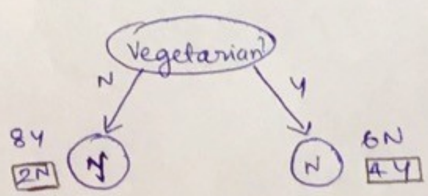
b) Adaboost would select Vegetarian? as the first stump. The stump predicts that if a person is a vegetarian, he/she does not drink coffee, and if the person is not a vegetarian he/she does drink coffee.

→ The stump would be:



→ The prediction is based on the ~~information gain~~ node purity.
→ The stump is chosen based on the information gain.

c) When vegetarian is chosen for the decision stump, there are 6 misclassifications.



So the error rate = $6/20$

coefficient of Classifier $= Z = \ln((1 - \text{error})/\text{error})$

$= \ln((1 - 6/20)/(1 - 6/20))$

$= 0.847$

weights for correctly classified entries $= wt * e^{-Z}$

$= 0.05 * e^{-0.847}$

$= 0.021$

weights for misclassified entries $= wt * e^{Z}$

$= 0.05 * e^{0.847}$

$= 0.117$

$\therefore$ It is visible that Adaboost gives higher weight to misclassified entries