Using Gini-Index to make a decision tree

Step 1)

Calculate Gini-index of Diagnosis

$$Gini(S) = 1 - \left[\left(\frac{5}{10}\right)^2 + \left(\frac{5}{10}\right)^2 \right] = 0.5$$

Step 2)

Calculate Gini-Index of all attributes individually.

a) Stiff neck

Stiff-Neck	Diagnosis
Y	P
N	P
Y	P
Y	P
Y	P
Y	N
N	N
N	N
N	N
N	N

if stiff neck = Yes
Positive = 4
Negative = 1

if stiff neck = No
Positive = 1
Negative = 4

Gini-index index Yes

$$1 - \left[\left(\frac{4}{5}\right)^2 + \left(\frac{1}{5}\right)^2 \right] = 0.32$$

Gini No

$$1 - \left[\left(\frac{1}{5}\right)^2 + \left(\frac{4}{5}\right)^2 \right] = 0.32$$

Weighted average of stiff-neck $= 0.32 \times (5/10) + 0.32(5/10) = 0.32$

b) Spots

Spots	Diagnosis
Y	P
Y	P
N	P
Y	P
Y	P
N	N
Y	N
Y	N
N	N
N	N

SPOTS = Y
P = 4, N = 2

SPOTS = N
P = 1, N = 3

Gini Yes

$$1 - \left[\left(\frac{4}{6}\right)^2 + \left(\frac{2}{6}\right)^2 \right] = \frac{4}{9}$$

Gini = NO

$$1 - \left[\left(\frac{1}{4}\right)^2 + \left(\frac{3}{4}\right)^2 \right] = \frac{3}{8}$$

Total weighted avg(Spots)

$$= \frac{4}{9} \times \left(\frac{6}{10}\right) + \frac{3}{8} \times \left(\frac{4}{10}\right) = \frac{5}{12}$$

$$= 0.41666\ldots$$

c) Headaches

Headaches	Diagnosis
Y	P
N	P
Y	P
N	P
Y	P
N	N
N	N
N	N
Y	N
N	N

Headaches = Y
P = 3, N = 1

Headaches = N
P = 2, N = 4

Gini-index Y

$$1-\left[\left(\frac{3}{4}\right)^2+\left(\frac{1}{4}\right)^2\right]=\frac{3}{8}$$

Gini-index

$$1-\left[\left(\frac{2}{6}\right)^2+\left(\frac{4}{6}\right)^2\right]=\frac{4}{9}$$

$$\frac{3}{8}\times\frac{4}{10}+\frac{4}{9}\times\left(\frac{6}{10}\right)=\frac{5}{12}$$

$$=0.41666$$

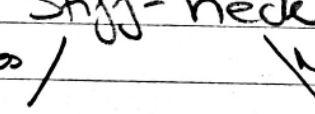
Step 3)

lowest Gini-index of attributs becomes root node

Headaches = 0.41666...
Spots = = 0.41666...
Stiff-neck = 0.32 ✓

Step 4)

split by Stiff neck



Cid	H	Sp	S	D
1	Y	Y	Y	P
3	Y	N	Y	P
4	N	Y	Y	P
5	Y	Y	Y	P
6	N	N	Y	N

Cid	H	Sp	S	D
2	N	Y	N	P
7	N	Y	N	N
8	N	Y	N	N
9	Y	N	N	N
10	N	N	N	N

Step(5)

Stiff-neck = N

Do Gini on No table

C	H	Sp	S	D
2	N	Y	N	P
7	N	Y	N	N
8	N	Y	N	N
9	Y	N	N	N
10	N	N	N	N

Split by all attributes except stiffneck as we have already calculated it

Gini-index D

P=1 , N=4

$$1-\left[\left(\frac{1}{5}\right)^2+\left(\frac{4}{5}\right)^2\right]=0.32$$

Gini-index ~~Stiff~~ Spots

Sp	D
Y	P
Y	N
Y	N
N	N
N	N

Sp = Y

P = 1 , N = 2

Sp = N

P = 0 , N = 2

$$Y = 1-\left[\left(\frac{1}{3}\right)^2+\left(\frac{2}{3}\right)^2\right]=4/9$$

$$N = 1-\left[\left(\frac{2}{2}\right)^2\right)=0$$

$$\frac{4}{9}\times\left(\frac{3}{5}\right)+0\times\left(\frac{2}{5}\right)=\frac{4}{15}$$

W(avg) = 0.2666 ....

Gini-index - Headache

H	D
N	P
N	N
N	N
Y	N
N	N

H = N
P = 1, N = 3

H = Y
P = 0, N = 1

Gini N

$$1 - \left(\left(\frac{1}{4}\right)^2 + \left(\frac{3}{4}\right)^2\right) = \frac{3}{8}$$

Gini Y

$$1 - \left(\left(\frac{1}{1}\right)^2\right) = 0$$

$$\frac{3}{8} \times \left(\frac{4}{5}\right) + 0 \times \left(\frac{1}{5}\right) = \frac{3}{10} = 0.3$$

Gini's

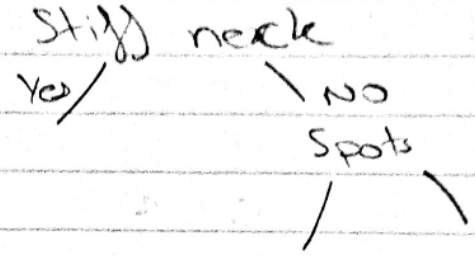
D = 0.32

H = 0.3

Sp = 0.266...

Step 6)

Split No section on spots



Spots yes

id	H	Sp	S	D
2	N	Y	N	D
7	N	Y	N	N
8	N	Y	N	N

Yes spots N

id	H	Sp	S	D
9	Y	N	N	N
10	N	N	N	N

No

4

Stop here

Spots yes

D

$= P=1, N=2$

$$1-\left[\left(\frac{1}{3}\right)^2+\left(\frac{2}{3}\right)^2\right]=\frac{4}{9}$$

S, Sp, H

$N = P=1, N=2, Y=0$ all outputs $= \frac{4}{9}$

$$N_{Sin}=1-\left(\left(\frac{1}{3}\right)^2+\left(\frac{2}{3}\right)^2\right)=\frac{4}{9}$$

$H_N = P=1, N=2$

$$N=1-\left(\left(\frac{1}{3}\right)^2+\left(\frac{2}{3}\right)^2\right)=\frac{4}{9}$$

Step 7

if stiff-neck = Yes

Cid	H	Sp	S	D
1	Y	Y	Y	P
3	Y	N	Y	P
4	N	Y	Y	P
5	Y	Y	Y	P
6	N	N	Y	N

D_{Gini}, P = 4, N = 1

$1-\left(\left(\frac{4}{5}\right)^2+\left(\frac{1}{5}\right)^2\right) = 0.32$

Sp gini,

Y, P = 3, N, P = 1, N = 1

$P = 1-\left(\left(\frac{3}{3}\right)^2\right) = 0$ $N = 1-\left(\left(\frac{1}{2}\right)^2+\left(\frac{1}{2}\right)^2\right) = 0.5$

$0 \times \left(\frac{3}{5}\right) + 0.5 \times \left(\frac{2}{5}\right) = \frac{1}{5} = 0.2$

H gini,

Y, P = 3 N, P = 1, N = 1

$P = 1-\left(\left(\frac{3}{3}\right)^2\right) = 0$ $N = 1-\left(\left(\frac{1}{2}\right)^2+\left(\frac{1}{2}\right)^2\right) = 0.5$

$0 \times \left(\frac{3}{5}\right) = 0 + 0.5 \times \left(\frac{2}{5}\right) = \frac{1}{5} = 0.2$

Split pH

~~Cid~~

Head

Yes

Cid	H	Sp	S	D
1	Y	Y	Y	P
3	Y	N	Y	P
5	Y	Y	Y	P

Stop

No

Cid	H	Sp	S	D
4	N	Y	Y	P
6	N	N	Y	N

Split Sp

Yes

Cid	H	Sp	S	D
1	Y	Y	Y	P
4	N	Y	Y	P
5	Y	Y	Y	P

Cid	H	Sp	S	D
3	Y	N	Y	P
6	N	N	Y	N

Gini Decision tree