

Aditya Jain

Answer images for question 3 and 4

Answer 3

Ans 3

A	B	C	y
T	T	T	F
T	T	F	T
T	F	T	T
T	F	F	F
F	T	T	T
F	T	F	F
F	F	T	F
F	F	F	T

decision tree \Rightarrow

(2)

\Rightarrow decision tree

$1 - 0.667 = 0.333$ 0.0625 $0.1276 + 0.0075$
 0.00 0.60

Fever	Diarrhea	Stomach pain	class = +	class = -	total
Yes	Yes	Yes	8	2	10
Yes	Yes	No	6	2	8
Yes	No	Yes	4	0	4
No	Yes	Yes	5	3	8
No	No	Yes	1	8	9
No	No	No	1	10	11
			25	25	50 total

① Overall gini = $0.5 \left(\frac{25}{50} \right)$
 ②

Decision Tree 1: Split by Fever
 - Yes: $\begin{matrix} + = 18 \\ - = 4 \end{matrix}$
 - No: $\begin{matrix} + = 7 \\ - = 21 \end{matrix}$
 $g_1 = 1 - \left(\frac{18}{22} \right)^2 - \left(\frac{4}{22} \right)^2 = 0.29$ $g_2 = 1 - \left(\frac{7}{28} \right)^2 - \left(\frac{21}{28} \right)^2 = 0.375$
 $g_{\text{Fever}} = \frac{22 \times 0.29}{50} + \frac{28 \times 0.375}{50} = 0.1351$

Decision Tree 2: Split by Diarrhea
 - Yes: $\begin{matrix} + = 19 \\ - = 7 \end{matrix}$
 - No: $\begin{matrix} + = 6 \\ - = 18 \end{matrix}$
 $g_1 = 1 - \left(\frac{19}{26} \right)^2 - \left(\frac{7}{26} \right)^2 = 0.39$ $g_2 = 1 - \left(\frac{6}{24} \right)^2 - \left(\frac{18}{24} \right)^2 = 0.375$
 $g_{\text{Diarrhea}} = \frac{26 \times 0.39}{50} + \frac{24 \times 0.375}{50} = 0.3776$

Decision Tree 3: Split by Stomach pain
 - Yes: $\begin{matrix} + = 18 \\ - = 13 \end{matrix}$
 - No: $\begin{matrix} + = 7 \\ - = 12 \end{matrix}$
 $g_1 = 1 - \left(\frac{18}{31} \right)^2 - \left(\frac{13}{31} \right)^2 = 0.49$ $g_2 = 1 - \left(\frac{7}{19} \right)^2 - \left(\frac{12}{19} \right)^2 = 0.47$
 $g_{\text{Stomach pain}} = \frac{31 \times 0.49}{50} + \frac{19 \times 0.47}{50} = 0.4776$

③ g_{Fever} is the lowest = 0.1351
 hence splitting by fever. [fever as root node]

for left Node

0.73
0.6625
0.182

Decision Tree 1:

```

graph TD
    A((Fever)) -- Yes --> B[diarrhea]
    A -- No --> C[Stomach pain]
    B -- Yes --> D["+ = 14  
- = 4"]
    B -- No --> E["+ = 4  
- = 0"]
    C -- Yes --> F["+ = 12  
- = 2"]
    C -- No --> G["+ = 4  
- = 2"]
        
```

$g_1 = 1 - \left(\frac{14}{18}\right)^2 - \left(\frac{4}{18}\right)^2 = 0.35$
 $g_2 = 1 - \left(\frac{4}{4}\right)^2 - \left(\frac{0}{4}\right)^2 = 0$
 $J_+ = \frac{18 \times 0.35}{22} = 0.3$

Decision Tree 2:

```

graph TD
    A((Fever)) -- Yes --> B[Stomach pain]
    A -- No --> C[diarrhea]
    B -- Yes --> D["+ = 12  
- = 2"]
    B -- No --> E["+ = 4  
- = 2"]
    C -- Yes --> F["+ = 8  
- = 18"]
    C -- No --> G["+ = 2  
- = 18"]
        
```

$g_1 = 1 - \left(\frac{12}{14}\right)^2 - \left(\frac{2}{14}\right)^2 = 0.24$
 $g_2 = 1 - \left(\frac{8}{8}\right)^2 - \left(\frac{2}{8}\right)^2 = 0.37$
 $J_+ = \frac{14 \times 0.24}{22} + \frac{8 \times 0.37}{22} = 0.28$

Stomach pain gini index is smaller so assigning left node as stomach pain.

Decision Tree 3:

```

graph TD
    A((Fever)) -- Yes --> B[Stomach pain]
    A -- No --> C[diarrhea]
    B -- Yes --> D["+ = 1"]
    B -- No --> E["+ = 3"]
    C -- Yes --> F["+ = 5"]
    C -- No --> G["+ = 2  
- = 18"]
        
```

$g_1 = 1 - \left(\frac{5}{8}\right)^2 - \left(\frac{3}{8}\right)^2 = 0.46$
 $g_2 = 1 - \left(\frac{2}{20}\right)^2 - \left(\frac{18}{20}\right)^2 = 0.18$
 $J_+ = \frac{8 \times 0.46}{28} + \frac{20 \times 0.18}{28} = 0.25$

Decision Tree 4:

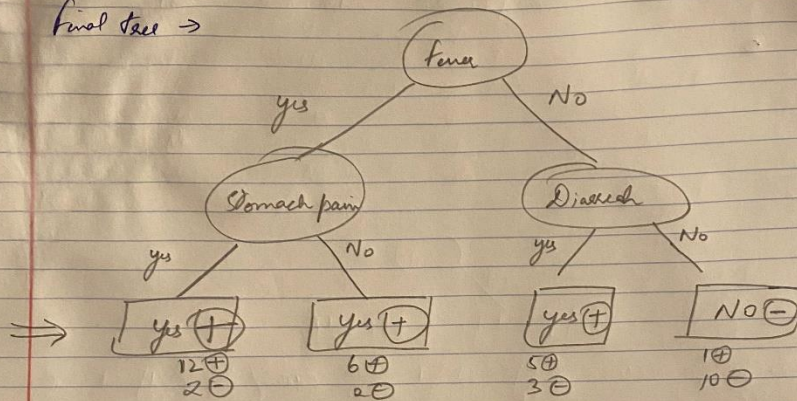
```

graph TD
    A((Fever)) -- Yes --> B[St pain]
    A -- No --> C[St pain]
    B -- Yes --> D["+ = 6  
- = 11"]
    B -- No --> E["+ = 10"]
    C -- Yes --> F["+ = 1"]
    C -- No --> G["+ = 10"]
        
```

$g_1 = 1 - \left(\frac{6}{17}\right)^2 - \left(\frac{11}{17}\right)^2 = 0.45$
 $g_2 = 1 - \left(\frac{1}{11}\right)^2 - \left(\frac{10}{11}\right)^2 = 0.16$
 $J_+ = \frac{17 \times 0.45}{28} + \frac{11 \times 0.16}{28} = 0.335$

gini of diarrhea is lower = 0.25

Final tree \rightarrow



training error $\Rightarrow \frac{2+2+3+1}{50} = 0.16 \text{ or } 16\%$

the decision tree will predict that the patient has the disease.
as all left child node predicts (+) action. is presence of disease.