HW 2 Theory May 2018

1

a.)
First we write a table over our training set:

	rich		married		healthy		
	Yes	No	Yes	No	Yes	No	total
c=1	3	1	2	2	3	1	4
c=0	1	3	1	3	1	3	4
total	4	4	3	5	4	4	8

And we see our Prior probabilites:
$P(c = 1) = 0.5 \quad (4/8)$
$P(c = 0) = 0.5 \quad (4/8)$

Then we find our evidence probabilites:
$P(rich) = P(!rich) = 0.5$
$P(married) = 0.375$
$P(healthy) = P(!healthy) = 0.5$

Then we find our likelihood probabilities for model (0,1,1):
$P(!rich|c = 1) = 0.25$
$P(married|c = 1) = 0.5$
$P(healthy|c = 1) = 0.75$

Now we want to calculate:

$$P(c = 1|!rich, married, healthy) = \frac{P(!rich|c = 1)P(married|c = 1)P(healthy|c = 1)P(c = 1)}{P(!rich)P(married)P(healthy)}$$

and

$$P(c = 0|!rich, married, healthy) = \frac{P(!rich|c = 0)P(married|c = 0)P(healthy|c = 0)P(c = 0)}{P(!rich)P(married)P(healthy)}$$

This is:

$$P(c=1|!rich, married, healthy) = \frac{0.25 * 0.5 * 0.75 * 0.5}{P(evidence)} = \frac{0.046875}{P(evidence)}$$

and:

$$P(c=0|!rich, married, healthy) = \frac{0.75 * 0.25 * 0.25 * 0.5}{P(evidence)} = \frac{0.0234375}{P(evidence)}$$

this give the normalized probability:

$$P(c=1|!rich, married, healthy) = \frac{0.046875}{0.046875 + 0.0234375} = 0.6667$$

Answer: The probability that a person who is not rich, married and healthy is "content" is 0.6667.

b.)

We want to calculate:

$$P(c=1|!rich, married) = \frac{0.25 * 0.5 * 0.5}{P(evidence)} = \frac{0.0625}{P(evidence)}$$
$$P(c=0|!rich, married) = \frac{0.75 * 0.25 * 0.5}{P(evidence)} = \frac{0.09375}{P(evidence)}$$

this give the normalized probability:

$$P(c=1|!rich, married) = \frac{0.0625}{0.0625 + 0.09375} = 0.4$$

Answer: The pobability that a person who is not rich and married is "content" is 0.4.

2

Note that multiple answers for this question is acceptable, one of them is as follows:

The Naive Bayes Assumption tells us that the x_is must be conditionally independent given y. We clearly see that the variables x_1, x_2 and x_3 are dependent. One of the variables must be one and the two other must be zero, i.e. You can't be under 20 and over 30 at the same time.

One way to solve this is by merge the vectors x_1, x_2 and x_3 to one vector x_1':
$x_1' = -1$ if costumer is under 20.
$x_1' = 0$ if customer is between 20 and 30.
$x_1' = 1$ if customer is over 30.