Assignment-4 MDL

2018111033

DOMS Page No. Date / /

Q4

- Usual Greedy algo is followed.

| | Angle | Scorpio | Speed | (Topple) |
|---|---|---|---|---|
| 2 | 1.5 | N | 220 | N |
| — | 4.5 | Y | 120 | Y |
| 2 | 3 | N | 120 | Y |
| | 5.5 | N | 117 | Y |
| 2 | 3.2 | N | 170 | N |
| | 5.2 | Y | 90 | Y |
| 2 | 1.85 | N | 120 | N |
| | 4.8 | Y | 147 | Y |
| 2 | 1.7 | N | 100 | Y |

Y : 3, N : 6 Y : 6, N : 3

$$H(\text{Topple}) = B\left(\frac{2}{3}\right) = B\left(\frac{p}{p+n}\right)$$

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

$$= -\left(\frac{2}{3}\log_2 2 - \frac{2}{3}\log_2 3 + \frac{1}{3}\log 1\right)$$

($\log 1 \to 0$, $\therefore -\frac{1}{3}\log_2 3$)

$$= -\left(\frac{2}{3} - \log_2 3\right)$$

$$= 1.584 - \frac{2}{3} = 0.918$$

Scorpio

| Outlook | Topple (count) Yes | No | Entropy → H |
|---|---|---|---|
| Yes | 3 | 0 | 0 |
| No | 3 | 3 | 1 |

$I_f$ (Scorpio) $= \frac{3}{9}$ H (yes) $+ \frac{6}{9}$ H (no)

$= \frac{3}{9} \cdot 0 + \frac{6}{9}$

$= \frac{2}{3} = 0.667$

gain $= 0.918 - 0.667 = 0.251$

B − IG

Angle : Threshold $= \frac{3.2 + 4.5}{2} = \frac{7.7}{2} = 3.85$

| Outlook | Topple (count) Yes | No | Entropy → H |
|---|---|---|---|
| > 3.85 | 4 | 0 | 0 |
| ≤ 3.85 | 2 | 3 | |

$I_f$ (Angle) $= \frac{4}{9} \cdot 0 + \frac{5}{9}$ H (≤ 3.85)

$= \frac{5}{9}\left(\frac{2}{5}\log\frac{2}{5} + \frac{3}{5}\log\frac{3}{5}\right)$

$$= -\frac{5}{9}\left(\frac{-2}{5}\log_2 4 + \frac{-3}{5}\log 6\right)$$

$$= \frac{-5}{9}\left(\frac{-4}{5} - \frac{3 \times 2.584}{5}\right)$$

$$= -\frac{5}{9}(-0.528 - 0.442)$$

$$= \frac{0.970 \times 5}{9} = 0.539$$

gain = $0.918 - 0.539 = 0.379$

Speed

Speed is pretty varied so taking mean of all observations as split point (threshold)

Threshold = 134

| Speed | Topple (count) | | H |
|---|---|---|---|
| Outlook | Yes | No | |
| >134 | 1 | 2 | 0.918 |
| ≤134 | 5 | 1 | 0.650 |

~~Choosing binary value for entry 120<as<T~~

$H(>134) = 0.918$

$H(\leq 134) = -\left(\frac{5}{6}\log\frac{5}{6} + \frac{1}{6}\log\frac{1}{6}\right) = -(-0.219 + (-0.431)) = 0.650$

$I_f(\text{Speed}) = \frac{1}{3} \cdot 0.918 + \frac{2}{3} \cdot 0.650$

$= 0.306 + 0.433$

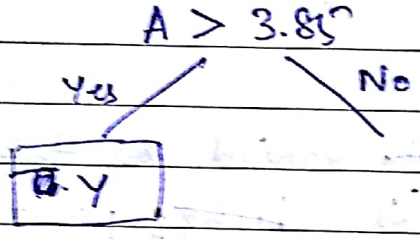
$= 0.739$

gain = $0.918 - 0.739 = 0.179$ :-

∴ Highest information gain is with angle

So we take root as angle.

Angle = A




| | Angle | Scorpio | Speed | Topple |
|---|---|---|---|---|
| | 1.5 | N | 220 | N |
| 3 | 3 | N | 120 | Y |
| | 3.2 | N | 170 | N |
| 3 | 1.85 | N | 120 | N |
| 3 | 1.7 | N | 100 | Y |

Ignoring scorpio now

$= B\left(\frac{2}{5}\right)$

$H(\text{Topple}) = 0.970$ (previous calc)

For angle,

let split point = $\frac{1.85+3}{2} = \frac{4.85}{2} \approx 2.43$

| Outlook | Topple (count) | | Entropy |
|---|---|---|---|
| | Yes | No | |
| > 2.43 | 1 | 1 | 1 |
| ≤ 2.43 | 1 | 2 | 0.918 |

→ (already calculated many times before)

$$If(Angle) = \frac{2}{5} \times 1 + \frac{3}{5} \times 0.918$$

$$= \frac{2 + 2.754}{5} = \frac{4.754}{5} = 0.951$$

$$gain = 0.970 - 0.951 = \boxed{0.019}$$

For speed

Let split point $= \frac{120+170}{2} = \frac{290}{2} = 145$

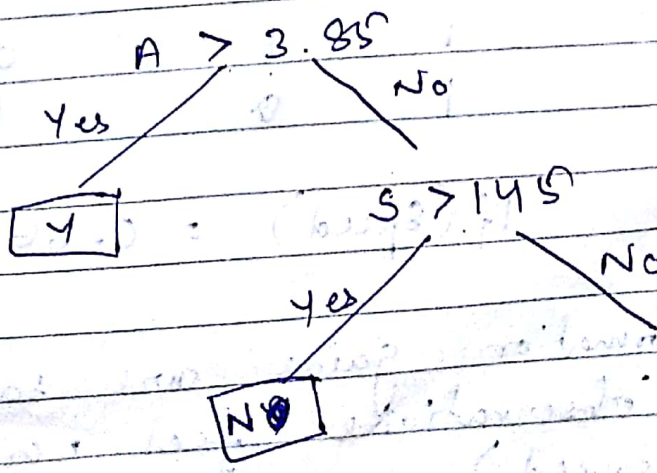
| Outlook | Topple (count) | | Entropy |
|---|---|---|---|
| | Yes | No | |
| > 145 | 0 | 2 | 0 |
| ≤ 145 | 2 | 1 | 0.918 |

$$If(speed) = \frac{2}{5} \times 0 + \frac{3}{5} \times 0.918 = \frac{2.754}{5} < If(angle)$$

$$gain = 0.97 - 0.551 = \boxed{0.419}$$

∴ we choose speed as breakpoint now

Speed (S)




Comment:

But this definitely feels like noisy data.

| Angle | Scorpio | Speed | Topple |
|---|---|---|---|
| 3 | N | 120 | Y |
| 1.85 | N | 120 | N |
| 1.7 | N | 100 | Y |

$H(\text{Topple}) = B\left(\frac{2}{3}\right)$

$= 0.918$

Angle: Let's take split point for angle as $\frac{1.85+3}{2}$

$= \frac{4.85}{2} = 2.43$

| Outlook | Topple (count) | | Entropy |
|---|---|---|---|
| | Yes | No | |
| $> 2.43$ | 1 | 0 | 0 |
| $\leq 2.43$ | 1 | 1 | 1 |

$If(\text{Angle}) = \frac{1}{3} \cdot 0 + \frac{2}{3} \cdot 1 = 0.667$

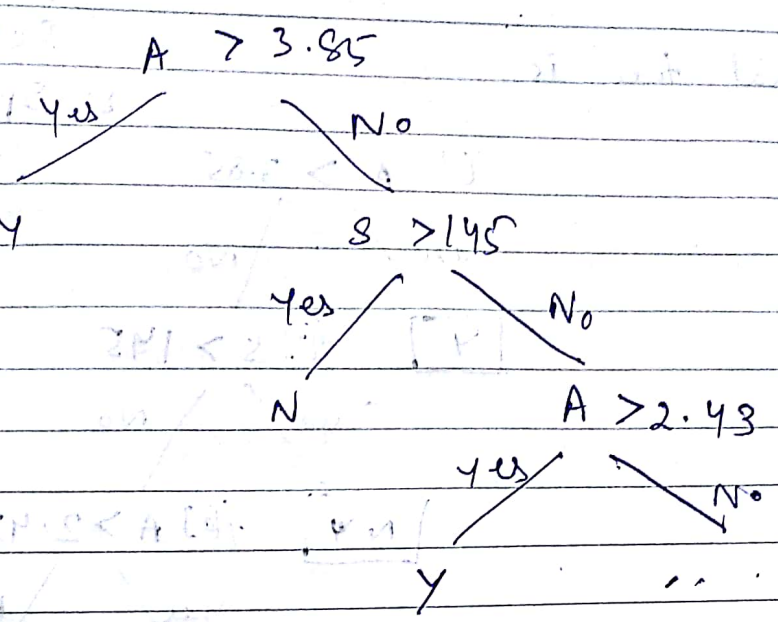
Speed

Let's take split point as 110

| Outlook | Topple (count) | | Entropy |
|---|---|---|---|
| | Yes | No | |
| $> 110$ | 1 | 1 | 0 |
| $\leq 110$ | 1 | 0 | 1 |

$If(\text{Speed}) = 0.667$

Both information gains are same but speed observations feel noisy (topple at low speed) so we will choose angle.




**Angle**

Let split = 1.77

| Outlook | Topple: Yes | Topple: No | Entropy |
|---|---|---|---|
| > 1.77 | 0 | 1 | 0 |
| ≤ 1.77 | 1 | 0 | 0 |

**Speed**

Let split = 110

| Outlook | Topple: Yes | Topple: No | Entropy |
|---|---|---|---|
| > 110 | 0 | 1 | 0 |
| ≤ 110 | 1 | 6 | 0 |

∵ entropies are same, we can choose either.

Comment: Both look like noise.

Choosing speed now (because angle has been selected more no of times

gain = H(topple) so it checks out.

Date / /

A : Take-off angle : f[0]
Scorpio : f[1]
S: Speed : f[2]

∴ final tree is