Q2. DECISION TREE USING INFORMATION GAIN.

PRATEEK MISHRA
IIT 2018199

Dataset.

	slope	y
1.	3.0	0
2.	2.0	1
3.	2	1
4.	3	0
5.	1	0
6.	1	0
7.	3	0
8.	1	0
9.	2	1
10.	3	1
11.	2	0
12.	2	1
13.	2	0
14.	1	0
15.	1	1
16.	1	0
17.	3	0
18.	1	1
19.	1	0
20.	1	1

Dataset sorted (ascending order slope)

	slope	y
1.		0
2.		0
3.		0
4.		0
5.		0
6.		0
7.		0
8.		0
9.		0
10.	2	1
11.	2	1
12.	2	0
13.	2	0
14.	2	0
15.	3	1
16.	3	0
17.	3	0
18.	3	0
19.	3	1
20.	3	1

Initial Entropy (I) = $-\left(\frac{13}{20}\log\left(\frac{13}{20}\right) + \frac{7}{20}\log\left(\frac{7}{20}\right)\right)$

Entropy (<1.5) = $-(1\log(1)) + 0\log(0) = 0$

Entropy (≥1.5) = $-\left(\frac{6}{11}\log\frac{6}{11} + \frac{5}{11}\log\left(\frac{5}{11}\right)\right)$

Information gain
for split value = I − (Entropy |>1.5) + Entropy(<1.5))
1.5

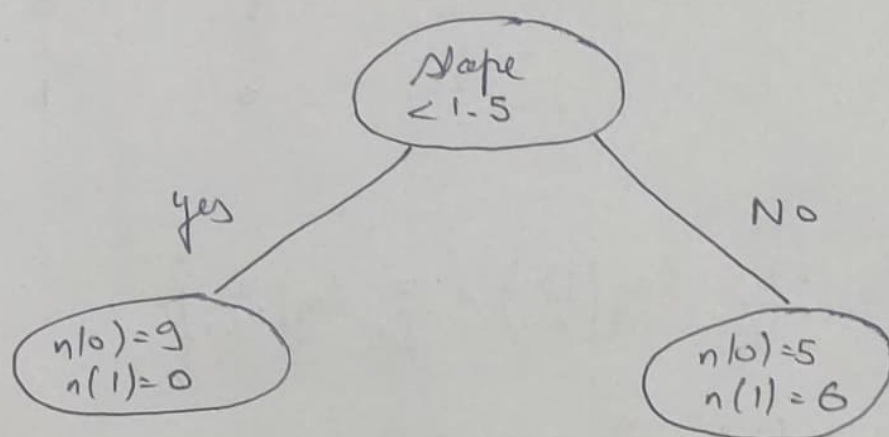$= I - \frac{6}{11}\log\frac{6}{11} + \frac{5}{11}\log\frac{5}{11}$

$$\text{Entropy} (<2.5) = -\left(\frac{4}{15} \log \frac{4}{15} + \frac{11}{15} \log \frac{11}{15} \right)$$

$$\text{Entropy} (>=2.5) = -\left(\frac{2}{5} \log \frac{2}{5} + \frac{3}{5} \log \frac{3}{5} \right)$$

IG for 2.5 = $I - (\text{Entropy} (>2.5) + \text{Entropy} (<2.5))$

$$= I + \frac{4}{15} \log \left(\frac{4}{15} \right) + \frac{11}{15} \log \left(\frac{11}{15} \right) + \frac{2}{5} \log \left(\frac{2}{5} \right) + \frac{3}{5} \log \left(\frac{3}{5} \right)$$

Thus   IG   for 1.5   >   IG   for 2.5

The data corresponding to slope < 1.5 belongs to only one class, so we do not need to split it further for data with slope ≠1.5

The only possible way of splitting for slope <2.5

$$\text{Entropy} (< 2.5) = -\left(\frac{4}{6} \log \frac{4}{5} + \frac{2}{6} \log \frac{2}{6} \right)$$

$$\text{Entropy} (> 2.5) = -\left(\frac{2}{5} \log \left(\frac{2}{5} \right) + \frac{3}{5} \log \frac{3}{5} \right)$$

$$IG = \text{Dataset Entropy } (I')$$
$$- \text{Entropy } (<2.5) - \text{Entropy } (>2.5)$$

$$= I' - \frac{4}{6} \log \frac{4}{6} + \frac{2}{6} \log \frac{2}{6} + \frac{2}{5} \log \frac{2}{5} + \frac{3}{5} \log \frac{3}{5}$$

and $I' = - \left(\frac{6}{11} \log \frac{6}{11} + \frac{5}{11} \log \frac{5}{11} \right)$

```
                    slope < 1.5
                   /           \
                  /             No
               yes               \
               /              slope < 1.5
             yes              /          \
              0            yes            No
                          /                \
                      result = 1        result - 0
```