

Home Work-2 - Q3

Data

Height	Diameter	Weight	Type
0.0888	0.0980	0.5764	Ceramic
0.0789	0.0853	0.271	Metal
0.131	0.0704	0.5096	Ceramic
0.112	0.15	0.373	Plastic
0.116	0.119	0.33	Plastic
0.0914	0.060	0.1952	Metal

→ Here is the above data for every 2 data items

→ First we take height into consideration and sort it into ascending order and we get data

Height	Diameter	Weight	Type
0.0789	0.0853	0.271	Metal
0.0888	0.0980	0.5764	Ceramic
0.0914	0.060	0.1952	Metal
0.112	0.15	0.373	Plastic
0.116	0.119	0.33	Plastic
0.131	0.0704	0.5096	Ceramic

Gini Index

→ (height > 0.0888)

$$\rightarrow 1 - \sum (p_i)^2$$

$$= 1 - 1 = 0 \text{ Gini-L}$$

$$\text{Gini-R} = 1 - \sum (p_i)^2$$

$$= 1 - \left(\frac{1}{5}\right)^2 - \left(\frac{3}{5}\right)^2$$

$$= 0.64$$

$$IG = 1 - \frac{1}{6} \times 0 - \frac{5}{6} \times 0.64$$

$$= 0.4667$$

Height > 0.0789

$$p_1 \rightarrow 1, p_2 \rightarrow 1$$

$$p_1 \rightarrow 1, p_2 \rightarrow 1, p_3 \rightarrow 2$$

$$1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2$$

$$1 - \frac{1}{4} - \frac{1}{4} = \frac{1}{2}$$

$$1 - \left(\frac{1}{4}\right)^2 - \left(\frac{1}{4}\right)^2 - \left(\frac{1}{2}\right)^2$$

$$1 - \frac{1}{16} - \frac{1}{16} - \frac{1}{4}$$

$$1 - \frac{1}{8} - \frac{1}{4} = 1 - \frac{3}{8}$$

$$= \frac{5}{8}$$

$$IG_2 = 1 - \left(\frac{2}{6} \times \frac{1}{2}\right) - \left(\frac{4}{6} \times \frac{5}{8}\right)$$

$$= 1 - \frac{1}{6} - \frac{20}{48}$$

$$= 1 - 0.41667$$

$$= 0.41667$$

IG. 3

$$x > 0.0914$$

$$P_1 \rightarrow 1, P_2 \rightarrow 2$$

$$P_3 \rightarrow 2, P_1 \rightarrow 1$$

Gini-L

Gini-R

$$= 1 - \left(\frac{1}{3}\right)^2 - \left(\frac{2}{3}\right)^2$$

$$1 - \left(\frac{1}{3}\right)^2 - \left(\frac{1}{3}\right)^2$$

$$= \frac{9 - 1 - 4}{9} = \frac{4}{9}$$

$$\frac{4}{9}$$

$$J.G. = 1 - \frac{1 \times 1}{2 \times 9} - \frac{1 \times 1}{2 \times 9}$$

$$= 1 - \frac{2}{9} - \frac{2}{9} = \frac{5}{9} = 0.556$$

IG. 4

$$x > 0.112$$

$$P_3 \rightarrow 1, P_2 \rightarrow 2, P_1 \rightarrow 1$$

$$P_3 \rightarrow 1, P_1 \rightarrow 1$$

Gini-L

$$1 - \left(\frac{1}{4}\right)^2 - \left(\frac{2}{4}\right)^2 - \left(\frac{1}{4}\right)^2$$

$$1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2$$

$$= 1 - 0.125 - 0.25$$

$$1 - \frac{1}{2} = \frac{1}{2} = 0.5$$

$$= 0.625$$

$$= 1 - \frac{5}{6} \times 0.625 - \frac{2}{6} \times 0.5$$

$$= 1 - 0.166 - 0.520$$

$$= 0.313$$

IG 510 T

$\swarrow \nearrow$ D. 116
 $P_1 \rightarrow 1, P_2 \rightarrow 2, P_3 \rightarrow 2$

$P_1 \rightarrow 1$

Gini - L

Gini - R

$$= 1 - \left(\left(\frac{1}{5} \right)^2 + \left(\frac{2}{5} \right)^2 + \left(\frac{2}{5} \right)^2 \right)$$

$$1 - (1)^2 = 0$$

$$JG = 1 - \frac{5 \times 0.64}{6}$$

$$JG = 0.466$$

~~Thus from above all JG_3 is greatest hence our first division will be~~

Height

→ Now we will repeat the same with Diameter

Height	Diameter	Weight	Type
0.0989	0.060	0.1732	Metal
0.131	0.0704	0.5096	Ceramic
0.0789	0.0853	0.271	Metal
0.088	0.0980	0.5764	Ceramic
0.116	0.119	0.33	Plastic
0.112	0.15	0.393	Plastic

$D > 0.060$

$P_2 \rightarrow 1$

$P_1 \rightarrow 2, P_2 \rightarrow 1, P_3 \rightarrow 2$

$$JG = 0.4667$$

IG_2

$D > 0.0704$

$P_1 \rightarrow 1, P_2 \rightarrow 1$

$P_1 \rightarrow 1, P_2 \rightarrow 1, P_3 \rightarrow 2$

$IG_2 = 0.41663$

 IG_3

$D > 0.0853$

$P_1 \rightarrow 1, P_2 \rightarrow 2$

$P_1 \rightarrow 1, P_3 \rightarrow 2$

$IG_3 = 0.556$

 IG_4

$D > 0.0980$

$P_1 \rightarrow 2, P_2 \rightarrow 2$

$P_3 \rightarrow 2$

$Gini_L = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2$

$Gini_R = 0$

$= 1 - \frac{1}{2} - \frac{1}{2}$

$\therefore I.G = 1 - \frac{1}{2} - \frac{1}{2} = \frac{2}{3} = 0.66$

 IG_5

$D > 0.119$

$P_1 \rightarrow 2, P_2 \rightarrow 2, P_3 \rightarrow 1$

$P_3 \rightarrow 1$

$IG = 0.466$

Thus separation with Diameter > 0.0980 has more ^{information} than separation with height > 0.0914

Now we perform same with Weight

Height

Diameter

Weight

Type

0.0914

0.060

0.1952

Metal

0.0704

0.0853

0.271

Metal

0.116

0.119

0.33

Plastic

0.112

0.15

0.373

Plastic

0.131

0.0704

0.5096

Ceramic

0.088

0.0980

0.5764

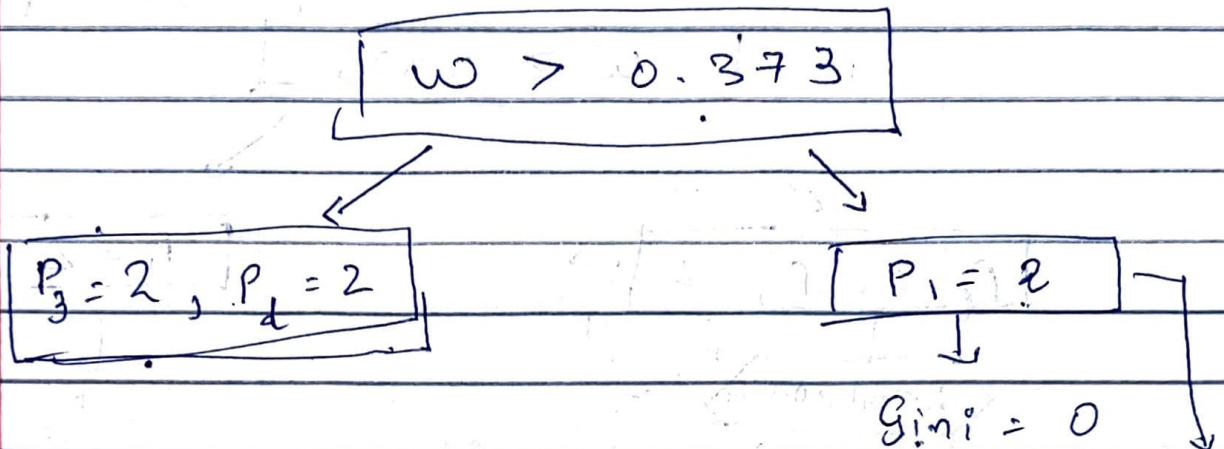
Ceramic

We can say that

$$w > 0.373$$

will have same of I.G of $D > 0.0980$

Thus we know perform the separation with $w > 0.373$ we get



Thus Now our Dataset is classified as ceramic

Height	Diameter	Type
0.0789	0.0853	Metal
0.0914	0.0980	Metal
0.112	0.15	Plastic
0.116	0.119	Plastic

Now Height > 0.0914



$$\therefore \text{Gini} = 0$$

$$\text{Gini} = 0$$

\therefore I.G \rightarrow Max

Thus next

separation will be height > 0.0914