Computer Science and Engineering Discipline

Assignment 2

Course Number: CSE 4221
Course Title: Pattern Recognition

Supervised by,

Dr. S. M. Mohidul Islam
Associate Professor,
Computer Science and Engineering Discipline,
Khulna University.

Submitted By,

Hafsa Sultana
Student ID: 170220
Year: 4th, Term = 2nd
CSE Discipline
Khulna University

17. Given the following labeled samples:

Sl	x	y	Class
1.	2.491	2.176	ω_1
2.	1.053	0.677	ω_1
3.	5.792	3.425	ω_1
4.	2.054	-1.467	ω_1
5.	0.550	4.020	ω_1
6.	4.218	-2.075	ω_2
7.	-1.156	-2.992	ω_2
8.	-4.435	1.408	ω_2
9.	-1.794	-2.838	ω_2
10.	-2.137	-2.473	ω_2
11.	-2.520	0.483	ω_3
12.	-1.163	3.161	ω_3
13.	-13.438	2.414	ω_3
14.	-4.467	2.298	ω_3
15.	-3.711	4.364	ω_3

(v) Create a Decision Tree model for the dataset and what class would be assigned to the feature vector (-2.799, 0.746)?
Using (i) ID3 algorithm (ii) C4.5 algorithm (iii) CART algorithm

Solution:

I. Using **ID3 algorithm**:

Dataset, sorted by X feature-

Sl	x	y	Class
13.	-13.438	2.414	ω_3
14.	-4.467	2.298	ω_3
8.	-4.435	1.408	ω_2
15.	-3.711	4.364	ω_3
11.	-2.520	0.483	ω_3

10.	-2.137	-2.473	ω_2
9.	-1.794	-2.838	ω_2
12.	-1.163	3.161	ω_3
7.	-1.156	-2.992	ω_2
5.	0.550	4.020	ω_1
2.	1.053	0.677	ω_1
4.	2.054	-1.467	ω_1
1.	2.491	2.176	ω_1
6.	4.218	-2.075	ω_2
3.	5.792	3.425	ω_1

The boundary value between each of these pairs is simply the average of their X values:

1. $\frac{d_{14} + d_8}{2} = \frac{-4.467 + (-4.435)}{2} = -4.451$

2. $\frac{d_8 + d_{15}}{2} = \frac{-4.435 + (-3.711)}{2} = -4.073$

3. $\frac{d_{11} + d_{10}}{2} = \frac{-2.520 + (-2.137)}{2} = -2.3285$

4. $\frac{d_9 + d_{12}}{2} = \frac{-1.794 + (-1.163)}{2} = -1.4785$

5. $\frac{d_{12} + d_7}{2} = \frac{-1.163 + (-1.156)}{2} = -1.1595$

6. $\frac{d_7 + d_5}{2} = \frac{-1.156 + 0.550}{2} = -0.303$

7. $\frac{d_1 + d_6}{2} = \frac{2.491 + 4.218}{2} = 3.3545$

8. $\frac{d_6 + d_3}{2} = \frac{4.218 + 5.792}{2} = 5.005$

$$Entropy(Class) = -p(\omega_1)log_2(p(\omega_1)) - p(\omega_2)log_2(p(\omega_2)) - p(\omega_3)log_2(p(\omega_3))$$

$$= -\left(\tfrac{5}{15}\right)log_2\left(\tfrac{5}{15}\right) - \left(\tfrac{5}{15}\right)log_2\left(\tfrac{5}{15}\right) - \left(\tfrac{5}{15}\right)log_2\left(\tfrac{5}{15}\right)$$

$$= 1.585$$

1. **For boundary value -4.451**

$$Entropy(Class|X < -4.451) = -\left(\frac{2}{2}\right)log_2\left(\frac{2}{2}\right) - \left(\frac{0}{2}\right)log_2\left(\frac{0}{2}\right) - \left(\frac{0}{2}\right)log_2\left(\frac{0}{2}\right) = 0$$

$$Entropy(Class|X \geq -4.451) = -\left(\frac{5}{13}\right)log_2\left(\frac{5}{13}\right) - \left(\frac{5}{13}\right)log_2\left(\frac{5}{13}\right) - \left(\frac{3}{13}\right)log_2\left(\frac{3}{13}\right)$$
$$= 1.549$$

$$Gain(Class|X <> -4.451) = 1.585 - \left[\frac{2}{15} \times 0 + \frac{13}{15} \times 1.549\right] = 0.2425$$

2. **For boundary value -4.073**

$$Entropy(Class|X < -4.073) = 0 - \left(\frac{1}{3}\right)log_2\left(\frac{1}{3}\right) - \left(\frac{2}{3}\right)log_2\left(\frac{2}{3}\right) = 0.918$$

$$Entropy(Class|X \geq -4.073) = -\left(\frac{5}{12}\right)log_2\left(\frac{5}{12}\right) - \left(\frac{4}{12}\right)log_2\left(\frac{4}{12}\right) - \left(\frac{3}{12}\right)log_2\left(\frac{3}{12}\right)$$
$$= 1.555$$

$$Gain(Class|X <> -4.073) = 1.585 - \left[\frac{3}{15} \times 0.918 + \frac{12}{15} \times 1.555\right] = 0.1574$$

3. **For boundary value -2.3285**

$$Entropy(Class|X < -2.3285) = 0 - \left(\frac{1}{5}\right)log_2\left(\frac{1}{5}\right) - \left(\frac{4}{5}\right)log_2\left(\frac{4}{5}\right) = 0.7219$$

$$Entropy(Class|X \geq -2.3285) = -\left(\frac{5}{10}\right)log_2\left(\frac{5}{10}\right) - \left(\frac{4}{10}\right)log_2\left(\frac{4}{10}\right) - \left(\frac{1}{10}\right)log_2\left(\frac{1}{10}\right)$$
$$= 1.361$$

$$Gain(Class|X <> -2.3285) = 1.585 - \left[\frac{5}{15} \times 0.7219 + \frac{10}{15} \times 1.361\right] = 0.437$$

4. For boundary value -1.4785

$$Entropy(Class|X < -1.4785) = \ 0 - \left(\frac{3}{7}\right) log_2\left(\frac{3}{7}\right) - \left(\frac{4}{7}\right) log_2\left(\frac{4}{7}\right) = 0.985$$

$$Entropy(Class|X \geq -1.4785) = \ -\left(\frac{5}{8}\right) log_2\left(\frac{5}{8}\right) - \left(\frac{2}{8}\right) log_2\left(\frac{2}{8}\right) - \left(\frac{1}{8}\right) log_2\left(\frac{1}{8}\right) = 1.299$$

$$Gain(Class|X <> -1.4785) = 1.585 - \left[\frac{7}{15} \times 0.985 + \frac{8}{15} \times 1.299\right] = 0.8922$$

5. For boundary value -1.1595

$$Entropy(Class|X < -1.1595) = \ 0 - \left(\frac{3}{8}\right) log_2\left(\frac{3}{8}\right) - \left(\frac{5}{8}\right) log_2\left(\frac{5}{8}\right) = 0.954$$

$$Entropy(Class|X \geq -1.1595) = \ -\left(\frac{5}{7}\right) log_2\left(\frac{5}{7}\right) - \left(\frac{2}{7}\right) log_2\left(\frac{2}{7}\right) + 0 = 0.863$$

$$Gain(Class|X <> -1.1595) = 1.585 - \left[\frac{8}{15} \times 0.954 + \frac{7}{15} \times 0.863\right] = 0.673$$

6. For boundary value -0.303

$$Entropy(Class|X < -0.303) = \ 0 - \left(\frac{4}{9}\right) log_2\left(\frac{4}{9}\right) - \left(\frac{5}{9}\right) log_2\left(\frac{5}{9}\right) = 0.991$$

$$Entropy(Class|X \geq -0.303) = \ -\left(\frac{5}{6}\right) log_2\left(\frac{5}{6}\right) - \left(\frac{1}{6}\right) log_2\left(\frac{1}{6}\right) + 0 = 0.650$$

$$Gain(Class|X <> -0.303) = 1.585 - \left[\frac{9}{15} \times 0.991 + \frac{6}{15} \times 0.650\right] = 0.7304$$

7. For boundary value 3.3545

$$Entropy(Class|X < 3.3545) = -\left(\frac{4}{13}\right)log_2\left(\frac{4}{13}\right) - \left(\frac{4}{13}\right)log_2\left(\frac{4}{13}\right) - \left(\frac{5}{13}\right)log_2\left(\frac{5}{13}\right) = 1.5766$$

$$Entropy(Class|X \geq 3.3545) = -\left(\frac{1}{2}\right)log_2\left(\frac{1}{1}\right) - \left(\frac{1}{2}\right)log_2\left(\frac{1}{2}\right) = 1$$

$$Gain(Class|X <> 3.3545) = 1.585 - \left[\frac{13}{15} \times 1.5766 + \frac{2}{15} \times 1\right] = 0.0853$$

8. For boundary value 5.005

$$Entropy(Class|X < 5.005) = -\left(\frac{9}{14}\right)log_2\left(\frac{9}{14}\right) - \left(\frac{5}{14}\right)log_2\left(\frac{5}{14}\right) - \left(\frac{5}{14}\right)log_2\left(\frac{5}{14}\right) = 1.5774$$

$$Entropy(Class|X \geq 5.005) = -\left(\frac{1}{1}\right)log_2\left(\frac{1}{1}\right) = 0$$

$$Gain(Class|X <> 5.005) = 1.585 - \left[\frac{14}{15} \times 1.5774 + \frac{1}{15} \times 0\right] = 0.1128$$

For feature X, the threshold ≥ -0.303 has the highest information gain of any of the candidate thresholds.

Dataset, sorted by Y feature.

SI	x	y	Class
7.	-1.136	-2.992	ω_2
9.	-1.794	-2.838	ω_2
10.	-2.137	-2.473	ω_2
6.	4.218	-2.075	ω_2
4.	2.054	-1.467	ω_1
11.	-2.520	0.483	ω_3
2.	1.053	0.677	ω_1

8.	-4.435	1.408	ω_2
1.	2.491	2.176	ω_1
14.	-4.467	2.298	ω_3
13.	-13.438	2.414	ω_3
12.	-1.163	3.161	ω_3
3.	5.792	3.421	ω_1
5.	0.550	4.020	ω_1
15.	-3.711	4.364	ω_3

The boundary value between each of these pairs is simply the average of their Y values:

1. $\frac{d_6 + d_4}{2} = \frac{-2.075 + (-1.467)}{2} = -1.771$
2. $\frac{d_4 + d_{10}}{2} = \frac{-1.467 + 0.483}{2} = -0.492$
3. $\frac{d_{10} + d_2}{2} = \frac{0.483 + 0.677}{2} = 0.58$
4. $\frac{d_2 + d_8}{2} = \frac{0.677 + 1.408}{2} = 1.0425$
5. $\frac{d_8 + d_1}{2} = \frac{1.408 + 2.176}{2} = 1.792$
6. $\frac{d_1 + d_{14}}{2} = \frac{2.176 + 2.298}{2} = 2.237$
7. $\frac{d_{12} + d_3}{2} = \frac{3.161 + 3.425}{2} = 3.293$
8. $\frac{d_5 + d_{15}}{2} = \frac{4.020 + 4.364}{2} = 4.192$

1. **For boundary value -1.771**

$Entropy(Class|Y < -1.771) = -\left(\frac{4}{4}\right) log_2\left(\frac{4}{4}\right) = 0$

$$Entropy(Class|Y \geq -1.771) = -\left(\frac{5}{11}\right) log_2\left(\frac{5}{11}\right) - \left(\frac{1}{11}\right) log_2\left(\frac{1}{11}\right) - \left(\frac{5}{11}\right) log_2\left(\frac{5}{11}\right) = 1.3486$$

$$Gain(Class|Y <> -1.771) = 1.585 - \left[\frac{4}{15} \times 0 + \frac{11}{15} \times 1.3486\right] = 0.596$$

2. For boundary value -0.492

$$Entropy(Class|Y < -0.492) = -\left(\frac{1}{5}\right) log_2\left(\frac{1}{5}\right) - \left(\frac{4}{5}\right) log_2\left(\frac{4}{5}\right) = 0.722$$

$$Entropy(Class|Y \geq -0.492) = -\left(\frac{4}{10}\right) log_2\left(\frac{4}{10}\right) - \left(\frac{1}{10}\right) log_2\left(\frac{1}{10}\right) - \left(\frac{5}{10}\right) log_2\left(\frac{5}{10}\right) = 1.361$$

$$Gain(Class|Y <> -0.492) = 1.585 - \left[\frac{5}{15} \times 0.722 + \frac{10}{15} \times 1.361\right] = 0.437$$

3. For boundary value 0.58

$$Entropy(Class|Y < 0.58) = -\left(\frac{1}{6}\right) log_2\left(\frac{1}{6}\right) - \left(\frac{4}{6}\right) log_2\left(\frac{4}{6}\right) - \left(\frac{1}{6}\right) log_2\left(\frac{1}{6}\right) = 1.252$$

$$Entropy(Class|Y \geq 0.58) = -\left(\frac{4}{9}\right) log_2\left(\frac{4}{9}\right) - \left(\frac{1}{9}\right) log_2\left(\frac{1}{9}\right) - \left(\frac{4}{9}\right) log_2\left(\frac{4}{9}\right) = 1.392$$

$$Gain(Class|Y <> 0.58) = 1.585 - \left[\frac{6}{15} \times 1.252 + \frac{9}{15} \times 1.392\right] = 0.249$$

4. For boundary value 1.0425

$$Entropy(Class|Y < 1.0425) = -\left(\frac{2}{7}\right) log_2\left(\frac{2}{2}\right) - \left(\frac{4}{7}\right) log_2\left(\frac{4}{7}\right) - \left(\frac{1}{7}\right) log_2\left(\frac{1}{7}\right) = 1.379$$

$$Entropy(Class|Y \geq 1.0425) = -\left(\frac{3}{8}\right) log_2\left(\frac{3}{8}\right) - \left(\frac{1}{8}\right) log_2\left(\frac{1}{8}\right) - \left(\frac{4}{8}\right) log_2\left(\frac{4}{8}\right) = 1.406$$

$$Gain(Class|Y <> 1.0425) = 1.585 - \left[\frac{7}{15} \times 1.379 + \frac{8}{15} \times 1.406\right] = 0.1916$$

5. **For boundary value 1.792**

$$Entropy(Class|Y < 1.792) = -\left(\frac{2}{8}\right) log_2\left(\frac{2}{8}\right) - \left(\frac{5}{8}\right) log_2\left(\frac{5}{8}\right) - \left(\frac{1}{8}\right) log_2\left(\frac{1}{8}\right) = 1.299$$

$$Entropy(Class|Y \geq 1.792) = -\left(\frac{3}{7}\right) log_2\left(\frac{3}{7}\right) - \left(\frac{4}{7}\right) log_2\left(\frac{4}{7}\right) = 0.985$$

$$Gain(Class|Y <> 1.792) = 1.585 - \left[\frac{8}{15} \times 1.299 + \frac{7}{15} \times 0.985\right] = 0.4325$$

6. **For boundary value 2.237**

$$Entropy(Class|Y < 2.237) = -\left(\frac{3}{9}\right) log_2\left(\frac{3}{9}\right) - \left(\frac{5}{9}\right) log_2\left(\frac{5}{9}\right) - \left(\frac{1}{9}\right) log_2\left(\frac{1}{9}\right) = 1.352$$

$$Entropy(Class|Y \geq 2.237) = -\left(\frac{2}{6}\right) log_2\left(\frac{2}{6}\right) - \left(\frac{4}{6}\right) log_2\left(\frac{4}{6}\right) = 0.9143$$

$$Gain(Class|Y <> 2.237) = 1.585 - \left[\frac{9}{15} \times 1.352 + \frac{6}{15} \times 0.9143\right] = 0.4065$$

7. **For boundary value 3.293**

$$Entropy(Class|Y < 3.293) = -\left(\frac{3}{12}\right) log_2\left(\frac{3}{12}\right) - \left(\frac{5}{12}\right) log_2\left(\frac{5}{12}\right) - \left(\frac{4}{12}\right) log_2\left(\frac{4}{12}\right) = 1.555$$

$$Entropy(Class|Y \geq 3.293) = -\left(\frac{2}{3}\right) log_2\left(\frac{2}{3}\right) - 0 - \left(\frac{1}{3}\right) log_2\left(\frac{1}{3}\right) = 0.9183$$

$$Gain(Class|Y <> 3.293) = 1.585 - \left[\frac{12}{15} \times 1.555 + \frac{3}{15} \times 0.9183\right] = 0.15734$$

8. **For boundary value 4.192**

$$Entropy(Class|Y < 4.192) = -\left(\frac{5}{14}\right) log_2\left(\frac{5}{14}\right) - \left(\frac{5}{14}\right) log_2\left(\frac{5}{14}\right) - \left(\frac{4}{14}\right) log_2\left(\frac{4}{14}\right) = 1.5774$$

$$Entropy(Class|Y \geq 4.192) = -0 - 0 - \left(\frac{1}{1}\right) log_2 \left(\frac{1}{1}\right) = 0$$

$$Gain(Class|Y <> 4.192) = 1.585 - \left[\frac{1}{15} \times 0 + \frac{14}{15} \times 1.5.774\right] = 0.1128$$

For feature Y, the threshold ≥ -1.771 has the highest information gain of any of the candidate thresholds.

Now, In X and Y feature the highest information gain is the threshold ≥ -0.303 of X feature. So, we use X ≥ -0.303 as the test at the root node of the tree.

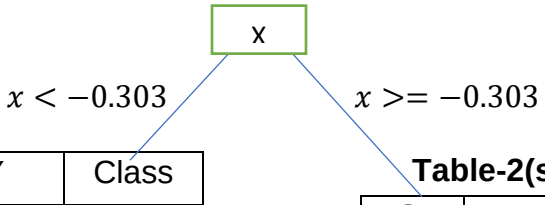

Table-1(sorted by X)

Sl	X	Y	Class
13.	-13.438	2.414	ω_3
14.	-4.467	2.298	ω_3
8.	-4.435	1.408	ω_2
15.	-3.711	4.364	ω_3
11.	-2.520	0.483	ω_3
10.	-2.137	-2.473	ω_2
9.	-1.794	-2.838	ω_2
12.	-1.163	3.161	ω_3
7.	-1.156	-2.992	ω_2

Table-2(sorted by X)

Sl	Y	Y	Class
5.	0.550	4.020	ω_1
2.	1.053	0.677	ω_1
4.	2.054	-1.467	ω_1
1.	2.491	2.176	ω_1
6.	4.218	-2.075	ω_2
3.	5.792	3.425	ω_1

From Table-1(X):

The boundary value between each of these pairs is simply the average of their X values:

1. $\frac{d_{14} + d_8}{2} = -4.451$

2. $\frac{d_8 + d_{15}}{2} = -4.073$

3. $\frac{d_{11} + d_{10}}{2} = -2.3285$

4. $\frac{d_9 + d_{12}}{2} = -1.4785$

5. $\frac{d_{12} + d_7}{2} = -1.1595$

$$Entropy(Class) = -p(\omega_1)log_2(p(\omega_1)) - p(\omega_2)log_2(p(\omega_2)) - p(\omega_3)log_2(p(\omega_3))$$

$$= -\left(\tfrac{0}{9}\right)log_2\left(\tfrac{0}{9}\right) - \left(\tfrac{4}{9}\right)log_2\left(\tfrac{4}{9}\right) - \left(\tfrac{5}{9}\right)log_2\left(\tfrac{5}{9}\right)$$

$$= 0.991$$

1. For boundary value -4.451

$$Entropy(Class|X < -4.451) = -\left(\frac{2}{2}\right)log_2\left(\frac{2}{2}\right) = 0$$

$$Entropy(Class|X \geq -4.451) = -\left(\frac{4}{7}\right)log_2\left(\frac{4}{7}\right) - \left(\frac{3}{7}\right)log_2\left(\frac{3}{7}\right) = 0.985$$

$$Gain(Class|X <> -4.451) = 0.991 - \left[\frac{2}{9} \times 0 + \frac{7}{9} \times 0.985\right] = 0.225$$

2. For boundary value -4.073

$$Entropy(Class|X < -4.073) = -\left(\frac{1}{3}\right)log_2\left(\frac{1}{3}\right) - \left(\frac{2}{3}\right)log_2\left(\frac{2}{3}\right) = 0.9183$$

$$Entropy(Class|X \geq -4.073) = -\left(\frac{3}{6}\right)log_2\left(\frac{3}{6}\right) - \left(\frac{3}{6}\right)log_2\left(\frac{3}{6}\right) = 1$$

$$Gain(Class|X <> -4.073) = 0.991 - \left[\frac{3}{9} \times 0.9183 + \frac{6}{9} \times 1\right] = 0.018$$

3. For boundary value -2.3285

$$Entropy(Class|X < -2.3285) = -\left(\frac{1}{5}\right)log_2\left(\frac{1}{5}\right) - \left(\frac{4}{5}\right)log_2\left(\frac{4}{5}\right) = 0.7219$$

$$Entropy(Class|X \geq -2.3285) = -\left(\frac{3}{4}\right) log_2\left(\frac{3}{4}\right) - \left(\frac{1}{4}\right) log_2\left(\frac{1}{4}\right) = 0.8113$$

$$Gain(Class|X <> -2.3285) = 0.991 - \left[\frac{5}{9} \times 0.7219 + \frac{4}{9} \times 0.8113\right] = 0.2294$$

4. For boundary value -1.4785

$$Entropy(Class|X < -1.4785) = -\left(\frac{3}{7}\right) log_2\left(\frac{3}{7}\right) - \left(\frac{4}{7}\right) log_2\left(\frac{4}{7}\right) = 0.985$$

$$Entropy(Class|X \geq -1.4785) = -\left(\frac{1}{2}\right) log_2\left(\frac{1}{2}\right) - \left(\frac{1}{2}\right) log_2\left(\frac{1}{2}\right) = 1$$

$$Gain(Class|X <> -1.4785) = 0.991 - \left[\frac{7}{9} \times 0.985 + \frac{2}{9} \times 1\right] = 0.0027$$

5. For boundary value -1.1595

$$Entropy(Class|X < -1.1595) = -\left(\frac{3}{8}\right) log_2\left(\frac{3}{8}\right) - \left(\frac{5}{8}\right) log_2\left(\frac{5}{8}\right) = 0.9544$$

$$Entropy(Class|X \geq -1.1595) = -\left(\frac{1}{1}\right) log_2\left(\frac{1}{1}\right) = 0$$

$$Gain(Class|X <> -1.1595) = 0.991 - \left[\frac{8}{9} \times 0.9544 + \frac{1}{9} \times 0\right] = 0.1426$$

For feature X, the threshold ≥ -2.3285 has the highest information gain of any of the candidate thresholds.

Table-1(sorted by Y feature)

Sl	x	y	Class
7.	-1.136	-2.992	ω_2
9.	-1.794	-2.838	ω_2
10.	-2.137	-2.473	ω_2
11.	-2.520	0.483	ω_3

8.	-4.435	1.408	ω_2
14.	-4.467	2.298	ω_3
13.	-13.438	2.414	ω_3
12.	-1.163	3.161	ω_3
15.	-3.711	4.364	ω_3

The boundary value between each of these pairs is simply the average of their Y values:

1. $\frac{d_{10} + d_{11}}{2} = -0.995$
2. $\frac{d_{11} + d_8}{2} = 0.946$
3. $\frac{d_8 + d_{14}}{2} = 1.853$

1. For boundary value -0.995

$$Entropy(Class|Y < -0.995) = -\left(\frac{3}{3}\right) log_2\left(\frac{3}{3}\right) = 0$$

$$Entropy(Class|Y \geq -0.995) = -\left(\frac{1}{6}\right) log_2\left(\frac{1}{6}\right) - \left(\frac{5}{6}\right) log_2\left(\frac{5}{6}\right) = 0.65$$

$$Gain(Class|Y <> -0.995) = 0.991 - \left[\frac{3}{9} \times 0 + \frac{6}{9} \times 0.65\right] = 0.558$$

2. For boundary value 0.946

$$Entropy(Class|Y < 0.946) = -\left(\frac{3}{4}\right) log_2\left(\frac{3}{4}\right) - \left(\frac{1}{4}\right) log_2\left(\frac{4}{4}\right) = 0.8113$$

$$Entropy(Class|Y \geq 0.946) = -\left(\frac{1}{5}\right) log_2\left(\frac{1}{5}\right) - \left(\frac{4}{5}\right) log_2\left(\frac{4}{5}\right) = 0.723$$

$$Gain(Class|Y <> 0.946) = 0.991 - \left[\frac{4}{9} \times 0.8113 + \frac{5}{9} \times 0.723\right] = 0.229$$

3. For boundary value 1.853

$$Entropy(Class|Y < 1.853) = -\left(\frac{4}{5}\right) log_2 \left(\frac{4}{5}\right) - \left(\frac{1}{5}\right) log_2 \left(\frac{1}{5}\right) = 0.722$$

$$Entropy(Class|Y \geq 1.853) = -\left(\frac{4}{4}\right) log_2 \left(\frac{4}{4}\right) = 0$$

$$Gain(Class|Y <> 1.853) = 0.991 - \left[\frac{5}{9} \times 0.722 + \frac{4}{9} \times 0\right] = 0.5899$$

For feature Y, the threshold ≥ 1.853 has the highest information gain of any of the candidate thresholds

In X and Y feature the highest information gain is the threshold ≥ 1.853 of Y feature. So, we use Y ≥ 1.853 as the test at the next node of the tree.

Now the decision tree is –

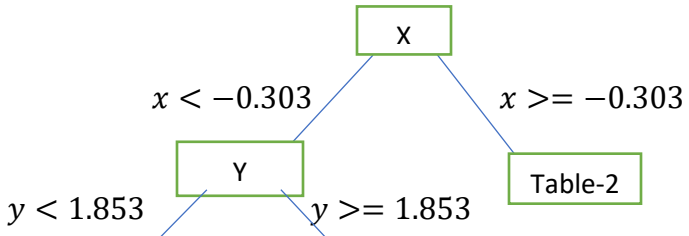

Table-3 (sorted by Y)

Sl	X	Y	Class
7.	-1.136	-2.992	ω_2
9.	-1.794	-2.838	ω_2
10.	-2.137	-2.473	ω_2
11.	-2.520	0.483	ω_3
8.	-4.435	1.408	ω_2

Table-4 (sorted by Y)

Sl	X	Y	Class
14.	-4.467	2.298	ω_3
13.	-13.438	2.414	ω_3
12.	-1.163	3.161	ω_3
15.	-3.711	4.364	ω_3

ω_3

From Table-3(Y):

The boundary value between each of these pairs is simply the average of their X values:

1. $\frac{d_{10} + d_{11}}{2} = -0.995$

$$2.\ \frac{d_{11}+d_8}{2}=0.945$$

$$Entropy(Class)=-\left(\tfrac{4}{5}\right)log_2\left(\tfrac{4}{5}\right)-\left(\tfrac{1}{5}\right)log_2\left(\tfrac{1}{5}\right)$$

$$=0.722$$

1. For boundary value -0.995

$$Entropy(Class|Y<-0.995)=\ -\left(\frac{3}{3}\right)log_2\left(\frac{3}{3}\right)=0$$

$$Entropy(Class|Y\geq-0.995)=\ -\left(\frac{1}{2}\right)log_2\left(\frac{1}{2}\right)-\left(\frac{1}{2}\right)log_2\left(\frac{1}{2}\right)=1$$

$$Gain(Class|Y<>-0.995)=0.722-\left[\frac{3}{5}\times0+\frac{2}{5}\times1\right]=0.322$$

2. For boundary value 0.945

$$Entropy(Class|Y<0.945)=\ -\left(\frac{3}{4}\right)log_2\left(\frac{3}{4}\right)\ -\left(\frac{1}{4}\right)log_2\left(\frac{1}{4}\right)=0.8113$$

$$Entropy(Class|Y\geq0.945)=\ -\left(\frac{1}{1}\right)log_2\left(\frac{1}{1}\right)=0$$

$$Gain(Class|Y<>0.945)=0.722-\left[\frac{4}{5}\times0.8113+\frac{1}{5}\times0\right]=0.073$$

For feature Y, the threshold ≥ -0.995has the highest information gain of any of the candidate thresholds.

Table-3 (sorted by X)

SI	x	y	Class
8.	-4.435	1.408	ω_2
11.	-2.520	0.483	ω_3
10.	-2.137	-2.473	ω_2
9.	-1.794	-2.838	ω_2
7.	-1.156	-2.992	ω_2

The boundary value between each of these pairs is simply the average of their X values:

1. $\frac{d_8 + d_{11}}{2} = -3.4775$

2. $\frac{d_{11} + d_{10}}{2} = -2.3285$

1. For boundary value -3.4775

$$Entropy(Class|X < -3.4775) = -\left(\frac{1}{1}\right) log_2\left(\frac{1}{1}\right) = 0$$

$$Entropy(Class|X \geq -3.4775) = -\left(\frac{3}{4}\right) log_2\left(\frac{3}{4}\right) - \left(\frac{1}{4}\right) log_2\left(\frac{1}{4}\right) = 0.8113$$

$$Gain(Class|X <> -3.4775) = 0.722 - \left[\frac{1}{5} \times 0 + \frac{4}{5} \times 0.8113\right] = 0.073$$

2. For boundary value -2.3285

$$Entropy(Class|X < -2.3285) = -\left(\frac{1}{2}\right) log_2\left(\frac{1}{2}\right) - \left(\frac{1}{2}\right) log_2\left(\frac{1}{2}\right) = 1$$

$$Entropy(Class|X \geq -2.3285) = -\left(\frac{3}{3}\right) log_2\left(\frac{3}{3}\right) = 0$$

$$Gain(Class|X <> -2.3285) = 0.722 - \left[\frac{2}{5} \times 1 + \frac{3}{5} \times 0\right] = 0.322$$

For feature X, the threshold ≥ -2.3285 has the highest information gain of any of the candidate thresholds

In X and Y feature the highest information gain is the threshold ≥ 1.853 of X and Y feature. So, we use anyone. Here we use X ≥ -2.3285 as the test at the next node of the tree.

Then we get two table. For X < -2.3285 , we get Table-5 and for X ≥ -2.3285 , we get Table-6. For Table-6, the class is ω_2.

Table-5(sorted by X)

Sl	X	Y	Class
8.	-4.435	1.408	ω_2
11.	-2.520	0.483	ω_3

Table-6(sorted by X)

Sl	X	Y	Class
10.	-2.137	-2.473	ω_2
9.	-1.794	-2.838	ω_2
7.	-1.156	-2.992	ω_2

From Table-5(x):

The boundary value between each of these pairs is simply the average of their X values:

$$1.\ \frac{d_8 + d_{11}}{2} = -3.4775$$

$$Entropy(Class) = -\left(\frac{1}{2}\right) log_2\left(\frac{1}{2}\right) - \left(\frac{1}{2}\right) log_2\left(\frac{1}{2}\right)$$

$$= 1$$

1. For boundary value -3.4775

$Entropy(Class|X < -3.4775) = 0\ (\omega_2)$

$Entropy(Class|X \geq -3.4775) = 0\ (\omega_3))$

$Gain(Class|X <> -3.4775) = 1 - 0 = 1$

So, we get two class by the threshold ≥ -3.4775 of X. For X <-3.4775 , The class is ω_2 and for X ≥ -3.4775 ,the class is ω_3 .

Now,
From the Table-2(X):

The boundary value between each of these pairs is simply the average of their X values:

$$1.\ \frac{d_1 + d_6}{2} = 3.3545$$

$$2.\ \frac{d_6 + d_3}{2} = 5.005$$

$$Entropy(Class) = -\left(\frac{5}{6}\right) log_2\left(\frac{5}{6}\right) - \left(\frac{1}{6}\right) log_2\left(\frac{1}{6}\right)$$

$$= 0.65$$

1. For boundary value 3.3545

$$Entropy(Class|X < 3.3545) = -\left(\frac{4}{4}\right) log_2\left(\frac{4}{4}\right) = 0$$

$$Entropy(Class|X \geq 3.3545) = -\left(\frac{1}{2}\right) log_2\left(\frac{1}{2}\right) - \left(\frac{1}{2}\right) log_2\left(\frac{1}{2}\right) = 1$$

$$Gain(Class|X <> 3.3545) = 0.65 - \left[\frac{4}{6} \times 0 + \frac{2}{6} \times 1\right] = 0.317$$

2. For boundary value 5.005

$$Entropy(Class|X < 5.005) = -\left(\frac{4}{5}\right) log_2\left(\frac{4}{5}\right) - \left(\frac{1}{5}\right) log_2\left(\frac{1}{5}\right) = 0.722$$

$$Entropy(Class|X \geq 5.005) = -\left(\frac{1}{1}\right) log_2\left(\frac{1}{1}\right) = 0$$

$$Gain(Class|X <> 5.005) = 0.65 - \left[\frac{5}{6} \times 0.722 + \frac{1}{6} \times 0\right] = 0.0483$$

For feature X, the threshold ≥ 3.3545 has the highest information gain of any of the candidate thresholds.

Table-2(sorted by y)

SI	Y	Y	Class
6.	4.218	-2.075	ω_2
4.	2.054	-1.467	ω_1
2.	1.053	0.677	ω_1
1.	2.491	2.176	ω_1
3.	5.792	3.425	ω_1
5.	0.550	4.020	ω_1

The boundary value between each of these pairs is simply the average of their Y values:

$$1.\ \frac{d_6 + d_4}{2} = -1.771$$

1. For boundary value -1.771

$$Entropy(Class|X < -1.771) = -\left(\frac{1}{1}\right) log_2 \left(\frac{1}{1}\right) = 0\ (\omega_2)$$

$$Entropy(Class|X \geq -1.771) = -\left(\frac{5}{5}\right) log_2 \left(\frac{5}{5}\right) = 0\ (\omega_1)$$

$$Gain(Class|X <> -1.771) = 0.65 - 0 = 0.65$$

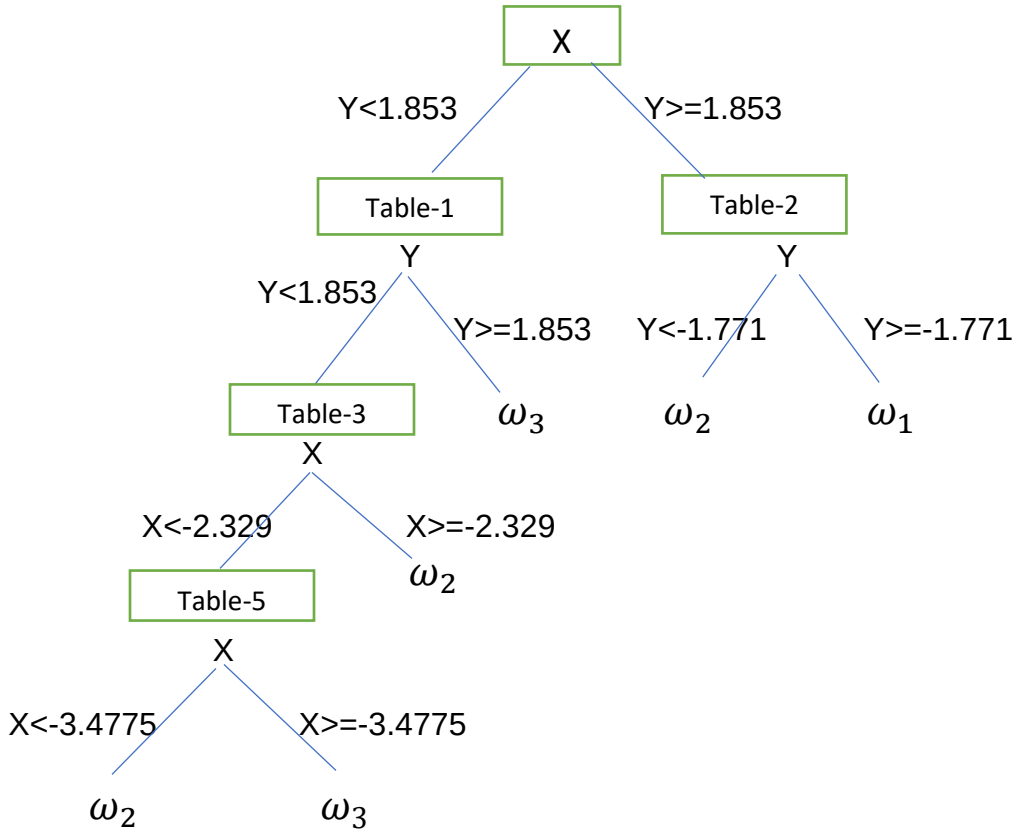

Fig: Final Decision Tree (ID3)

Test: X=-2.799, Y=0.746

Result: X=-2.799 < 0.303
Go Table-1, then Y = 0.746 < 1.853
Go Table-3, then X = -2.799 < -2.329
Go Table-5, then X = -2.799>=-3.4775

So, Result Class = ω_3

II. Using **C4.5 algorithm**:

Dataset, sorted by X feature which is same as ID3.

Sl	x	y	Class
13.	-13.438	2.414	ω_3
14.	-4.467	2.298	ω_3
8.	-4.435	1.408	ω_2
15.	-3.711	4.364	ω_3
11.	-2.520	0.483	ω_3
10.	-2.137	-2.473	ω_2
9.	-1.794	-2.838	ω_2
12.	-1.163	3.161	ω_3
7.	-1.156	-2.992	ω_2
5.	0.550	4.020	ω_1
2.	1.053	0.677	ω_1
4.	2.054	-1.467	ω_1
1.	2.491	2.176	ω_1
6.	4.218	-2.075	ω_2
3.	5.792	3.425	ω_1

The boundary value between each of these pairs is simply the average of their X values:

1. $\frac{d_{14} + d_8}{2} = \frac{-4.467 + (-4.435)}{2} = -4.451$

2. $\frac{d_8 + d_{15}}{2} = \frac{-4.435 + (-3.711)}{2} = -4.073$

3. $\frac{d_{11} + d_{10}}{2} = \frac{-2.520 + (-2.137)}{2} = -2.3285$

4. $\frac{d_9 + d_{12}}{2} = \frac{-1.794 + (-1.163)}{2} = -1.4785$

5. $\frac{d_{12} + d_7}{2} = \frac{-1.163 + (-1.156)}{2} = -1.1595$

6. $\frac{d_7 + d_5}{2} = \frac{-1.156 + 0.550}{2} = -0.303$

7. $\frac{d_1 + d_6}{2} = \frac{2.491 + 4.218}{2} = 3.3545$

8. $\frac{d_6 + d_3}{2} = \frac{4.218 + 5.792}{2} = 5.005$

$$Entropy(Class) = -p(\omega_1)log_2(p(\omega_1)) - p(\omega_2)log_2(p(\omega_2)) - p(\omega_3)log_2(p(\omega_3))$$

$$= -\left(\tfrac{5}{15}\right)log_2\left(\tfrac{5}{15}\right) - \left(\tfrac{5}{15}\right)log_2\left(\tfrac{5}{15}\right) - \left(\tfrac{5}{15}\right)log_2\left(\tfrac{5}{15}\right)$$

$$= 1.585$$

1. For boundary value -4.451

$Gain(Class|X <> -4.451) = 0.2425$ [$by\ ID3$]

$$SplitInfo(Class| <> -4.451) = -\left(\frac{2}{15}\right)log_2\left(\frac{2}{15}\right) - \left(\frac{13}{15}\right)log_2\left(\frac{13}{15}\right) = 0.567$$

$$Gain\ Ratio(Class| <> -4.451) = \frac{0.2425}{0.567} = 0.428$$

2. For boundary value -4.073

$Gain(Class|X <> -4.073) = 0.1574$ [$by\ ID3$]

$SplitInfo(Class|X <> -4.073) = -\left(\frac{3}{15}\right) log_2\left(\frac{3}{15}\right) - \left(\frac{12}{15}\right) log_2\left(\frac{12}{15}\right) = 0.722$

$Gain\ Ratio(Class|X <> -4.073) = \frac{0.1574}{0.722} = 0.218$

3. For boundary value -2.3285

$Gain(Class|X <> -2.3285) = 0.437$ [$by\ ID3$]

$SplitInfo(Class|X <> -2.3285) = -\left(\frac{5}{15}\right) log_2\left(\frac{5}{15}\right) - \left(\frac{10}{15}\right) log_2\left(\frac{10}{15}\right) = 0.918$

$Gain\ Ratio(Class|X <> -2.3285) = \frac{0.437}{0.918} = 0.476$

4. For boundary value -1.4786

$Gain(Class|X <> -1.4786) = 0.437$ [$by\ ID3$]

$SplitInfo(Class|X <> -1.4786) = -\left(\frac{5}{15}\right) log_2\left(\frac{5}{15}\right) - \left(\frac{10}{15}\right) log_2\left(\frac{10}{15}\right) = 0.918$

$Gain\ Ratio(Class|X <> -1.4786) = \frac{0.437}{0.918} = 0.476$

5. For boundary value -1.1595

$Gain(Class|X <> -1.1595) = 0.673$ [$by\ ID3$]

$SplitInfo(Class|X <> -1.1595) = -\left(\frac{8}{15}\right) log_2\left(\frac{8}{15}\right) - \left(\frac{7}{15}\right) log_2\left(\frac{7}{15}\right) = 0.997$

$Gain\ Ratio(Class|X <> -1.1595) = \frac{0.673}{0.997} = 0.675$

6. For boundary value -0.303

$Gain(Class|X <> -0.303) = 0.7304\ [by\ ID3]$

$SplitInfo(Class|X <> -0.303) = -\left(\frac{9}{15}\right) log_2\left(\frac{9}{15}\right) - \left(\frac{6}{15}\right) log_2\left(\frac{6}{15}\right) = 0.971$

$Gain\ Ratio(Class|X <> -0.303) = \frac{0.7304}{0.971} = 0.752$

7. For boundary value 3.3545

$Gain(Class|X <> 3.3545) = 0.0853\ [by\ ID3]$

$SplitInfo(Class|X <> 3.3545) = -\left(\frac{13}{15}\right) log_2\left(\frac{13}{15}\right) - \left(\frac{2}{15}\right) log_2\left(\frac{2}{15}\right) = 0.567$

$Gain\ Ratio(Class|X <> 3.3545) = \frac{0.0853}{0.567} = 0.1504$

8. For boundary value 5.005

$Gain(Class|X <> 5.005) = 0.1128\ [by\ ID3]$

$SplitInfo(Class|X <> 5.005) = -\left(\frac{14}{15}\right) log_2\left(\frac{14}{15}\right) - \left(\frac{1}{15}\right) log_2\left(\frac{1}{15}\right) = 0.3534$

$Gain\ Ratio(Class|X <> 5.005) = \frac{0.1128}{0.3534} = 0.3192$

So, for boundary value -0.303, we get the highest gain ratio.

Dataset, sorted by Y feature which is same as ID3.

Sl	x	y	Class
7.	-1.136	-2.992	ω_2
9.	-1.794	-2.838	ω_2
10.	-2.137	-2.473	ω_2

6.	4.218	-2.075	ω_2
4.	2.054	-1.467	ω_1
10.	-2.520	0.483	ω_3
2.	1.053	0.677	ω_1
8.	-4.435	1.408	ω_2
1.	2.491	2.176	ω_1
14.	-4.467	2.298	ω_3
13.	-13.438	2.414	ω_3
12.	-1.163	3.161	ω_3
3.	5.792	3.421	ω_1
5.	0.550	4.020	ω_1
15.	-3.711	4.364	ω_3

The boundary value between each of these pairs is simply the average of their Y values:

1. $\frac{d_6 + d_4}{2} = \frac{-2.075 + (-1.467)}{2} = -1.771$
2. $\frac{d_4 + d_{10}}{2} = \frac{-1.467 + 0.483}{2} = -0.492$
3. $\frac{d_{10} + d_2}{2} = \frac{0.483 + 0.677}{2} = 0.58$
4. $\frac{d_2 + d_8}{2} = \frac{0.677 + 1.408}{2} = 1.0425$
5. $\frac{d_8 + d_1}{2} = \frac{1.408 + 2.176}{2} = 1.792$
6. $\frac{d_1 + d_{14}}{2} = \frac{2.176 + 2.298}{2} = 2.237$
7. $\frac{d_{12} + d_3}{2} = \frac{3.161 + 3.425}{2} = 3.293$
8. $\frac{d_5 + d_{15}}{2} = \frac{4.020 + 4.364}{2} = 4.192$

1. **For boundary value -1.771**

$Gain(Class|Y <> -1.771) = 0.596\ [by\ ID3]$

$SplitInfo(Class|Y <> -1.771) = -\left(\frac{4}{15}\right) log_2\left(\frac{4}{15}\right) - \left(\frac{11}{15}\right) log_2\left(\frac{11}{15}\right) = 0.837$

$Gain\ Ratio(Class|Y <> -1.771) = \frac{0.596}{0.837} = 0.712$

2. **For boundary value -0.492**

$Gain(Class|Y <> -0.492) = 0.437\ [by\ ID3]$

$SplitInfo(Class|Y <> -0.492) = -\left(\frac{5}{15}\right) log_2\left(\frac{5}{15}\right) - \left(\frac{10}{15}\right) log_2\left(\frac{10}{15}\right) = 0.9183$

$Gain\ Ratio(Class|Y <> -0.492) = \frac{0.437}{0.9183} = 0.476$

3. **For boundary value 0.58**

$Gain(Class|Y <> 0.58) = 0.249\ [by\ ID3]$

$SplitInfo(Class|Y <> 0.58) = -\left(\frac{6}{15}\right) log_2\left(\frac{6}{15}\right) - \left(\frac{9}{15}\right) log_2\left(\frac{9}{15}\right) = 0.971$

$Gain\ Ratio(Class|Y <> 0.58) = \frac{0.249}{0.971} = 0.256$

4. **For boundary value 1.0425**

$Gain(Class|Y <> 1.0425) = 0.1916\ [by\ ID3]$

$SplitInfo(Class|Y <> 1.0425) = -\left(\frac{7}{15}\right) log_2\left(\frac{7}{15}\right) - \left(\frac{8}{15}\right) log_2\left(\frac{8}{15}\right) = 0.997$

$Gain\ Ratio(Class|Y <> 1.0425) = \frac{0.1916}{0.997} = 0.192$

5. For boundary value 1.792

$$Gain(Class|Y <> 1.792) = 0.4325\ [by\ ID3]$$

$$SplitInfo(Class|Y <> 1.792) = -\left(\frac{8}{15}\right) log_2\left(\frac{8}{15}\right) - \left(\frac{7}{15}\right) log_2\left(\frac{7}{15}\right) = 0.997$$

$$Gain\ Ratio(Class|Y <> 1.792) = \frac{0.4325}{0.997} = 0.434$$

6. For boundary value 2.237

$$Gain(Class|Y <> 2.237) = 0.4065\ [by\ ID3]$$

$$SplitInfo(Class|Y <> 2.237) = -\left(\frac{9}{15}\right) log_2\left(\frac{9}{15}\right) - \left(\frac{6}{15}\right) log_2\left(\frac{6}{15}\right) = 0.971$$

$$Gain\ Ratio(Class|Y <> 2.237) = \frac{0.4065}{0.971} = 0.419$$

7. For boundary value 3.293

$$Gain(Class|Y <> 3.293) = 0.1473\ [by\ ID3]$$

$$SplitInfo(Class|Y <> 3.293) = -\left(\frac{12}{15}\right) log_2\left(\frac{12}{15}\right) - \left(\frac{3}{15}\right) log_2\left(\frac{3}{15}\right) = 0.722$$

$$Gain\ Ratio(Class|Y <> 3.293) = \frac{0.1473}{0.722} = 0.218$$

8. For boundary value 4.192

$$Gain(Class|Y <> 4.192) = 0.1128\ [by\ ID3]$$

$$SplitInfo(Class|Y <> 4.192) = -\left(\frac{14}{15}\right) log_2\left(\frac{14}{15}\right) - \left(\frac{1}{15}\right) log_2\left(\frac{1}{15}\right) = 0.3534$$

$$Gain\ Ratio(Class|Y <> 4.192) = \frac{0.1128}{0.3534} = 0.3192$$

For feature Y, the threshold ≥ -1.771 has the highest information gain ratio of any of the candidate thresholds.

Now, In X and Y feature the highest information gain ratio is the threshold ≥ -0.303 of X feature. So, we use X ≥ -0.303 as the test at the root node of the tree.

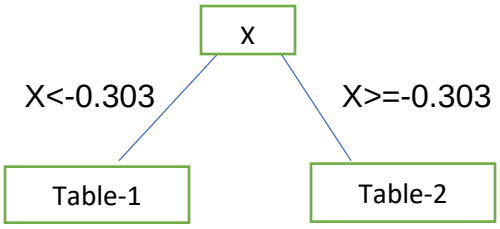

Table-1(for X feature) [From ID3]

The boundary value between each of these pairs is simply the average of their X values:

1. $\frac{d_{14} + d_8}{2} = -4.451$
2. $\frac{d_8 + d_{15}}{2} = -4.073$
3. $\frac{d_{11} + d_{10}}{2} = -2.3285$
4. $\frac{d_9 + d_{12}}{2} = -1.4785$
5. $\frac{d_{12} + d_7}{2} = -1.1595$

$$Entropy(Class) = -p(\omega_1)log_2(p(\omega_1)) - p(\omega_2)log_2(p(\omega_2)) - p(\omega_3)log_2(p(\omega_3))$$

$$= -\left(\frac{0}{9}\right)log_2\left(\frac{0}{9}\right) - \left(\frac{4}{9}\right)log_2\left(\frac{4}{9}\right) - \left(\frac{5}{9}\right)log_2\left(\frac{5}{9}\right)$$

$$= 0.991$$

1. For boundary value -4.451

$Gain(Class|X <> -4.451) = 0.225$ [By ID3]

$$SplitInfo(Class|X <> -4.451) = -\left(\frac{7}{9}\right)log_2\left(\frac{7}{9}\right) - \left(\frac{2}{9}\right)log_2\left(\frac{2}{9}\right) = 0.764$$

$$Gain\ Ratio(Class|X <> -4.451) = \frac{0.225}{0.764} = 0.295$$

2. For boundary value -4.073

$$Gain(Class|X <> -4.073) = 0.018\ [by\ ID3]$$

$$SplitInfo(Class|X <> -4.073) = -\left(\frac{3}{9}\right) log_2\left(\frac{3}{9}\right) - \left(\frac{6}{9}\right) log_2\left(\frac{6}{9}\right) = 0.918$$

$$Gain\ Ratio(Class|X <> -4.073) = \frac{0.018}{0.918} = 0.0196$$

3. For boundary value -2.3285

$$Gain(Class|X <> -2.3285) = 0.2294\ [by\ ID3]$$

$$SplitInfo(Class|X <> -2.3285) = -\left(\frac{5}{9}\right) log_2\left(\frac{5}{9}\right) - \left(\frac{4}{9}\right) log_2\left(\frac{4}{9}\right) = 0.991$$

$$Gain\ Ratio(Class|X <> -2.3285) = \frac{0.2294}{0.991} = 0.231$$

4. For boundary value -1.4785

$$Gain(Class|X <> -1.4785) = 0.0027\ [by\ ID3]$$

$$SplitInfo(Class|X <> -1.4785) = -\left(\frac{7}{9}\right) log_2\left(\frac{7}{9}\right) - \left(\frac{2}{9}\right) log_2\left(\frac{2}{9}\right) = 0.764$$

$$Gain\ Ratio(Class|X <> -1.4785) = \frac{0.0027}{0.764} = 0.0035$$

5. For boundary value -1.1595

$$Gain(Class|X <> -1.1595) = 0.1426\ [by\ ID3]$$

$$SplitInfo(Class|X <> -1.1595) = -\left(\frac{8}{9}\right) log_2\left(\frac{8}{9}\right) - \left(\frac{1}{9}\right) log_2\left(\frac{1}{9}\right) = 0.5033$$

$$Gain\ Ratio(Class|X <> -1.1595) = \frac{0.1426}{0.5033} = 0.283$$

For feature X, the threshold ≥ -4.451 has the highest information gain ratio of any of the candidate thresholds.

From Table-1(for Y feature) [From ID3]

The boundary value between each of these pairs is simply the average of their Y values:

1. $\frac{d_{10} + d_{11}}{2} = -0.995$
2. $\frac{d_{11} + d_{8}}{2} = 0.946$
3. $\frac{d_{8} + d_{14}}{2} = 1.853$

1. For boundary value -0.995

$Gain(Class|Y <> -0.995) = 0.558\ [by\ ID3]$

$SplitInfo(Class|Y <> -0.995) = -\left(\frac{3}{9}\right) log_2\left(\frac{3}{9}\right) - \left(\frac{6}{9}\right) log_2\left(\frac{6}{9}\right) = 0.918$

$Gain\ Ratio(Class|Y <> -0.995) = \frac{0.558}{0.918} = 0.608$

2. For boundary value 0.946

$Gain(Class|Y <> 0.946) = 0.229\ [by\ ID3]$

$SplitInfo(Class|Y <> 0.946) = -\left(\frac{4}{9}\right) log_2\left(\frac{4}{9}\right) - \left(\frac{5}{9}\right) log_2\left(\frac{5}{9}\right) = 0.991$

$Gain\ Ratio(Class|Y <> 0.946) = \frac{0.229}{0.991} = 0.2310$

3. For boundary value 1.853

$Gain(Class|Y <> 1.853) = 0.5899\ [by\ ID3]$

$$SplitInfo(Class|Y <> 1.853) = -\left(\frac{4}{9}\right) log_2\left(\frac{4}{9}\right) - \left(\frac{5}{9}\right) log_2\left(\frac{5}{9}\right) = 0.991$$

$$Gain\ Ratio(Class|Y <> 1.853) = \frac{0.5899}{0.991} = 0.595$$

For feature Y, the threshold ≥ -0.995has the highest information gain ratio of any of the candidate thresholds.

Now, In X and Y feature the highest information gain ratio is the threshold $\geq -0.995$of Y feature. So, we use Y ≥ -0.995as the test at the next node of the tree.

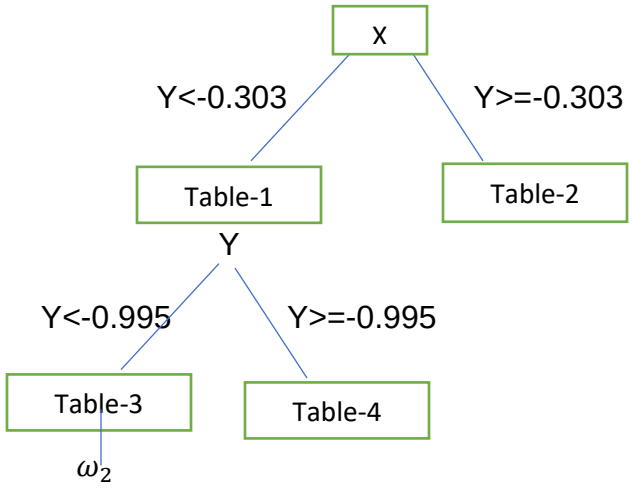

Table-4 (Sorted by Y)

Sl	Y	Y	Class
11.	-2.520	0.483	ω_3
8.	-4.435	1.408	ω_2
14.	-4.467	2.298	ω_3
13.	-13.438	2.414	ω_3
12.	-1.163	3.161	ω_3
15.	-3.711	4.364	ω_3

The boundary value between each of these pairs is simply the average of their Y values:

1. $\frac{d_{11} + d_8}{2} = 0.945$

2. $\frac{d_8 + d_{14}}{2} = 1.853$

$$Entropy(Class) = -\left(\tfrac{5}{6}\right) log_2 \left(\tfrac{5}{6}\right) - \left(\tfrac{1}{6}\right) log_2 \left(\tfrac{1}{6}\right) = 0.65$$

1. For boundary value 0.945

$$Entropy(Class|Y < 0.945) = -\left(\frac{1}{1}\right) log_2 \left(\frac{1}{1}\right) = 0$$

$$Entropy(Class|Y \geq 0.945) = -\left(\frac{1}{5}\right) log_2 \left(\frac{1}{5}\right) - \left(\frac{4}{5}\right) log_2 \left(\frac{4}{5}\right) = 0.722$$

$$Gain(Class|Y <> 0.945) = 0.65 - \left[\frac{1}{6} \times 0 + \frac{5}{6} \times 0.722\right] = 0.0483$$

$$SplitInfo(Class|Y <> 0.945) = -\left(\frac{1}{6}\right) log_2 \left(\frac{1}{6}\right) - \left(\frac{5}{6}\right) log_2 \left(\frac{5}{6}\right) = 0.65$$

$$Gain\ Ratio(Class|Y <> 0.945) = \frac{0.0483}{0.65} = 0.074$$

2. For boundary value 1.853

$$Entropy(Class|Y < 1.853) = -\left(\frac{1}{2}\right) log_2 \left(\frac{1}{2}\right) - \left(\frac{1}{2}\right) log_2 \left(\frac{1}{2}\right) = 1$$

$$Entropy(Class|Y \geq 1.853) = -\left(\frac{4}{4}\right) log_2 \left(\frac{4}{4}\right) = 0$$

$$Gain(Class|Y <> 1.853) = 0.65 - \left[\frac{2}{6} \times 1 + \frac{4}{6} \times 0\right] = 0.317$$

$$SplitInfo(Class|Y <> 1.853) = -\left(\frac{2}{6}\right) log_2 \left(\frac{2}{6}\right) - \left(\frac{4}{6}\right) log_2 \left(\frac{4}{6}\right) = 0.918$$

$$Gain\ Ratio(Class|Y <> 1.853) = \frac{0.317}{0.918} = 0.345$$

Table-4 (Sorted by X)

Sl	Y	Y	Class
13.	-13.438	2.414	ω_3
14.	-4.467	2.298	ω_3
8.	-4.435	1.408	ω_2
15.	-3.711	4.364	ω_3
11.	-2.520	0.483	ω_3
12.	-1.163	3.161	ω_3

The boundary value between each of these pairs is simply the average of their X values:

1. $\frac{d_{14} + d_8}{2} = -4.451$
2. $\frac{d_8 + d_{15}}{2} = -4.073$

1. For boundary value -4.451

$$Entropy(Class|X < -4.451) = -\left(\frac{2}{2}\right) log_2\left(\frac{2}{2}\right) = 0$$

$$Entropy(Class|X \geq -4.451) = -\left(\frac{1}{4}\right) log_2\left(\frac{1}{4}\right) - \left(\frac{3}{4}\right) log_2\left(\frac{3}{4}\right) = 0.8113$$

$$Gain(Class|X <> -4.451) = 0.65 - \left[\frac{2}{6} \times 0 + \frac{4}{6} \times 0.8113\right] = 0.109$$

$$SplitInfo(Class|X <> -4.451) = -\left(\frac{2}{6}\right) log_2\left(\frac{2}{6}\right) - \left(\frac{4}{6}\right) log_2\left(\frac{4}{6}\right) = 0.918$$

$$Gain\ Ratio(Class|X <> -4.451) = \frac{0.109}{0.918} = 0.119$$

2. For boundary value -4.073

$$Entropy(Class|X < -4.073) = -\left(\frac{1}{3}\right) log_2\left(\frac{1}{3}\right) - \left(\frac{2}{3}\right) log_2\left(\frac{2}{3}\right) = 0.918$$

$$Entropy(Class|X \geq -4.073) = -\left(\frac{3}{3}\right) log_2\left(\frac{3}{3}\right) = 0$$

$$Gain(Class|X <> -4.073) = 0.65 - \left[\frac{3}{6} \times 0.918 + \frac{3}{6} \times 0\right] = 0.191$$

$$SplitInfo(Class|X <> -4.073) = -\left(\frac{3}{6}\right) log_2\left(\frac{3}{6}\right) - \left(\frac{3}{6}\right) log_2\left(\frac{3}{6}\right) = 1$$

$$Gain\ Ratio(Class|X <> -4.073) = \frac{0.191}{1} = 0.191$$

In X and Y feature the highest information gain ratio is the threshold $\geq 1.853$of Y feature. So, we use Y ≥ 1.853 as the test at the next node of the tree.

Then we get two table. For Y < 1.853 , we get Table-5 and for Y ≥ 1.853 , we get Table-6. For Table-6, the class is ω_3.

Table-5 (sorted by Y)

SI	X	Y	Class
11.	-2.520	0.483	ω_3
8.	-4.435	1.408	ω_2

The boundary value between each of these pairs is simply the average of their Y values:

$$1.\ \frac{d_{11} + d_8}{2} = 0.946$$

$$Entropy(Class|Y <> 0.946) = 0$$

$$Gain(Class|Y <> 0.946) = 1$$

$$SplitInfo(Class|Y <> 0.946) = 1$$

$$Gain\ Ratio(Class|Y <> 0.946) = \frac{1}{1} = 1$$

Then we get two class. For Y < 0.946 ,class is ω_3 and for Y ≥ 0.946 class is ω_2.

Table-2 (For X feature) [From ID3]

The boundary value between each of these pairs is simply the average of their X values:

1. $\frac{d_1 + d_6}{2} = 3.3545$

2. $\frac{d_6 + d_3}{2} = 5.005$

1. For boundary value 3.3545

$Gain(Class|X <> 3.3545) = 0.317\ [by\ ID3]$

$SplitInfo(Class|X <> 3.3545) = -\left(\frac{2}{6}\right) log_2\left(\frac{2}{6}\right) - \left(\frac{4}{6}\right) log_2\left(\frac{4}{6}\right) = 0.918$

$Gain\ Ratio(Class|X <> 3.3545) = \frac{0.317}{0.918} = 0.345$

2. For boundary value 5.005

$Gain(Class|X <> 5.005) = 0.0483\ [by\ ID3]$

$SplitInfo(Class|X <> 5.005) = -\left(\frac{1}{6}\right) log_2\left(\frac{1}{6}\right) - \left(\frac{5}{6}\right) log_2\left(\frac{5}{6}\right) = 0.65$

$Gain\ Ratio(Class|X <> 5.005) = \frac{0.0483}{0.65} = 0.075$

Table-2 (For Y feature) [From ID3]

The boundary value between each of these pairs is simply the average of their Y values:

1. $\frac{d_6 + d_4}{2} = -1.771$

1. For boundary value -1.771

$Gain(Class|Y <> -1.771) = 0.65\ [by\ ID3]$

$SplitInfo(Class|Y <> -1.771) = -\left(\frac{1}{6}\right) log_2\left(\frac{1}{6}\right) - \left(\frac{5}{6}\right) log_2\left(\frac{5}{6}\right) = 0.65$

$$Gain\ Ratio(Class|Y <> -1.771) = \frac{0.65}{0.65} = 1$$

So, the final Decision tree is-

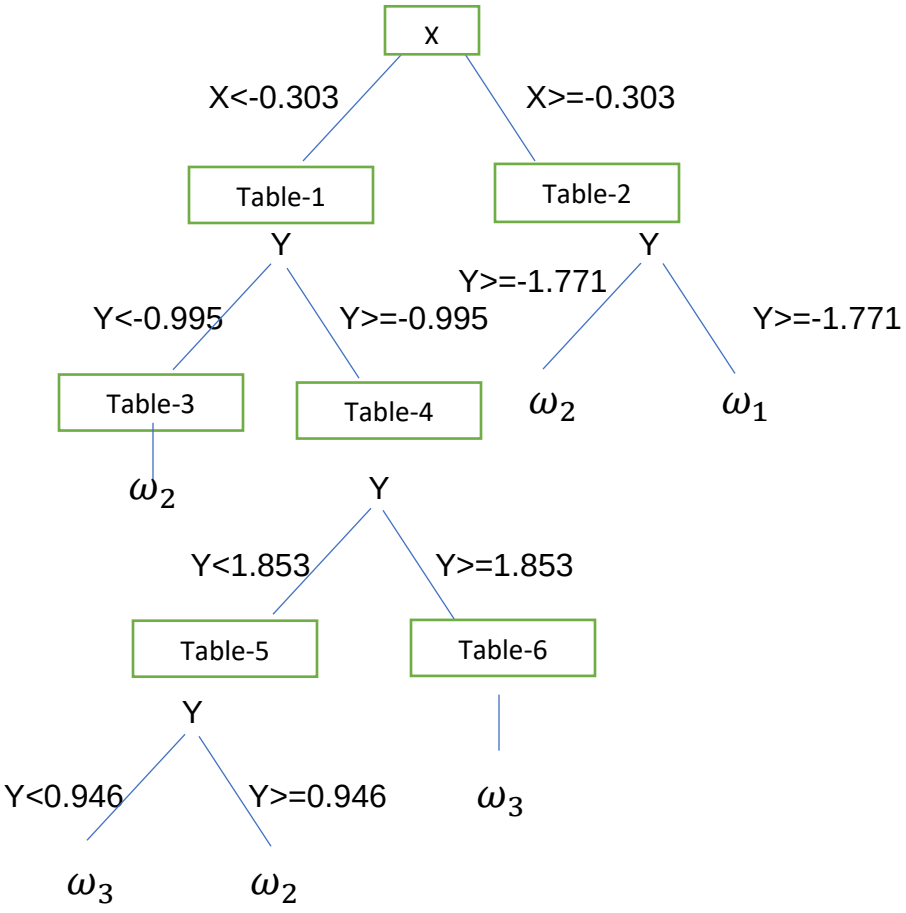

Fig: Final Decision Tree (C4.5)

Test:

Test: X=-2.799, Y=0.746

Result: X=-2.799 < 0.303
Go Table-1, then X = -2.799 < -0.303
Go Table-4, then Y = 0.746 >= -0.995
Go Table-5, then Y = 0.746 < 1.853
then Y = 0.746 <0. 946

So, Result Class = ω_3.

III. Using **CART algorithm**:

Dataset, sorted by X feature which is same as ID3 and C4.5.

Sl	x	y	Class
13.	−13.438	2.414	ω_3
14.	-4.467	2.298	ω_3
8.	-4.435	1.408	ω_2
15.	-3.711	4.364	ω_3
11.	-2.520	0.483	ω_3
10.	-2.137	-2.473	ω_2
9.	-1.794	-2.838	ω_2
12.	-1.163	3.161	ω_3
7.	-1.156	-2.992	ω_2
5.	0.550	4.020	ω_1
2.	1.053	0.677	ω_1
4.	2.054	-1.467	ω_1
1.	2.491	2.176	ω_1
6.	4.218	-2.075	ω_2
3.	5.792	3.425	ω_1

The boundary value between each of these pairs is simply the average of their X values:

1. $\frac{d_{14} + d_8}{2} = \frac{-4.467 + (-4.435)}{2} = -4.451$
2. $\frac{d_8 + d_{15}}{2} = \frac{-4.435 + (-3.711)}{2} = -4.073$
3. $\frac{d_{11} + d_{10}}{2} = \frac{-2.520 + (-2.137)}{2} = -2.3285$
4. $\frac{d_9 + d_{12}}{2} = \frac{-1.794 + (-1.163)}{2} = -1.4785$

5. $\frac{d_{12}+d_7}{2}=\frac{-1.163+(-1.156)}{2}=-1.1595$

6. $\frac{d_7+d_5}{2}=\frac{-1.156+0.550}{2}=-0.303$

7. $\frac{d_1+d_6}{2}=\frac{2.491+4.218}{2}=3.3545$

8. $\frac{d_6+d_3}{2}=\frac{4.218+5.792}{2}=5.005$

1. For boundary value -4.451

$$Gini(Class|X <> -4.451)$$
$$=\frac{2}{15}\left[1-\left(\frac{0}{2}\right)^2-\left(\frac{0}{2}\right)^2-\left(\frac{1}{2}\right)^2\right]+\frac{13}{15}\left[1-\left(\frac{5}{13}\right)^2-\left(\frac{5}{13}\right)^2-\left(\frac{3}{13}\right)^2\right]$$
$$=0.564$$

2. For boundary value -4.073

$$Gini(Class|X <> -4.073)$$
$$=\frac{3}{15}\left[1-\left(\frac{0}{3}\right)^2-\left(\frac{1}{3}\right)^2-\left(\frac{2}{3}\right)^2\right]+\frac{12}{15}\left[1-\left(\frac{5}{12}\right)^2-\left(\frac{4}{12}\right)^2-\left(\frac{3}{12}\right)^2\right]$$
$$=0.6109$$

3. For boundary value -2.3285

$$Gini(Class|X <> -2.3285)$$
$$=\frac{5}{15}\left[1-\left(\frac{0}{5}\right)^2-\left(\frac{1}{5}\right)^2-\left(\frac{4}{5}\right)^2\right]+\frac{10}{15}\left[1-\left(\frac{5}{10}\right)^2-\left(\frac{4}{10}\right)^2-\left(\frac{1}{10}\right)^2\right]$$
$$=0.4933$$

4. For boundary value -1.4785

$$Gini(Class|X <> -1.4785)=\frac{7}{15}\left[1-\left(\frac{0}{7}\right)^2-\left(\frac{3}{7}\right)^2-\left(\frac{4}{7}\right)^2\right]+\frac{8}{15}\left[1-\left(\frac{5}{8}\right)^2-\left(\frac{2}{8}\right)^2-\left(\frac{1}{8}\right)^2\right]$$
$$=0.5119$$

5. For boundary value -1.1595

$$Gini(Class|X <> -1.1595) = \frac{8}{15}\left[1 - \left(\frac{0}{8}\right)^2 - \left(\frac{3}{8}\right)^2 - \left(\frac{5}{8}\right)^2\right] + \frac{7}{15}\left[1 - \left(\frac{5}{7}\right)^2 - \left(\frac{2}{7}\right)^2 - \left(\frac{0}{7}\right)^2\right]$$

$$= 0.4405$$

6. For boundary value -0.303

$$Gini(Class|X <> -0.303) = \frac{9}{15}\left[1 - \left(\frac{0}{9}\right)^2 - \left(\frac{4}{9}\right)^2 - \left(\frac{5}{9}\right)^2\right] + \frac{6}{15}\left[1 - \left(\frac{5}{6}\right)^2 - \left(\frac{1}{6}\right)^2 - \left(\frac{0}{6}\right)^2\right]$$

$$= 0.407$$

7. For boundary value 3.3545

$$Gini(Class|X <> 3.3545)$$
$$= \frac{13}{15}\left[1 - \left(\frac{4}{13}\right)^2 - \left(\frac{4}{13}\right)^2 - \left(\frac{5}{13}\right)^2\right] + \frac{2}{15}\left[1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 - \left(\frac{0}{2}\right)^2\right]$$

$$= 0.641$$

8. For boundary value 5.005

$$Gini(Class|X <> 5.005)$$
$$= \frac{14}{15}\left[1 - \left(\frac{4}{14}\right)^2 - \left(\frac{5}{14}\right)^2 - \left(\frac{5}{14}\right)^2\right] + \frac{1}{15}\left[1 - \left(\frac{1}{1}\right)^2 - \left(\frac{0}{1}\right)^2 - \left(\frac{0}{1}\right)^2\right]$$

$$= 0.619$$

So, for boundary value -0.303 of X, we get the lowest Gini index.

Dataset, sorted by Y feature which is same as ID3 and C4.5.

Sl	x	y	Class
7.	-1.136	-2.992	ω_2
9.	-1.794	-2.838	ω_2
10.	-2.137	-2.473	ω_2
6.	4.218	-2.075	ω_2

4.	2.054	-1.467	ω_1
11.	-2.520	0.483	ω_3
2.	1.053	0.677	ω_1
8.	-4.435	1.408	ω_2
1.	2.491	2.176	ω_1
14.	-4.467	2.298	ω_3
13.	-13.438	2.414	ω_3
12.	-1.163	3.161	ω_3
3.	5.792	3.421	ω_1
5.	0.550	4.020	ω_1
15.	-3.711	4.364	ω_3

The boundary value between each of these pairs is simply the average of their Y values:

1. $\frac{d_6 + d_4}{2} = \frac{-2.075 + (-1.467)}{2} = -1.771$
2. $\frac{d_4 + d_{10}}{2} = \frac{-1.467 + 0.483}{2} = -0.492$
3. $\frac{d_{10} + d_2}{2} = \frac{0.483 + 0.677}{2} = 0.58$
4. $\frac{d_2 + d_8}{2} = \frac{0.677 + 1.408}{2} = 1.0425$
5. $\frac{d_8 + d_1}{2} = \frac{1.408 + 2.176}{2} = 1.792$
6. $\frac{d_1 + d_{14}}{2} = \frac{2.176 + 2.298}{2} = 2.237$
7. $\frac{d_{12} + d_3}{2} = \frac{3.161 + 3.425}{2} = 3.293$
8. $\frac{d_5 + d_{15}}{2} = \frac{4.020 + 4.364}{2} = 4.192$

1. **For boundary value -1.771**

$$Gini(Class|Y <> -1.771)$$
$$= \frac{4}{15}\left[1 - \left(\frac{0}{4}\right)^2 - \left(\frac{4}{4}\right)^2 - \left(\frac{0}{4}\right)^2\right] + \frac{11}{15}\left[1 - \left(\frac{5}{11}\right)^2 - \left(\frac{1}{11}\right)^2 - \left(\frac{5}{11}\right)^2\right]$$
$$= 0.424$$

2. **For boundary value -0.492**

$$Gini(Class|Y <> -0.492)$$
$$= \frac{5}{15}\left[1 - \left(\frac{1}{5}\right)^2 - \left(\frac{4}{5}\right)^2 - \left(\frac{0}{5}\right)^2\right] + \frac{10}{15}\left[1 - \left(\frac{4}{10}\right)^2 - \left(\frac{1}{10}\right)^2 - \left(\frac{5}{10}\right)^2\right]$$
$$= 0.493$$

3. **For boundary value 0.58**

$$Gini(Class|Y <> 0.58) = \frac{6}{15}\left[1 - \left(\frac{1}{6}\right)^2 - \left(\frac{4}{6}\right)^2 - \left(\frac{1}{6}\right)^2\right] + \frac{9}{15}\left[1 - \left(\frac{4}{9}\right)^2 - \left(\frac{1}{9}\right)^2 - \left(\frac{4}{9}\right)^2\right]$$
$$= 0.556$$

4. **For boundary value 1.0425**

$$Gini(Class|Y <> 1.0425) = \frac{7}{15}\left[1 - \left(\frac{2}{7}\right)^2 - \left(\frac{4}{7}\right)^2 - \left(\frac{1}{7}\right)^2\right] + \frac{8}{15}\left[1 - \left(\frac{3}{8}\right)^2 - \left(\frac{1}{8}\right)^2 - \left(\frac{4}{8}\right)^2\right]$$
$$= 0.583$$

5. **For boundary value 1.792**

$$Gini(Class|Y <> 1.792) = \frac{8}{15}\left[1 - \left(\frac{2}{8}\right)^2 - \left(\frac{5}{8}\right)^2 - \left(\frac{1}{8}\right)^2\right] + \frac{7}{15}\left[1 - \left(\frac{3}{7}\right)^2 - \left(\frac{0}{7}\right)^2 - \left(\frac{4}{7}\right)^2\right]$$
$$= 0.512$$

6. **For boundary value 2.237**

$$Gini(Class|Y <> 2.237) = \frac{9}{15}\left[1 - \left(\frac{3}{9}\right)^2 - \left(\frac{5}{9}\right)^2 - \left(\frac{1}{9}\right)^2\right] + \frac{6}{15}\left[1 - \left(\frac{2}{6}\right)^2 - \left(\frac{0}{6}\right)^2 - \left(\frac{4}{6}\right)^2\right]$$
$$= 0.519$$

7. For boundary value 3.293

$$Gini(Class|Y <> 3.293) = \frac{12}{15}\left[1 - \left(\frac{3}{12}\right)^2 - \left(\frac{5}{12}\right)^2 - \left(\frac{4}{12}\right)^2\right] + \frac{3}{15}\left[1 - \left(\frac{2}{3}\right)^2 - \left(\frac{0}{3}\right)^2 - \left(\frac{1}{3}\right)^2\right]$$

$$= 0.611$$

8. For boundary value 4.192

$$Gini(Class|Y <> 4.192) = \frac{14}{15}\left[1 - \left(\frac{5}{14}\right)^2 - \left(\frac{5}{14}\right)^2 - \left(\frac{4}{14}\right)^2\right] + \frac{1}{15}\left[1 - \left(\frac{0}{1}\right)^2 - \left(\frac{0}{1}\right)^2 - \left(\frac{1}{1}\right)^2\right]$$

$$= 0.619$$

So, for boundary value -0.303 of X, we get the lowest Gini index.

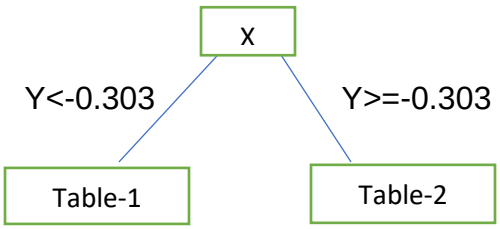

Table-1(for X feature) [From ID3]

The boundary value between each of these pairs is simply the average of their X values:

1. $\frac{d_{14} + d_8}{2} = -4.451$
2. $\frac{d_8 + d_{15}}{2} = -4.073$
3. $\frac{d_{11} + d_{10}}{2} = -2.3285$
4. $\frac{d_9 + d_{12}}{2} = -1.4785$
5. $\frac{d_{12} + d_7}{2} = -1.1595$

1. **For boundary value -4.451**

$$Gini(Class|X <> -4.451) = \frac{2}{9}\left[1 - \left(\frac{0}{2}\right)^2 - \left(\frac{0}{2}\right)^2 - \left(\frac{2}{2}\right)^2\right] + \frac{7}{9}\left[1 - \left(\frac{0}{7}\right)^2 - \left(\frac{4}{7}\right)^2 - \left(\frac{3}{7}\right)^2\right]$$

$$= 0.381$$

2. **For boundary value -4.073**

$$Gini(Class|X <> -4.073) = \frac{3}{9}\left[1 - \left(\frac{0}{3}\right)^2 - \left(\frac{1}{3}\right)^2 - \left(\frac{2}{3}\right)^2\right] + \frac{6}{9}\left[1 - \left(\frac{0}{6}\right)^2 - \left(\frac{3}{6}\right)^2 - \left(\frac{3}{6}\right)^2\right]$$

$$= 0.4815$$

3. **For boundary value -2.3285**

$$Gini(Class|X <> -2.3285) = \frac{5}{9}\left[1 - \left(\frac{0}{5}\right)^2 - \left(\frac{1}{5}\right)^2 - \left(\frac{4}{5}\right)^2\right] + \frac{4}{9}\left[1 - \left(\frac{0}{4}\right)^2 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2\right]$$

$$= 0.344$$

4. **For boundary value -1.4785**

$$Gini(Class|X <> -1.4785) = \frac{7}{9}\left[1 - \left(\frac{0}{7}\right)^2 - \left(\frac{3}{7}\right)^2 - \left(\frac{4}{7}\right)^2\right] + \frac{2}{9}\left[1 - \left(\frac{0}{2}\right)^2 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2\right]$$

$$= 0.111$$

5. **For boundary value -1.1595**

$$Gini(Class|X <> -1.1595) = \frac{8}{9}\left[1 - \left(\frac{0}{8}\right)^2 - \left(\frac{3}{8}\right)^2 - \left(\frac{5}{8}\right)^2\right] + \frac{1}{9}\left[1 - \left(\frac{0}{1}\right)^2 - \left(\frac{1}{1}\right)^2 - \left(\frac{0}{1}\right)^2\right]$$

$$= 0.4167$$

So, for boundary value -1.4785 of X, we get the lowest Gini index.

Table-1(for Y feature) [From ID3]

The boundary value between each of these pairs is simply the average of their Y values:

1. $\frac{d_{10} + d_{11}}{2} = -0.995$

2. $\frac{d_{11} + d_8}{2} = 0.946$

3. $\frac{d_8 + d_{14}}{2} = 1.853$

1. For boundary value -0.995

$$Gini(Class|Y <> -0.995) = \frac{3}{9}\left[1 - \left(\frac{0}{3}\right)^2 - \left(\frac{3}{3}\right)^2 - \left(\frac{0}{3}\right)^2\right] + \frac{6}{9}\left[1 - \left(\frac{0}{6}\right)^2 - \left(\frac{1}{6}\right)^2 - \left(\frac{5}{6}\right)^2\right]$$

$$= 0.185$$

2. For boundary value 0.946

$$Gini(Class|Y <> 0.946) = \frac{4}{9}\left[1 - \left(\frac{0}{4}\right)^2 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2\right] + \frac{5}{9}\left[1 - \left(\frac{0}{5}\right)^2 - \left(\frac{1}{5}\right)^2 - \left(\frac{4}{5}\right)^2\right]$$

$$= 0.344$$

3. For boundary value 1.853

$$Gini(Class|Y <> 1.853) = \frac{5}{9}\left[1 - \left(\frac{0}{5}\right)^2 - \left(\frac{4}{5}\right)^2 - \left(\frac{1}{5}\right)^2\right] + \frac{4}{9}\left[1 - \left(\frac{0}{4}\right)^2 - \left(\frac{0}{4}\right)^2 - \left(\frac{4}{4}\right)^2\right]$$

$$= 0.178$$

Here, for boundary value -1.4785 of X, we get the lowest Gini index. So, we use X -1.4785 as the test at the next node of the tree.

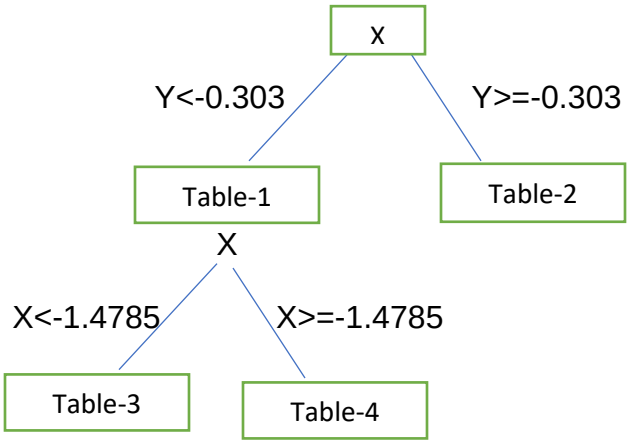

Table-3(sorted by X)

Sl	X	Y	Class
13.	-13.438	2.414	ω_3
14.	-4.467	2.298	ω_3
8.	-4.435	1.408	ω_2
15.	-3.711	4.364	ω_3
11.	-2.520	0.483	ω_3
10.	-2.137	-2.473	ω_2
9.	-1.794	-2.838	ω_2

Table-4(sorted by X)

Sl	Y	Y	Class
12.	-1.163	3.161	ω_3
7.	-1.156	-2.992	ω_2

From Table-4(X):

The boundary value between each of these pairs is simply the average of their X values:

$$1.\ \frac{d_{12}+d_7}{2} = -1.1595$$

1. For boundary value -1.1595

$$Gini(Class|X <> -1.1595) = \frac{1}{2}\left[1-\left(\frac{0}{1}\right)^2-\left(\frac{0}{1}\right)^2-\left(\frac{1}{1}\right)^2\right]+\frac{1}{2}\left[1-\left(\frac{0}{1}\right)^2-\left(\frac{0}{1}\right)^2-\left(\frac{1}{1}\right)^2\right]$$

$$= 0$$

Then we get two class. For X < -1.1595,class is ω_3 and for X ≥ -1.1595 class is ω_2.

From Table-3(X):

The boundary value between each of these pairs is simply the average of their X values:

$$1.\ \frac{d_{14}+d_8}{2} = -4.451$$

$$2.\ \frac{d_8+d_{15}}{2} = -4.073$$

$$3.\ \frac{d_{11}+d_{10}}{2} = -2.3285$$

1. For boundary value -4.451

$$Gini(Class|X <> -4.451) = \frac{2}{7}\left[1 - \left(\frac{0}{2}\right)^2 - \left(\frac{0}{2}\right)^2 - \left(\frac{2}{2}\right)^2\right] + \frac{5}{7}\left[1 - \left(\frac{0}{5}\right)^2 - \left(\frac{3}{5}\right)^2 - \left(\frac{2}{5}\right)^2\right]$$

$$= 0.3428$$

2. For boundary value -4.073

$$Gini(Class|X <> -4.073) = \frac{3}{7}\left[1 - \left(\frac{0}{3}\right)^2 - \left(\frac{1}{3}\right)^2 - \left(\frac{2}{3}\right)^2\right] + \frac{4}{7}\left[1 - \left(\frac{0}{4}\right)^2 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2\right]$$

$$= 0.4762$$

3. For boundary value -2.3285

$$Gini(Class|X <> -2.3285) = \frac{5}{7}\left[1 - \left(\frac{0}{5}\right)^2 - \left(\frac{1}{5}\right)^2 - \left(\frac{4}{5}\right)^2\right] + \frac{2}{7}\left[1 - \left(\frac{0}{2}\right)^2 - \left(\frac{2}{2}\right)^2 - \left(\frac{0}{2}\right)^2\right]$$

$$= 0.2285$$

So, for boundary value $-2.3285$of X, we get the lowest Gini index.

Table-3(sorted by Y feature)

Sl	x	y	Class
9.	-1.794	-2.838	ω_2
10.	-2.137	-2.473	ω_2
11.	-2.520	0.483	ω_3
8.	-4.435	1.408	ω_2
14.	-4.467	2.298	ω_3
13.	-13.438	2.414	ω_3
15.	-3.711	4.364	ω_3

The boundary value between each of these pairs is simply the average of their Y values:

1. $\frac{d_{10} + d_{11}}{2} = -0.995$

2. $\frac{d_{11} + d_8}{2} = 0.9455$

3. $\frac{d_8 + d_{14}}{2} = 1.853$

1. For boundary value -0.995

$$Gini(Class|Y <> -0.995) = \frac{2}{7}\left[1 - \left(\frac{0}{2}\right)^2 - \left(\frac{0}{2}\right)^2 - \left(\frac{2}{2}\right)^2\right] + \frac{5}{7}\left[1 - \left(\frac{0}{5}\right)^2 - \left(\frac{1}{5}\right)^2 - \left(\frac{4}{5}\right)^2\right]$$

$$= 0.229$$

2. For boundary value 0.9455

$$Gini(Class|Y <> 0.9455) = \frac{3}{7}\left[1 - \left(\frac{0}{3}\right)^2 - \left(\frac{2}{3}\right)^2 - \left(\frac{1}{3}\right)^2\right] + \frac{4}{7}\left[1 - \left(\frac{0}{4}\right)^2 - \left(\frac{1}{4}\right)^2 - \left(\frac{3}{4}\right)^2\right]$$

$$= 0.405$$

3. For boundary value 1.853

$$Gini(Class|Y <> 1.853) = \frac{4}{7}\left[1 - \left(\frac{0}{4}\right)^2 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2\right] + \frac{3}{7}\left[1 - \left(\frac{0}{3}\right)^2 - \left(\frac{0}{3}\right)^2 - \left(\frac{3}{3}\right)^2\right]$$

$$= 0.214$$

Here, for boundary value 1.853 of Y, we get the lowest Gini index. So, we use Y 1.853 as the test at the next node of the tree.

Then we get two table. For Y < 1.853 , we get Table-5 and for Y ≥ 1.853 , we get Table-6. For Table-6, the class is ω_3.

Table-5(sorted by Y)

SI	X	Y	Class
9.	-1.794	-2.838	ω_2
10.	-2.137	-2.473	ω_2
11.	-2.520	0.483	ω_3
8.	-4.435	1.408	ω_2

Table-6(sorted by Y)

SI	Y	Y	Class
13.	-13.438	2.414	ω_3
14.	-4.467	2.298	ω_3
15.	-3.711	4.364	ω_3

ω_3

Table-5(sorted by Y feature)

The boundary value between each of these pairs is simply the average of their Y values:

1. $\frac{d_{10} + d_{11}}{2} = -0.995$

2. $\frac{d_{11} + d_8}{2} = 0.9455$

1. For boundary value -0.995

$$Gini(Class|Y <> -0.995) = \frac{2}{4}\left[1 - \left(\frac{0}{2}\right)^2 - \left(\frac{2}{2}\right)^2 - \left(\frac{0}{2}\right)^2\right] + \frac{2}{4}\left[1 - \left(\frac{0}{2}\right)^2 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2\right]$$

$$= 0.25$$

2. For boundary value 0.9455

$$Gini(Class|Y <> -0.995) = \frac{3}{4}\left[1 - \left(\frac{0}{3}\right)^2 - \left(\frac{2}{3}\right)^2 - \left(\frac{1}{3}\right)^2\right] + \frac{1}{4}\left[1 - \left(\frac{0}{1}\right)^2 - \left(\frac{0}{1}\right)^2 - \left(\frac{1}{1}\right)^2\right]$$

$$= 0.333$$

So, for boundary value $-0.995$of Y, we get the lowest Gini index.

Table-5(sorted by X feature)

SI	X	Y	Class
8.	-4.435	1.408	ω_2
11.	-2.520	0.483	ω_3
10.	-2.137	-2.473	ω_2
9.	-1.794	-2.838	ω_2

The boundary value between each of these pairs is simply the average of their X values:

1. $\frac{d_8 + d_{11}}{2} = -3.4775$

$$2.\ \frac{d_{11}+d_{10}}{2}=-2.3285$$

1. For boundary value -3.4775

$$Gini(Class|X <> -3.4775)=\frac{1}{4}\left[1-\left(\frac{0}{1}\right)^2-\left(\frac{0}{1}\right)^2-\left(\frac{1}{1}\right)^2\right]+\frac{3}{4}\left[1-\left(\frac{0}{3}\right)^2-\left(\frac{2}{3}\right)^2-\left(\frac{1}{3}\right)^2\right]$$

$$=0.333$$

2. For boundary value -2.3285

$$Gini(Class|X <> -2.3285)=\frac{2}{4}\left[1-\left(\frac{0}{2}\right)^2-\left(\frac{0}{2}\right)^2-\left(\frac{1}{2}\right)^2\right]+\frac{2}{4}\left[1-\left(\frac{0}{2}\right)^2-\left(\frac{2}{2}\right)^2-\left(\frac{1}{2}\right)^2\right]$$

$$=0.25$$

Here, for boundary value $-2.3285$of X, we get the lowest Gini index from X and Y feature. So, we use X -2.3285 as the test at the next node of the tree.

Then we get two table. For X < -2.3285, we get Table-7 and for X $\geq -2.3285$of , we get the class of ω_2.

Table-7(sorted by X feature)

Sl	X	Y	Class
8.	-4.435	1.408	ω_2
11.	-2.520	0.483	ω_3

The boundary value between each of these pairs is simply the average of their X values:

$$1.\ \frac{d_8+d_{11}}{2}=-3.4775$$

1. For boundary value -3.4775

$$Gini(Class|X <> -3.4775)=\frac{1}{2}\left[1-\left(\frac{0}{1}\right)^2-\left(\frac{0}{1}\right)^2-\left(\frac{1}{1}\right)^2\right]+\frac{1}{2}\left[1-\left(\frac{0}{1}\right)^2-\left(\frac{0}{1}\right)^2-\left(\frac{1}{1}\right)^2\right]$$

$$=0$$

Then we get two class. For X < -3.4775, the class of ω_2 and for X ≥ -3.4775 of, we get the class of ω_3.

Table-2(sorted by X feature)

SI	X	Y	Class
5.	0.550	4.020	ω_1
2.	1.053	0.677	ω_1
4.	2.054	-1.467	ω_1
1.	2.491	2.176	ω_1
6.	4.218	-2.075	ω_2
3.	5.792	3.425	ω_1

The boundary value between each of these pairs is simply the average of their X values:

1. $\frac{d_1 + d_6}{2} = 3.3545$
2. $\frac{d_6 + d_3}{2} = 5.005$

1. For boundary value 3.3545

$$Gini(Class|X <> 3.3545) = \frac{4}{6}\left[1 - \left(\frac{0}{4}\right)^2 - \left(\frac{4}{4}\right)^2 - \left(\frac{0}{4}\right)^2\right] + \frac{2}{6}\left[1 - \left(\frac{0}{2}\right)^2 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2\right]$$

$$= 0.167$$

2. For boundary value 5.005

$$Gini(Class|X <> 5.005) = \frac{5}{6}\left[1 - \left(\frac{0}{5}\right)^2 - \left(\frac{4}{5}\right)^2 - \left(\frac{1}{5}\right)^2\right] + \frac{1}{6}\left[1 - \left(\frac{0}{1}\right)^2 - \left(\frac{1}{1}\right)^2 - \left(\frac{0}{1}\right)^2\right]$$

$$= 0.267$$

So, for boundary value 3.3545 of X, we get the lowest Gini index

Table-2(sorted by Y feature)

Sl	X	Y	Class
6.	4.218	-2.075	ω_2
4.	2.054	-1.467	ω_1
2.	1.053	0.677	ω_1
1.	2.491	2.176	ω_1
3.	5.792	3.425	ω_1
5.	0.550	4.020	ω_1

The boundary value between each of these pairs is simply the average of their Y values:

$$1.\ \frac{d_6 + d_4}{2} = -1.771$$

1. For boundary value -1.771

$$Gini(Class|X <> -1.771) = \frac{1}{6}\left[1 - \left(\frac{0}{1}\right)^2 - \left(\frac{0}{1}\right)^2 - \left(\frac{1}{1}\right)^2\right] + \frac{5}{6}\left[1 - \left(\frac{0}{5}\right)^2 - \left(\frac{5}{5}\right)^2 - \left(\frac{0}{5}\right)^2\right]$$

$$= 0$$

Here, for boundary value -1.771 of Y, we get the lowest Gini index. So, we use the threshold of -1.771 in feature Y as the test at the next node of the tree.

Then we get two class. For Y < -1.771 , the class of ω_2 and for X ≥ -1.771 of, we get the class of ω_1.

Now the final Decision Tree is –

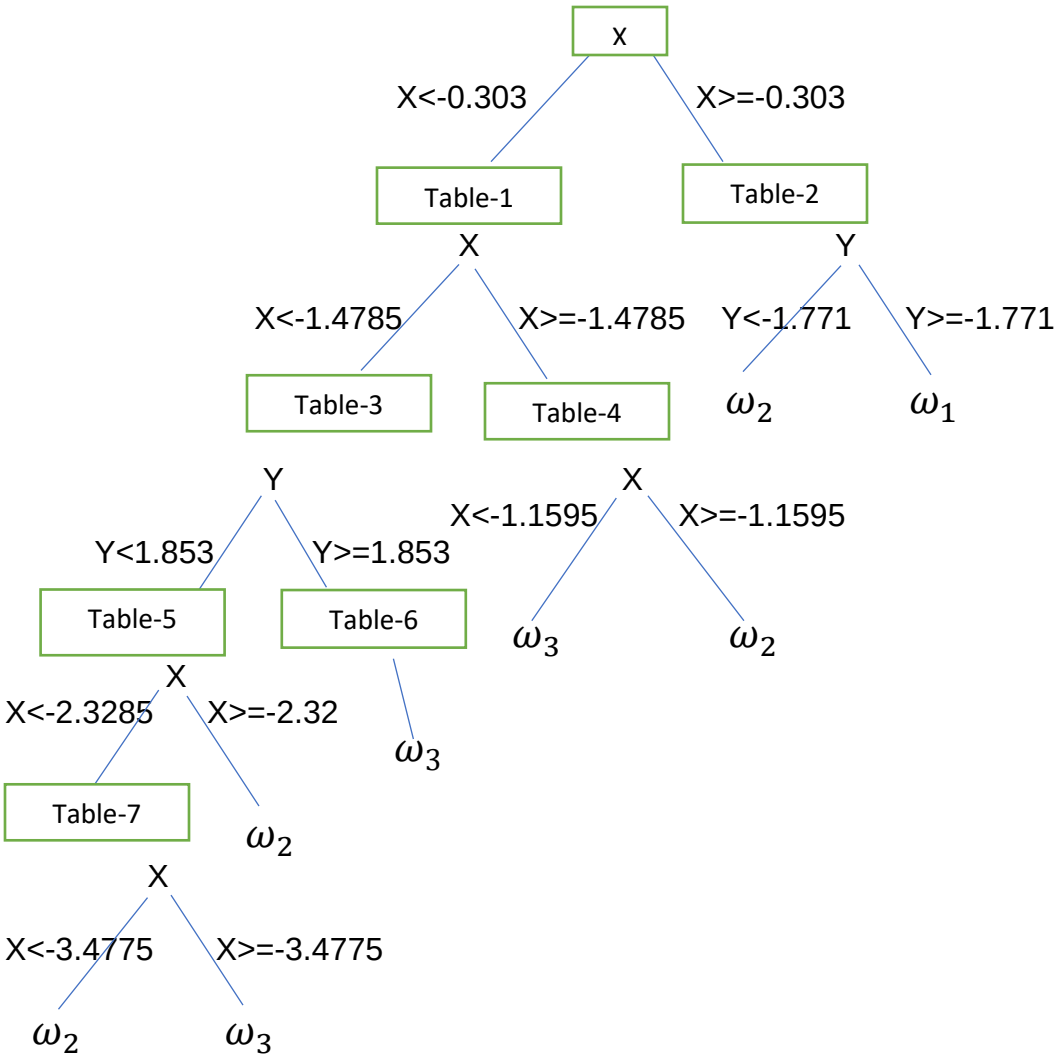

Fig: Final Decision Tree (CART)

Test:

Test: X=-2.799, Y=0.746

Result: X=-2.799 < 0.303
Go Table-1, then X = -2.799 < -0.303
Go Table-3, then X = -2.799 < -1.4785
Go Table-5, then Y = 0.746 < 1.853
Go Table-7, then X = -2.799 >= -3.4775, then get ω_3

So, Result Class = ω_3.