Assignment - Week-3

Bayesian Classification

1. Given study of pleas & prison sentences,

$P(\text{prison}) = 0.42$

$P(\text{guilty} \mid \text{prison}) = 0.38$

$P(\text{guilty} \mid \overline{\text{prison}}) = 0.5$

a) Probability of a person who was not sent to the prison

i.e., $P(\overline{\text{prison}})$

$$P(\overline{\text{prison}}) = 1 - P(\text{prison})$$
$$= 1 - 0.42$$
$$= 0.58$$

$$\boxed{P(\overline{\text{prison}}) = 0.58}$$

b) given subject entered a guilty plea, what is the probability that subject was not sent to prison.

i.e., $P(\overline{\text{prison}} \mid \text{Guilty})$ (probabity of not sent to prison given that the subject pleaded guilty).

According to Bayes theorem

$$P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X)}$$

Similarly,

$$P(\overline{\text{Prison}} \mid \text{Guilty}) = \frac{P(\text{Guilty} \mid \overline{\text{prison}}) \cdot P(\overline{\text{prison}})}{P(\text{Guilty})}$$

based on 1(a) $P(\overline{\text{prison}}) = ~~0.58~~ \; 0.58$

Given $P(\text{Guilty} \mid \overline{\text{prison}}) = 0.5$

we need to calculate,

$$P(\text{Guilty}) = P(\text{Guilty}|\text{prison}) \cdot P(\text{prison}) + P(\text{Guilty}|\overline{\text{prison}}) \cdot P(\overline{\text{prison}})$$

$$P(\text{Guilty}) = (0.38 \times 0.42) + (0.5 \times 0.58)$$

$$= 0.1596 + 0.29$$

$$= 0.4496.$$

using $P(\text{Guilty})$ in our actual equation

$$P(\overline{\text{prison}} \mid \text{Guilty}) = \frac{0.5 \times 0.58}{0.4496}$$

$$\boxed{P(\overline{\text{prison}}|\text{Guilty}) = 0.645}$$

1c) probability of getting some one sent to prison

$$\boxed{P(\text{prison}) = 0.42}$$

directly given in question

1d) Subject entered a guilty plea, probability of ~~not~~ sent to prison

i.e., $P(\text{~~prison~~}|$

$$P(\text{prison}|\text{Guilty}) = \frac{P(\text{Guilty}|\text{prison}) \cdot P(\text{prison})}{P(\text{Guilty})}$$

from 1(b) $P(\text{Guilty}) = 0.4496$

from question $P(\text{prison}) = 0.42$,

$P(\text{Guilty}|\text{prison}) = 0.38$

Substituting the values;

$$P(\text{prison}|\text{Guilty}) = \frac{0.38 \times 0.42}{0.4496}$$

$$\boxed{P(\text{prison}|\text{Guilty}) = 0.355}$$

This can also be calculated by $1 - P(\overline{\text{prison}}|\text{Guilty})$

$= 1 - 0.645 = \boxed{0.355}$

2) Given table, (attached)

we have 20 data points with 4 features.

i.e., Gender 2-categories (M, F)

Car type 3- categories (Family, sports, luxury)

Shirt size 4- categories (small, medium, large, extralarge)

classes 2-categories (C0, C1)

a) P(Gender = M | class = C0)

This can be obtained using frequency table & likelihood table. (using data table)

frequency:

Gender	Class C0	Class C1
M	6	4
F	4	6

Probability table

Gender	C0	C1
M	6/10	4/10
F	4/10	6/10

P(Gender=M | class C0 = 6/10

= 0.6

Customer ID	Gender	Car Type	Shirt Size	Class
1	M	Family	Small	C0
2	M	Sports	Medium	C0
3	M	Sports	Medium	C0
4	M	Sports	Large	C0
5	M	Sports	Extra Large	C0
6	M	Sports	Extra Large	C0
7	F	Sports	Small	C0
8	F	Sports	Small	C0
9	F	Sports	Medium	C0
10	F	Luxury	Large	C0
11	M	Family	Large	C1
12	M	Family	Extra Large	C1
13	M	Family	Medium	C1
14	M	Luxury	Extra Large	C1
15	F	Luxury	Small	C1
16	F	Luxury	Small	C1
17	F	Luxury	Medium	C1
18	F	Luxury	Medium	C1
19	F	Luxury	Medium	C1
20	F	Luxury	Large	C1

$P(\text{Gender} = F \mid \text{class} = C_1)$

Using the same probability table

$$\Rightarrow \boxed{P(F|C_1) = 6/10 = 0.6}$$

- $P(\text{Car typ} = \text{family} \mid \text{class} = C_0)$

this can be obtained by the frequency & probabity table of car type vs class

frequency:

Car type	C_0	C_1
family	1	3
sports	8	0
luxury	1	7

Probability

car type	C_0	C_1
family	1/10	3/10
sports	8/10	0/10
luxury	1/10	7/10

$$\boxed{P(\text{family} \mid C_0) = 1/10 = 0.1}$$

$P(\text{Car type} = \text{family} \mid \text{class} = C_1)$

$\Rightarrow$ Using the previous probability table

$$P(\text{Family} \mid C_1) = 3/10 = 0.3$$

$\therefore$ $P(\text{shirt size} = \text{Medium} \mid \text{class} = C_0)$

this can be obtained using class vs shirt size frequency & Probability table.

Shirt size	C_0	C_1
Small	3	2
Medium	3	4
Large	2	2
Extra large	2	2

Probability

shirt size	C_0	C_1
Small	3/10	2/10
medium	3/10	4/10
Large	2/10	2/10
Exlra large	2/10	2/10

based on this probability table

$$P(\text{shirt size} = \text{medium} / \text{class} = C_0)$$

$$= 3/10 = 0.3$$

$$\Rightarrow P(\text{shirt size} = \text{Medium} / \text{class} = C_1)$$

$$= 4/10 = 0.4$$

2 (b) class of

$$P(\text{Gender} = F / \text{cartyp} = \text{family} / \text{shirt size} = \text{Medium})$$

here we need to clasify if gender F, cartype family & shirt size medium is class C_0 & C_1 using Naive bayes

$$P(\text{class} / \text{gender} = F, \text{cartype} = \text{family}, \text{shirt} = \text{medium})$$

$$\propto P(\text{gender} = F / \text{class}) \cdot P(\text{cartye} = \text{family} / \text{class}) * P(\text{shirt} = \text{medium} / \text{class})$$

we have 10 data point C_1 & 10 datapoint C_0

$$P(\text{Class} = C_0) = P(\text{class} = C_1) = 0.5$$

from our probability table we can calculate probabilites of each class against each feature

$P(\text{Gender} = F \mid \text{class} = C_0) = 4/10 = 0.4$

$P(\text{Gender} = F \mid \text{class} = C_1) = 6/10 = 0.6$

$P(\text{Cartype} = \text{family} \mid \text{class} = C_0) = 1/10 = 0.1$

$P(\text{Cartype} = \text{family} \mid \text{class} = C_1) = 3/10 = 0.3$

$P(\text{shirttype} = \text{medium} \mid \text{class } C_0) = 3/10 = 0.3$

$P(\text{shirt type} = \text{medium} \mid \text{class } C_1) = 4/10 = 0.4$

$P(\text{class} = C_0 \mid \text{Gender} = F, \text{Cartype} = \text{family}, \text{shirt type} = \text{medium})$

$= P(\text{Gender} = F \mid \text{class} = C_0) \cdot P(\text{Cartype} = \text{family} \mid \text{class } C_0) \cdot P(\text{shirt} = \text{medium} \mid \text{class } C_0)$

$= 0.4 \times 0.1 \times 0.3$

$= 0.012$

$\boxed{P(\text{class} = C_0) = 0.012}$

Ily for class C_1

$P(\text{class} = C_1 \mid F, \text{family}, \text{medium})$

$= P(\text{Gender} = F \mid \text{class} = C_1) \cdot P(\text{Cartype} = \text{family} \mid \text{class } C_1) \cdot P(\text{Shirt} = \text{medium} \mid \text{class } C_1)$

$= 0.6 \times 0.3 \times 0.4$

$P(C_1) = 0.072$

as $P(\text{class } C_1) > P(\text{class } C_0)$

$0.072 > 0.012$

it is more likely belong to class 'C1'

thus the

class of $P(\text{Gender} = F \mid \text{cartype} = \text{family} \mid \text{shirtsize} = \text{medium})$

is **'C1'**