Intro to Data Science HW 2

Ben Green

Contents

March 23, 2015

Question 1

1a.

The probability of a bit in the array remaining 0 is:

$$e^{\frac{-20}{99}} \tag{1}$$

which comes out to .819. So, 1-.819 = .181 is the expected fraction of 1's.

1b.

The expected fraction of 0's is 1-.181 = .819

Question 2

The false positive rate is:

$$(1 - e^{\frac{-3*2}{11}}) = (1 - e^{\frac{-6}{11}}) \tag{2}$$

Question 3

a	**b**	**c**	**a**	**d**	**e**	**a**	**c**	**b**	**b**
1									
.9	1								
.81	.9	1							
	.81	.9	1.81						
	.729	.81	1.1629	1					
	.6561	.729	1.4661	.9	1				
	.5905	.6561		.81	.9	2.4661			
	.5314			.729	.81	2.219	1.6561		
				.6561	.729	1.997	1.4904	1.5314	
				.5905	.6561	1.797	1.3413		2.5314

3a.

D is dropped, as its popularity score is .5905 at the end of the stream.

3b.

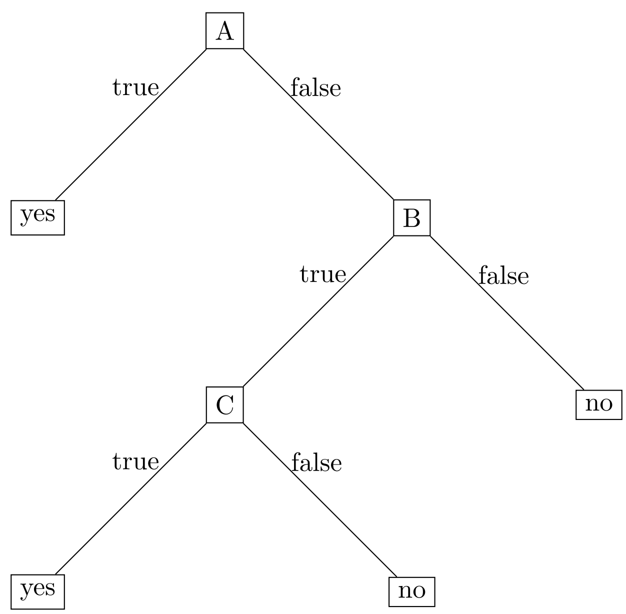
The most popular element is B, with a score of 2.5314.

Question 4

(3x +7) mod 11
X=1, 10 mod 11 = 10, 1010
x=2, 13 mod 11 = 2, 0010
X=3, 16 mod 11 = 5, 0101
X=4, 19 mod 11 = 8, 1000
X=5, 22 mod 11 = 0, 0000
X=6, 25 mod 11 = 3, 0011
X=7, 28 mod 11 = 6, 0110
X=8, 31 mod 11 = 9, 1001
X=9, 34 mod 11 = 1, 0001
X=10, 37 mod 11 = 4, 0100

With set {10 9 1 7}, 10 would have to be in the set since the estimate of the number of distinct elements is 2ˆr where r is the max tail length in the set.

Question 5



Question 6

6a.

$$H[12+,9-] = -\frac{12}{21}log_2\frac{12}{21} - \frac{9}{21}log_2\frac{9}{21} = -.5714 * -.8074 - .4286 * -1.2223 = .9852 \tag{3}$$

6b.

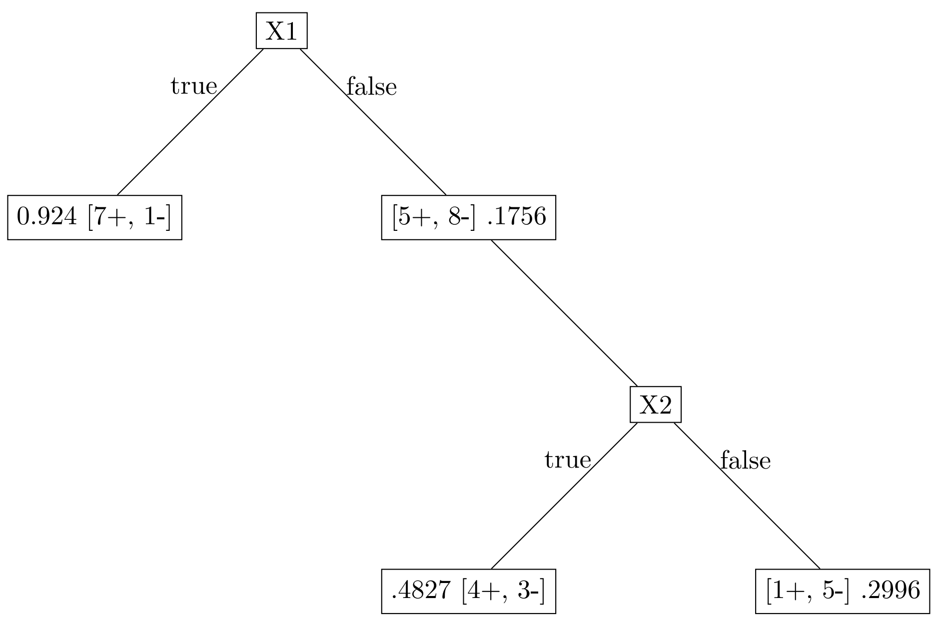
$$IG(X1) = T[7+,1-]F[5+,8-] \tag{4}$$

$$.9582 - \frac{8}{21} * 1.3925 - \frac{13}{21} * .6920 = .0265 \tag{5}$$

$$IG(X2) = T[7+,3-]F[5+,6-] \tag{6}$$

$$.9852 - \frac{10}{21} * 1.0704 - \frac{11}{21} * 9329 = .013 \tag{7}$$

6c.



6d.

$$\frac{4}{7} = .5714 \tag{8}$$

Question 7

Assuming that you've picked door number 1, there are three (equally likely) possible scenarios:

1. You pick the door with the prize, and the other two doors are empty.
2. Both your door and door number 2 are empty.
3. Both your door and door number 3 are empty.

Overall, there are two groups of possibilities - that you've picked a winning door (1/3) or you've picked an empty door (2/3).

If an empty door is revealed, it must either be door 2 or 3, because you picked door number 1. Now you are presented with the same two groups - but the second group's doors now have probabilities 0 and 2/3 of containing the prize, when originally they were 1/3 each.

Since your door (door number 1) has a probability of 1/3 and the other closed door has a probability of 2/3, you should switch doors.