CSIS734-01 Data Mining & Predictive Analytics
Garth Mortensen, mort0052@stthomas.edu

Graduate Program in Software
CSIS 734-01: Data Mining & Predictive Analytics
Assignment #7 (100 points)
Due Date: April 28th, 2018

Following table has the training data for our classification task.

ID	A	B	C	D	class
1	0	0	1	1	Y
2	0	0	1	1	Y
3	1	1	0	0	Y
4	1	1	0	0	Y
5	1	0	1	0	X
6	1	1	0	1	X
7	0	1	0	0	Y
8	0	1	0	0	Y
9	0	1	1	0	X
10	0	1	0	1	X
R1	1	0	1	1	?
R2	0	1	1	1	?

We want to predict the classes of the following two records:

R1 = {A = 1, B = 0, C = 1, D = 1}, and

R2 = {A = 0, B = 1, C =1 , D = 1}.

Use Naïve Bayes Theorem to determine the classes of the two records. Please include your calculations.

ID	A	B	C	D	Class	
1	0	1	0	1	N	
2	0	1	1	0	N	
3	1	0	1	0	N	
4	1	1	0	1	N	
5	0	0	1	1	Y	
6	0	0	1	1	Y	
7	0	1	0	0	Y	
8	0	1	0	0	Y	
9	1	1	0	0	Y	
10	1	1	0	0	Y	10 Total observations
R1	1	0	1	1	?	
R2	0	1	1	1	?	

Count 1	4	7	4	4	6 total Ys
Count 0	6	3	6	6	4 total Ns

Given d = (S, C, H, W) as a new instance.

Probability that the new instance d belong to class Yes.

P(Y | d) = (P(d | Yes) x P(Y)) / (P(d))

P(Y | d) = Posterior probability d belongs to class Yes, if d has attributes (sunny, cool, high humidity, windy)?

(P(d | Ye Likelihood our trainings have attributes like d (sunny, cool, high humidity, windy) & belong to class Yes?

P(Y) = Prior probability of trainings belong to class Yes in our training set.

P(d) = Predictor prior probability

r1	a	b	c	d
	1	0	1	1

Step 1 - Solve numerators

r1 Yes P(Yes |((2/6) * (2/6) * (2/6) * (2/6) * 6/10) / P(r1)
= 0.007 / P(r1)

r1 No P(No | ((2/4) * (1/4) * (2/4) * (2/4) * 4/10) / P(r1)
= 0.013 / P(r1)

Step 2 - Sum numerators

Sum 0.007 0.013 = 0.020

Step 3 - Normalization

P(Yes | d) = 0.007 /P(r1) * 1/sum = 0.372
P(No | d) = 0.013 /P(r1) * 1/sum = 0.628 greater

63% probability No. Class
R1 1 0 1 1 N

Assumption: Observations are independent and attributes are uncorrelated.

r2	a	b	c	d
	0	1	1	1

Step 1 - Solve numerators

r2 Yes P(Yes |((4/6) * (4/6) * (2/6) * (2/6) * (6/10)) / P(r2)
= 0.03 / P(r2)

r2 No P(No | ((2/4) * (3/4) * (2/4) * (2/4) * (4/10) / P(r2)
= 0.038 / P(r2)

Step 2 - Sum numerators

Sum 0.03 0.038 = 0.067

Step 3 - Normalization

P(Yes | d) = 0.030 /P(r1) * 1/sum = 0.441
P(No | d) = 0.038 /P(r1) * 1/sum = 0.559 greater

56% probability No. Class
R2 0 1 1 1 N

Assumption: Observations are independent and attributes are uncorrelated.