Sai Parthish. M
1002022847

Assignment - 4.

Task 1 :-

Given joint probability distribution for a domain of two variables

	Color = Red	Color = Green	Color = Blue
Vechile = Car	0.1184	0.1280	0.0736
Vechile = van	0.0444	0.0480	0.0276
Vechile = Truck	0.1554	0.1680	0.0966
Vechile = SUV	0.0518	0.0560	0.0322

Sol :- P(A and B) = P(A)*P(B)
where P(A) = colour is green and
P(B) = Vehicle is truck.
from Baye's theorem,
P(A) = 0.1280 + 0.0480 + 0.1680 + 0.0560
= 0.4
add all green color values =, and then
$\neg P(A) = 1 - P(A) = 1 - 0.4 = 0.6$
$P(B) = 1/4 = 0.25$
For calculating given condition

$$\frac{P(\text{Color is not } \cancel{\text{green}} \text{ green} \wedge \text{Vechile is Truck})}{P(\text{Vechile is Truck})}$$

$$\Rightarrow \frac{0.1554 + 0.0966}{0.1554 + 0.1680 + 0.0966} =$$

$$\Rightarrow \frac{0.252}{0.42} = 0.6$$

1002022847

Part b:

Check if color are totally independent from each other

P(color is green) = 1 - P(color is green)
P(color is green) = 0.4.
P(color is not green) = 1 - 0.4 = 0.6

P(color is not green) = P(Color is not given / vechile is truck)

if the color is not green and P(color is not given / vechile is truck) then they are totally independent of each other.

Task 2:-

given In, a certain probability problem we have 11 variables A B1, B2 . . . . , B10. Variables has 7 Rules and Each of Variable B1 . . B10 have 8 possible values. Given that Each Bi is Conditionally Independent of all other 9 Bj variables (with j! = i) given A.

1002022847

Part a:-

Given, 11 variables : A, B, B_2 . . . , B_{10}
A has 7 values
B_1 to B_{10} has 8 possible values, each B_i is
conditionally independent
possible . of A = 7
possible value of B = 8^{10}

7×8^{10} is the total numbers to be
stored in joint distribution table = 7×8^{10}.

Part b:-

The most space-efficient way
of representation for that joint probability
distribution of these 11 are
$P(B/A) = 7 \times 8 = 56$ values or
~~for~~ $7 \times 7 = 49$ values
for the 10 variables: $49 \times 10 = 490$
we need to calculate $P(A) = (7-1) = 6$
Total Space = $490 + 6 = 496$.

Part c:-

Yes, this scenario follow
the Naive-Bayes model.

1002022847

Task-4

given table

Class	A	B	C
X	1	2	1
X	2	1	2
X	3	2	2
X	1	3	3
X	1	2	1
Y	2	1	2
Y	3	1	1
Y	2	2	2
Y	3	3	1
Y	2	1	1

The ~~information~~

Entropy before split $x=5, y=5$

and splitting with A

for $A=1$

$x=3, y=0$

$Ha = -\frac{3}{3} \log_2 3/3 - 0/3 \log_2 0/3 = 0$

for $A=2=$ $x=1, y=3$ $Hb = 0.8113$

Hc -- $A=3 = 0.9183$ $\therefore I_A = \underline{\underline{0.4}}$

Splitting it with B.

For B=1 $H_d = 1/4 \log_2 1/4 - 3/4 \log_2 3/4$

$= 0.8113$

for B=2

$X = 3, Y = 1$

For B=3, ~~He~~.

$H_f = -1/2 \log 1/2 - 1/2 \log 1/2 = 1$

$I_B = H - (4/10\, H_d - 4/10\, H_c - \frac{2}{10} H_f$

0.151 //

Splitting with C.

for C=1 , $X = 2, Y = 3$

$H_g = -2/5 \log 2/5 - 3/5 \log 3/5$

$= 0.971$

for C=2 $X = 2, Y = 2$

$H_h = -1/1 \log 1/1 - 0/1 \log 0/1 = 0$

$I_C = H - 5/10\, H_g - 4/10\, H_h - 1/10\, H_i$

$= 0.1145 \approx 0.115$

Therefore, A is the best attribute.

1002022847

Task-5

Class	A	B	C
X	25	24	31
X	22	14	24
X	28	22	25
X	24	13	30
X	26	20	24
Y	20	31	17
Y	18	32	14
Y	21	25	20
Y	13	32	15
Y	12	27	18

gain = Parent node, where
The gine looks as $-1/2 \log 1/2$
$- 1/2 \log 1/2$
$\Rightarrow 1$

for the treshold is 15.
$-0/2 \log 0/2 - 2/2 \log 2/2 = 0$

if theshold is 20 $\Rightarrow$ 0

if treshold is 25 $\Rightarrow$ 0.286

Grain at A $= 1 - (0.286 + 0 + 0)$
$= 0.713$

Case B:-

if treshold is 15 $\Rightarrow -2/2 \log 2/2 - 0/2 \log 0/2$
$= 0$

If threshold is 1.0 => $-3/3 \log 3/3 - 0/3 \log 0/3$
$= 0$

If threshold is 15 => $-5/6 \log 5/6 - 1/6 \log 1/6$
$= 0.065 + 0.124$
$= 0.199$

at $B = 1 - (0 + 0 + 0.195)$
$= 0.805$

Case (c) :- at 'c'
If threshold is 15 => $-0/2 \log 0/2 - 2/2 \log 2/2 = 0$
If threshold is 20 => $-0/5 \log 0/5 - 5/5 \log 5/5 = 0$
If threshold is 25 => $-3/8 \log 3/8 - 5/8 \log 5/8$
~~gain at C = 1 - (0+~~ $= 0.286$
gain at C :- $1 - (0 + 0 + 0.286)$
$= 0.713$
Attribute "B" achieves the highest information gain at the root.