Problem 3: Naïve Bayes Classifiers

x1	x2	x3	x4	x5	y
know author?	is long?	has 'research'	has 'grade'	has 'lottery'	⇒ read?
0	0	1	1	0	-1
1	1	0	1	0	-1
0	1	1	1	1	-1
1	1	1	1	0	-1
0	1	0	0	0	-1
1	0	1	1	1	1
0	0	1	0	0	1
1	0	0	0	0	1
1	0	1	1	0	1
1	1	1	1	1	-1

1. COMPUTE ALL THE PROBABILITIES NECESSARY FOR A NAÏVE BAYES CLASSIFIER

THE CLASS PROBABILITY P(Y)

Y= => read?

	Count	P(y)
Y=1	4	P(y=1)=2/5
Y=-1	6	P(y=-1)=3/5
Total	10	

INDIVIDUAL FEATURE PROBABILITIES P(XI|Y)

X1 = know author?

	Y=1	Y=-1	P(x1|y=1)	P(x1|y=-1)	P(x1)
X1=0	1	3	P(x1=0|y=1) = 1/4	P(x1=0|y=-1) = 1/2	P(x1=0)=2/5
X1=1	3	3	P(x1=1|y=1) = 3/4	P(x1=1|y=-1) = 1/2	P(x1=1)=3/5
Total	4	6	1	1	1

X2=is long?

	Y=1	Y=-1	P(x2|y=1)	P(x2|y=-1)	P(x2)
X2=0	4	1	P(x2=0|y=1) = 1	P(x2=0|y=-1) = 1/6	P(x2=0)=1/2
X2=1	0	5	P(x2=1|y=1) = 0	P(x2=1|y=-1) = 5/6	P(x2=1)=1/2
Total	4	6	1	1	1

X3=has 'research'?

	Y=1	Y=-1	P(x3|y=1)	P(x3|y=-1)	P(x3)
X3=0	1	2	P(x3=0|y=1) = 1/4	P(x3=0|y=-1) = 1/3	P(x3=0)=3/10
X3=1	3	4	P(x3=1|y=1) = 3/4	P(x3=1|y=-1) = 2/3	P(x3=1)=7/10
Total	4	6	1	1	1

X4=has 'grade?

	Y=1	Y=-1	P(x4|y=1)	P(x4|y=-1)	P(x4)
X4=0	2	1	P(x4=0|y=1) = 1/2	P(x4=0|y=-1) = 1/6	P(x4=0)=3/10
X4=1	2	5	P(x4=1|y=1) = 1/2	P(x4=1|y=-1) = 5/6	P(x4=1)=7/10
Total	4	6	1	1	1

X5=has 'lottery?

	Y=1	Y=-1	P(x5|y=1)	P(x5|y=-1)	P(x5)

X5=0	3	4	P(x5=0|y=1) = 3/4	P(x5=0|y=-1) = 2/3	P(x5=0)=7/10
X5=1	1	2	P(x5=1|y=1) = 1/4	P(x5=1|y=-1) = 1/3	P(x5=1)=3/10
Total	4	6	1	1	1

2. x = (0 0 0 0 0) and x = (1 1 0 1 0)

Naïve Bayes theorem applied to the above data gives us:

$$P(y|x1,x2,x3,x4,x5) = \frac{P(x1|y)P(x2|y)P(x3|y)P(x4|y)P(x5|y)P(y)}{P(x1,x2,x3,x4,x5)}$$

Since we need to compare between $y = 0$, $y = 1$ and the denominator is the same, we can ignore it.

x = (0 0 0 0 0)

$P(y = 1|x1 = 0, x2 = 0, x3 = 0, x4 = 0, x5 = 0)$

$= P(x1 = 0|y = 1)P(x2 = 0|y = 1)P(x3 = 0|y = 1)P(x4 = 0|y = 1)P(x5 = 0|y = 1)P(y = 1)$

$= 0.009375$ --- **(1)**

And,

$P(y = -1|x1 = 0, x2 = 0, x3 = 0, x4 = 0, x5 = 0)$

$= P(x1 = 0|y = -1)P(x2 = 0|y = -1)P(x3 = 0|y = -1)P(x4 = 0|y = -1)P(x5 = 0|y = -1)P(y = -1)$

$$= \frac{1}{2} * \frac{1}{6} * \frac{1}{3} * \frac{1}{6} * \frac{2}{3} * \frac{3}{5}$$

$= 0.00185$ --- **(2)**

From (1) and (2), it is clear that $P(y = 1|x1 = 0, x2 = 0, x3 = 0, x4 = 0, x5 = 0) >$
$P(y = -1|x1 = 0, x2 = 0, x3 = 0, x4 = 0, x5 = 0)$

Hence, the class the class $Y = 1$ would be predicted. i.e. "Read"

x = (1 1 0 1 0)

$P(y = 1|x1 = 1, x2 = 1, x3 = 0, x4 = 1, x5 = 0)$

$= P(x1 = 1|y = 1)P(x2 = 1|y = 1)P(x3 = 0|y = 1)P(x4 = 1|y = 1)P(x5 = 0|y = 1)P(y = 1)$

$$= \frac{3}{4} * 0 * \frac{1}{4} * \frac{1}{2} * \frac{3}{4} * \frac{2}{5}$$

$= 0$ --- **(1)**

And,

$P(y = -1|x1 = 1, x2 = 1, x3 = 0, x4 = 1, x5 = 0)$

$= P(x1 = 1|y = -1)P(x2 = 1|y = -1)P(x3 = 0|y = -1)P(x4 = 1|y = -1)P(x5 = 0|y = -1)P(y = -1)$

$$= \frac{1}{2} * \frac{5}{6} * \frac{1}{3} * \frac{5}{6} * \frac{2}{3} * \frac{3}{5}$$

$= 0.0463$ --- **(2)**

From (1) and (2), it is clear that $P(y = -1|x1 = 1, x2 = 1, x3 = 0, x4 = 1, x5 = 0) > P(y = 1|x1 = 1, x2 = 1, x3 = 0, x4 = 1, x5 = 0)$

Hence, the class the class $Y = -1$ would be predicted. i.e. "Discard"

3. POSTERIOR PROBABILITY THAT Y = +1 GIVEN THE OBSERVATION X = (1 1 0 1 0).

From above (2), it is clear that

$P(y = 1|x1 = 1, x2 = 1, x3 = 0, x4 = 1, x5 = 0) = 0$
(Since the numerator is 0, we can be sure that the posterior probability is 0 irrespective of the denominator)

4. "JOINT" BAYES CLASSIFIER

Since the data is not complete, i.e. we do not have the complete joint distribution of input data across 5 features (2^5=32 data points required). Hence, the joint Bayes classifier might assign a zero probability to new data that does not exist in the training data. For example X=(0, 0, 0, 0, 0)

5. Losing access to the address book

Since the Individual feature probabilities and the class probability will remain the same for the underlying data, we do not need to retrain the data. We can discard the P(x1/y) and P(x1) values. However, we still need to change the classifier to use the new formula to predict the class. The new Naïve Bayes formula would be:

$$P(y|x2, x3, x4, x5) = \frac{P(x2|y)P(x3|y)P(x4|y)P(x5|y)P(y)}{P(x2, x3, x4, x5)}$$

Since we do not have x1 values, we shall ignore them in predicting the class.

Statement of Collaboration

I have not collaborated with anyone for this homework and have maintained the UCI code of honesty.