Sudharsan Srinivasan
1001755919

1.

a. We need 3 matrices namely Nominal, ordinal and Numeric. Another dij matrix to find distance.

b. Nominal Matrix: (dissimilarity)

	1	2	3	4	5
1	0				
2	1	0			
3	1	1	0		
4	0	1	1	0	
5	1	0	1	1	0

Condition is

$$d = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y \end{cases}$$

Ordinal Matrix → Assign values to ordinal data
Fair = 1, Good = 2.
Excellent = 3

	1	2	3	4	5
1	0				
2	1	0			
3	0.5	0.5	0		
4	0	1	0.5	0	
5	1	0	0.5	1	0

Condition:

$$d = \frac{|x-y|}{n-1}$$

$n = 3$
$n-1 = 2$

$$d = \frac{|x-y|}{2}$$

Numeric Matrix:

Min = 20
Max = 64
diff = 64 − 20 = 44

Sudharsan Srinivasan
1001755919

	1	2	3	4	5
1	0				
2	0.522	0			
3	0.431	0.954	0		
4	0.159	0.363	0.590	0	
5	0.568	0.045	1	0.049	0

$d(2,1) = \frac{|22-45|}{44} = 0.522$

$d(3,1) = \frac{|64-45|}{44} = 0.431$

$d(4,1) = \frac{|38-45|}{44} = 0.159$

$d(5,1) = \frac{|20-45|}{44} = 0.568$

$d(3,2) = \frac{|64-22|}{44} = 0.954$

$d(4,2) = \frac{|38-22|}{44} = 0.363$, $d(4,3) = \frac{|38-64|}{44} = 0.590$

$d(5,2) = \frac{|20-22|}{44} = 0.045$ $d(5,3) = \frac{|20-64|}{44} = 1$

$d(5,4) = 0.049$

c. d_{ij}

	1	2	3	4	5
1	0				
2	0.840	0			
3	0.643	0.818	0		
4	0.053	0.7876	0.6967	0	
5	0.856	0.015	0.823	0.803	0

$d(1,1) = 0$

$d(2,1) = (1+1+0.522)/3 = 0.840$

$d(3,1) = (1+0.5+0.431)/3 = 0.643$

$d(3,2) = (1+0.5+0.954)/3 = 0.818$

Sudharsan Srinivasan
1001755919

$d(4,1) = (0 + 0.154)/3 = 0.053$

$d(4,2) = (2 + 0.363)/3 = 0.7876$

$d(4,3) = (1 + 0.5 + 0.590)/3 = 0.6967$

$d(5,1) = (1 + 1 + 0.568)/3 = 0.856$

$d(5,2) = (0.045)/3 = 0.015$

The value of $d(1,3) = 0.643$

Sudharsan Srinivasan
1001755919

2. Gini Index:

Gini Index $1 - \sum P_i^2$

Age Attribute :- consider mean as 33

Two groups

≤ 33 and > 33

		High	Low	Total
Age:	≤ 33	3	1	4
	> 33	1	1	2

Gini Index (≤ 33) $= 1 - (3/4)^2 - (1/4)^2 = 6/16$

Gini Index (> 33) $= 1 - (1/2)^2 - (1/2)^2 = 1/2$

Gini Index (Age) $= (4/6)(6/16) + (1/6)(1/2) = 5/12$

Car type Attribute

		High	Low	Total
Car type	family	2	1	3
	Sports	2	0	2
	Truck	0	1	1

Gini (Family) $= 1 - (2/3)^2 - (1/3)^2 = 4/9$

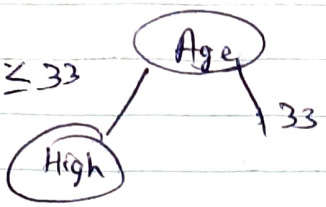
Gini (Sports) $= 1 - (2/2)^2 = 0$

Gini (Truck) $= 1 - 1 = 0$

Gini (Car Type) $= 3/6\ (4/9) + 0 + 0 = 2/3$

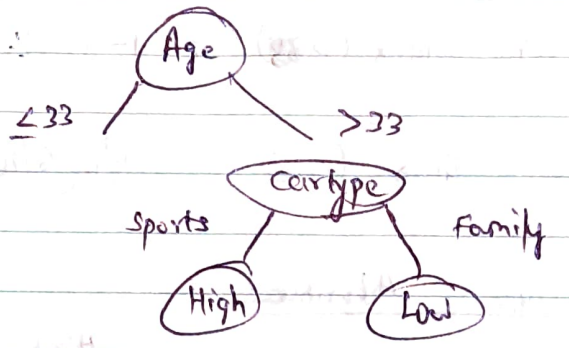
Sudharsan Srinivasan
1001755914

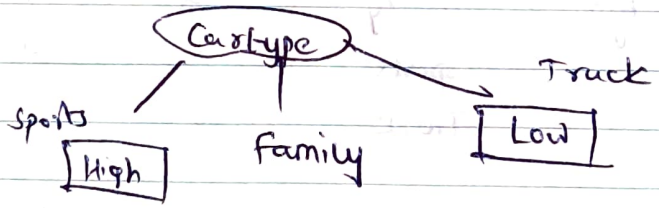
Since Gini for both attributes are different we consider Age as Root



Data for > 33

Tid	Age	Cartype	Class
2	43	Sports	High
3	68	family	Low

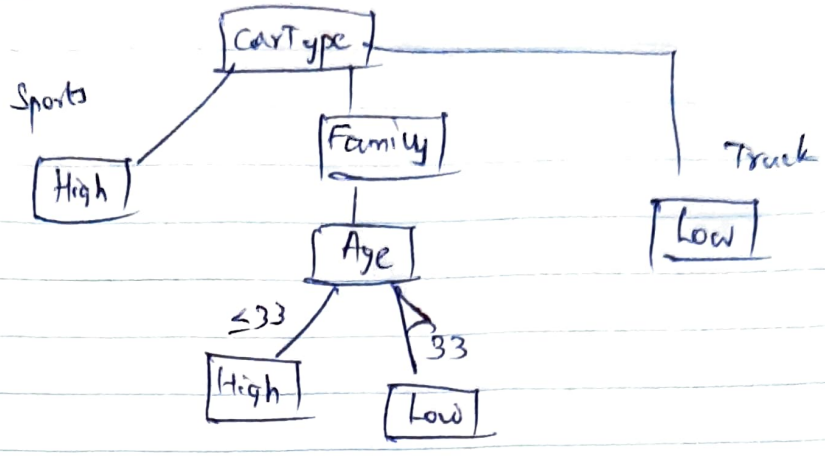




Data for family

Tid	Age	CarType	Class
0	23	Family	High
3	68	family	Low
5	20	family	High

Sudharsan Srinivasan
1001751919



(2) Type of Attributes

Tid → Nominal
Age → Ratio
Cartype → Ordinal

(3) Stop splitting if every node has the class attribute values

(4) age = 40 cartype = family => class = Low

Sudharsan Srinivasan
1001755919

3. Bayes predict class for x_4

Given $z = \{x_4, \text{excellent}, 4\}$

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

(i) $$P(x_4|h_1) = \frac{P(h_1|x_4)\cdot P(x_4)}{P(h_1)} = \frac{1 \times 1/10}{6/10} = 1/6$$

$$P(\text{Excellent}|h_1) = \frac{P(h_1|\text{excellent})\,P(\text{Excellent})}{P(h_1)}$$

$$= \frac{(1)(3/10)}{6/10} = 1/2$$

$$P(4|h_1) = \frac{P(h1|4)\,P(4)}{P(h_1)} = \frac{(1)\cdot 2/10}{6/10} = 1/3$$

$$\boxed{P(z|h_1) = 1/6 \times 1/2 \times 1/3 = 1/36} \quad \text{— (1)}$$

(ii) for class h_2 $z = \{x_4, \text{excellent}, 4\}$

$$P(x_4|h_2) = \frac{(0)\ (1/10)}{2/10} = 0$$

$$P(\text{Excellent}|h_2) = \frac{P(h_2|\text{Excellent})\,P(\text{Excellent})}{P(h_2)} = \frac{0(3/10)}{2/10} = 0$$

$$P(4|h_2) = \frac{P(h_2|4)\,P(4)}{P(h_2)} = \frac{(0)\ (2/10)}{(2/10)} = 0$$

$$\boxed{P(z|h_2) = 0} \quad \text{— (2)}$$

Sudharsan Srinivasan
1001755919

(iii) for class h_3 $z = \{x_4, \text{excellent}, 4\}$

$$P(x_4 | h_3) = \frac{P(h_3 | x_4)\, P(x_4)}{P(h_3)} = \frac{0 \times 1/10}{1/10} = 0$$

$$P(\text{Excellent} | h_3) = \frac{0 \times 3/10}{1/10} = 0$$

$$P(4 | h_3) = \frac{0 \times 2/10}{1/10} = 0$$

$$\therefore \boxed{P(z | h_3) = 0} \quad \text{—— (3)}$$

(iv) for class h_4 $z = \{x_4, \text{excellent}, 4\}$

$$P(x_4 | h_4) = \frac{P(h_4 | x_4)\, P(x_4)}{P(h_4)} = \frac{0 \times 1/10}{1/10} = 0$$

$$P(\text{Excellent} | h_4) = \frac{0 \times 3/10}{1/10} = 0$$

$$P(4 | h_4) = \frac{0 \times 2/10}{1/10} = 0$$

$$\therefore \boxed{P(z | h_4) = 0} \quad \text{—— (4)}$$

from ①, ②, ③ & ④

$$P(z | h_1) = 4/36$$
$$P(z | h_2) = 0$$
$$P(z | h_3) = 0$$
$$P(z | h_4) = 0$$

Sudharsan Srinivasan
1001755919

4. KNN with K=5 & Euclidean Distance

From the dataset, find the euclidean distance for all

formula = $\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$

1. $P_1(158, 58)$ $P_2(161, 61)$
E.D. = $\sqrt{(161-158)^2+(61-58)^2}$ = $\sqrt{18}$ = 4.2464

2. $P_1(158, 59)$ $P_2(161, 61)$
E.D = $\sqrt{(161-158)^2+(61-59)^2}$ = $\sqrt{13}$ = 3.6055

3. $P_1(158, 63)$ $P_2(161, 61)$
E.D = $\sqrt{(161-158)^2+(61-63)^2}$ = $\sqrt{13}$ = 3.6055

4. $P_1(160, 59)$ $P_2(161, 61)$
E.D. = $\sqrt{(161-160)^2+(61-59)^2}$ = $\sqrt{5}$ = 2.2360

5. $P_1(160, 60)$ $P_2(161, 61)$
E.D = $\sqrt{(161-160)^2+(61-60)^2}$ = $\sqrt{2}$ = 1.4142

6. $P_1(163, 60)$ $P_2(161, 61)$
E.D. = $\sqrt{(161-163)^2+(61-60)^2}$ = $\sqrt{5}$ = 2.2360

7. $P_1(163, 61)$ $P_2(161, 61)$
E.D. = $\sqrt{(161-163)^2+(61-61)^2}$ = $\sqrt{4}$ = 2

8. $P_1(160, 64)$ $P_2(161, 61)$
E.D. = $\sqrt{(161-160)^2+(61-64)^2}$ = $\sqrt{10}$ = 3.1622

Sudharsan Srinivasan
1001755919

9. $P_1(163, 64)$ $P_2(161, 61)$

$E.D. = \sqrt{(161-163)^2 + (61-64)^2} = \sqrt{13} = 3.6055$

10. $P_1(165, 61)$ $P_2(161, 61)$

$E.D = \sqrt{(161-165)^2 + (61-61)^2} = \sqrt{16} = 4$

11. $P_1(165, 62)$ $P_2(161, 61)$

$E.D. = \sqrt{(161-165)^2 + (61-62)^2} = \sqrt{17} = 4.1231$

12. $P_1(165, 65)$ $P_2(161, 61)$

$E.D. = \sqrt{(161-165)^2 + (61-65)^2} = \sqrt{32} = 5.6568$

13. $P_1(168, 62)$ $P_2(161, 61)$

$E.D. = \sqrt{(161-168)^2 + (61-62)^2} = \sqrt{50} = 7.071$

14. $P_1(168, 63)$ $P_2(161, 61)$

$E.D. = \sqrt{(161-168)^2 + (61-63)^2} = \sqrt{53} = 7.2801$

15. $P_1(168, 66)$ $P_2(161, 61)$

$E.D. = \sqrt{(161-168)^2 + (61-66)^2} = \sqrt{74} = 8.6023$

16. $P_1(170, 63)$ $P_2(161, 61)$

$E.D. = \sqrt{(161-170)^2 + (61-63)^2} = \sqrt{85} = 9.2195$

17. $P_1(170, 64)$ $P_2(161, 61)$

$E.D = \sqrt{(161-170)^2 + (61-64)^2} = \sqrt{90} = 9.4868$

18. $P_1(170, 68)$ $P_2(161, 61)$

$E.D. = \sqrt{(161-170)^2 + (61-68)^2} = \sqrt{133} = 11.4017$

Sudharsan Srinivasan

1001755919

Here $k=5$, we take 5 points close to $(161, 61)$

Points	Distance	T-shirt Size
P(160, 60)	1.4142	M
P(163, 61)	2	M
P(160, 59)	2.23606	M
P(163, 60)	2.23606	M
P(160, 64)	3.1622	L

Here we have 4M and 1L

$\therefore$ $(161, 61) \rightarrow$ M (T-shirt size)