

Machine Learning Coursework

Part 1

Notes:

+ : Yes

– : No

Gini Index

Root Node:

Root Node (14) : 7 Positive, 7 Negative

Headache+ (6) : 4 Positive, 2 Negative ; Headache– (8) : 3 Positive, 5 Negative

Spots+ (8) : 5 Positive, 3 Negative ; Spots– (6) : 2 Positive, 4 Negative

Stiff Neck+ (8) : 6 Positive, 2 Negative ; Stiff Neck– (6) : 1 Positive, 5 Negative

$$\begin{aligned}I(\text{Root}) &= 1 - (7/14)^2 - (7/14)^2 \\ &= 0.5\end{aligned}$$

$$\begin{aligned}I(\text{Headache+}) &= 1 - (4/6)^2 - (2/6)^2 \\ &= 4/9\end{aligned}$$

$$\begin{aligned}I(\text{Headache–}) &= 1 - (3/8)^2 - (5/8)^2 \\ &= 0.46875\end{aligned}$$

$$\begin{aligned}I(\text{Spots+}) &= 1 - (5/8)^2 - (3/8)^2 \\ &= 0.46875\end{aligned}$$

$$\begin{aligned}I(\text{Spots–}) &= 1 - (2/6)^2 - (4/6)^2 \\ &= 4/9\end{aligned}$$

$$\begin{aligned}I(\text{Stiff Neck+}) &= 1 - (6/8)^2 - (2/8)^2 \\ &= 0.375\end{aligned}$$

$$\begin{aligned}I(\text{Stiff Neck–}) &= 1 - (1/6)^2 - (5/6)^2 \\ &= 5/18\end{aligned}$$

$$\begin{aligned}\text{Gain}(\text{Root, Headache}) &= 0.5 - (6/14)(4/9) - (8/14)(0.46875) \\ &= 0.04167\end{aligned}$$

$$\begin{aligned}\text{Gain}(\text{Root, Spots}) &= 0.5 - (8/14)(0.46875) - (6/14)(4/9) \\ &= 0.04167\end{aligned}$$

$$\begin{aligned}\text{Gain}(\text{Root, Stiff Neck}) &= 0.5 - (8/14)(0.375) - (6/14)(5/18) \\ &= 0.1667\end{aligned}$$

*** Split at Stiff Neck**

Stiff Neck+:

Stiff Neck+ (8) : 6 Positive, 2 Negative

Headache+ (4) : 4 Positive, 0 Negative ; Headache- (4) : 2 Positive, 2 Negative

Spots+ (5) : 4 Positive, 1 Negative ; Spots- (3) : 2 Positive, 1 Negative

$$I(\text{Stiff Neck+}) = 1 - (6/8)^2 - (2/8)^2 \\ = 0.375$$

$$I(\text{Headache+}) = 1 - (4/4)^2 - 0 \\ = 0$$

$$I(\text{Headache-}) = 1 - (2/4)^2 - (2/4)^2 \\ = 0.5$$

$$I(\text{Spots+}) = 1 - (4/5)^2 - (1/5)^2 \\ = 0.32$$

$$I(\text{Spots-}) = 1 - (2/3)^2 - (1/3)^2 \\ = 4/9$$

$$\text{Gain}(\text{Root, Headache}) = 0.375 - 0 - (4/8)(0.5) \\ = 0.125$$

$$\text{Gain}(\text{Root, Spots}) = 0.375 - (5/8)(0.32) - (3/8)(4/9) \\ = 0.008333$$

*** Stiff Neck+ split at Headache**

Stiff Neck-:

Stiff Neck- (6) : 1 Positive, 5 Negative

Headache+ (2) : 0 Positive, 2 Negative ; Headache- (4) : 1 Positive, 3 Negative

Spots+ (3) : 1 Positive, 2 Negative ; Spots- (3) : 0 Positive, 3 Negative

$$I(\text{Stiff Neck-}) = 1 - (1/6)^2 - (5/6)^2 \\ = 5/18$$

$$I(\text{Headache+}) = 1 - 0 - (2/2)^2 \\ = 0$$

$$I(\text{Headache-}) = 1 - (1/4)^2 - (3/4)^2 \\ = 0.375$$

$$I(\text{Spots+}) = 1 - (1/3)^2 - (2/3)^2 \\ = 4/9$$

$$I(\text{Spots-}) = 1 - 0 - (3/3)^2 \\ = 0$$

$$\text{Gain}(\text{Root, Headache}) = 5/18 - 0 - (4/6)(0.375) \\ = 0.02778$$

$$\text{Gain}(\text{Root, Spots}) = 5/18 - (3/6)(4/9) - 0 \\ = 0.0556$$

*** Stiff Neck- split at Spots**

Stiff Neck+ -> Headache+:

Headache+ (4): 4 Positive, 0 Negative

Spots+ (2) : 2 Positive, 0 Negative; Spots- (2) : 2 Positive, 0 Negative

*** Stop at Headache+**

**** No Negative diagnosis for Headache+**

Stiff Neck+ -> Headache- Split:

Headache- (4): 2 Positive, 2 Negative

Spots+ (3) : 2 Positive, 1 Negative ; Spots- (1) : 0 Positive, 1 Negative

$$\begin{aligned}I(\text{Headache-}) &= 1 - (2/4)^2 - (2/4)^2 \\ &= 0.5\end{aligned}$$

$$\begin{aligned}I(\text{Spots+}) &= 1 - (2/3)^2 - (1/3)^2 \\ &= 4/9\end{aligned}$$

$$\begin{aligned}I(\text{Spots-}) &= 1 - 0 - (1/1)^2 \\ &= 0\end{aligned}$$

*** Split at Spots and Stop**

**** No more attributes**

Stiff Neck- -> Spots+:

Spots+ (3) : 1 Positive, 2 Negative

Headache- (3): 1 Positive, 2 Negative

$$\begin{aligned}I(\text{Spots+}) &= 1 - (1/3)^2 - (2/3)^2 \\ &= 4/9\end{aligned}$$

$$\begin{aligned}I(\text{Headache-}) &= 1 - (1/3)^2 - (2/3)^2 \\ &= 4/9\end{aligned}$$

*** Split at Headache and Stop**

**** No more attributes**

Stiff Neck- -> Spots-:

Spots- (3) : 0 Positive, 3 Negative

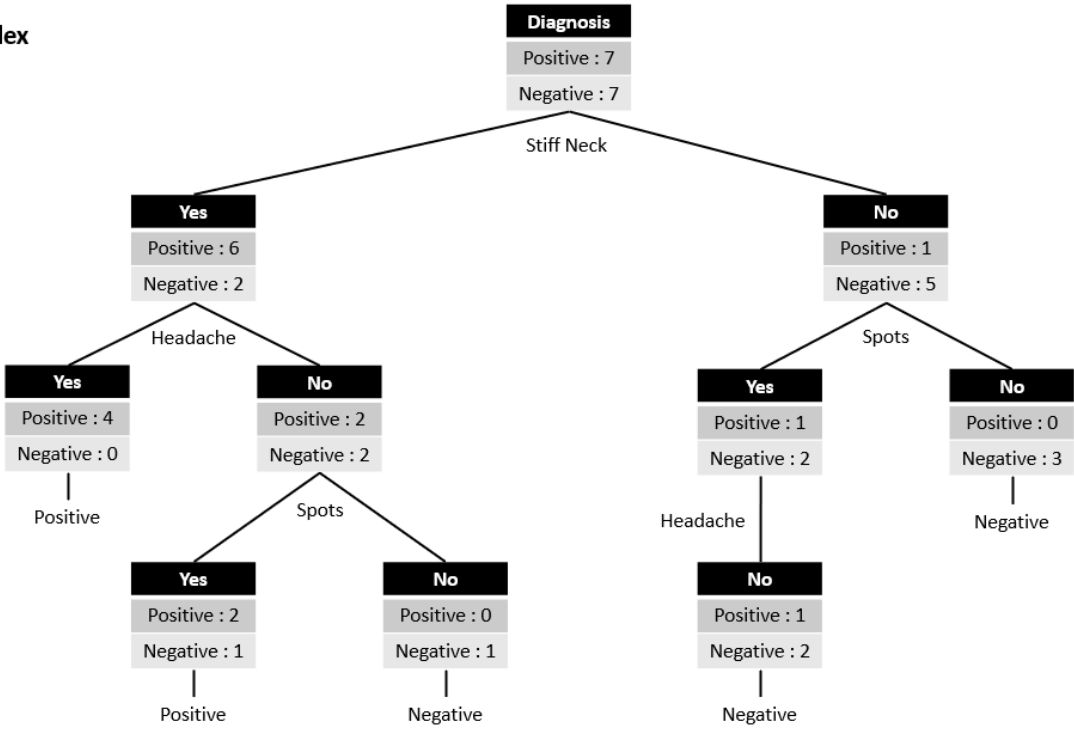
Headache+ (2): 0 Positive, 2 Negative

Headache- (1): 0 Positive, 1 Negative

*** Stop at Spots-**

**** No Positive diagnosis for Spots-**

Gini Index



Chi-squared Statistic

Root Node:

Root Node (14) : 7 Positive, 7 Negative

Headache+ (6) : 4 Positive, 2 Negative ; Headache- (8) : 3 Positive, 5 Negative

Spots+ (8) : 5 Positive, 3 Negative ; Spots- (6) : 2 Positive, 4 Negative

Stiff Neck+ (8) : 6 Positive, 2 Negative ; Stiff Neck- (6) : 1 Positive, 5 Negative

$$P(\text{Root} = \text{Positive}) = \frac{7}{14}$$

$$P(\text{Root} = \text{Negative}) = \frac{7}{14}$$

Headache :

Headache	Positive [Observed (Expected)]	Negative [Observed (Expected)]
Yes	4 (6 * $\frac{7}{14} = 3$)	2 (6 * $\frac{7}{14} = 3$)
No	3 (8 * $\frac{7}{14} = 4$)	5 (8 * $\frac{7}{14} = 4$)

$$\begin{aligned} \chi^2 &= \frac{(4 - 3)^2}{3} + \frac{(2 - 3)^2}{3} + \frac{(3 - 4)^2}{4} + \frac{(5 - 4)^2}{4} \\ &= 1.667 \end{aligned}$$

Spots :

Spots	Positive [Observed (Expected)]	Negative [Observed (Expected)]
Yes	5 (8 * $\frac{7}{14} = 4$)	3 (8 * $\frac{7}{14} = 4$)
No	2 (6 * $\frac{7}{14} = 3$)	4 (6 * $\frac{7}{14} = 3$)

$$\begin{aligned} \chi^2 &= \frac{(5 - 4)^2}{4} + \frac{(3 - 4)^2}{4} + \frac{(2 - 3)^2}{3} + \frac{(4 - 3)^2}{3} \\ &= 1.667 \end{aligned}$$

Stiff Neck :

Stiff Neck	Positive [Observed (Expected)]	Negative [Observed (Expected)]
Yes	6 (8 * $\frac{7}{14} = 4$)	2 (8 * $\frac{7}{14} = 4$)
No	1 (6 * $\frac{7}{14} = 3$)	5 (6 * $\frac{7}{14} = 3$)

$$\begin{aligned} \chi^2 &= \frac{(6 - 4)^2}{4} + \frac{(2 - 4)^2}{4} + \frac{(1 - 3)^2}{3} + \frac{(5 - 3)^2}{3} \\ &= 4.667 \end{aligned}$$

* Split at Stiff Neck

Stiff Neck+:

Stiff Neck+ (8) : 6 Positive, 2 Negative

Headache+ (4) : 4 Positive, 0 Negative ; Headache- (4) : 2 Positive, 2 Negative

Spots+ (5) : 4 Positive, 1 Negative ; Spots- (3) : 2 Positive, 1 Negative

$$P(\text{Stiff Neck+} = \text{Positive}) = \frac{6}{8}$$

$$P(\text{Stiff Neck+} = \text{Negative}) = \frac{2}{8}$$

Headache :

Headache	Positive [Observed (Expected)]	Negative [Observed (Expected)]
Yes	4 (4 * $\frac{6}{8} = 3$)	0 (4 * $\frac{2}{8} = 1$)
No	2 (4 * $\frac{6}{8} = 3$)	2 (4 * $\frac{2}{8} = 1$)

$$\begin{aligned} \chi^2 &= \frac{(4 - 3)^2}{3} + \frac{(0 - 1)^2}{1} + \frac{(2 - 3)^2}{3} + \frac{(2 - 1)^2}{1} \\ &= 2.667 \end{aligned}$$

Spots :

Spots	Positive [Observed (Expected)]	Negative [Observed (Expected)]
Yes	4 (5 * $\frac{6}{8} = 3.75$)	1 (5 * $\frac{2}{8} = 1.25$)
No	2 (3 * $\frac{6}{8} = 2.25$)	1 (3 * $\frac{2}{8} = 0.75$)

$$\begin{aligned} \chi^2 &= \frac{(4 - 3.75)^2}{3.75} + \frac{(1 - 1.25)^2}{1.25} + \frac{(2 - 2.25)^2}{2.25} + \frac{(1 - 0.75)^2}{0.75} \\ &= 0.1778 \end{aligned}$$

* Stiff Neck+ split at Headache

Stiff Neck–:

Stiff Neck– (6) : 1 Positive, 5 Negative

Headache+ (2) : 0 Positive, 2 Negative ; Headache– (4) : 1 Positive, 3 Negative

Spots+ (3) : 1 Positive, 2 Negative ; Spots– (3) : 0 Positive, 3 Negative

$$P(\text{Stiff Neck–} = \text{Positive}) = \frac{1}{6}$$

$$P(\text{Stiff Neck–} = \text{Negative}) = \frac{5}{6}$$

Headache :

Headache	Positive [Observed (Expected)]	Negative [Observed (Expected)]
Yes	0 ($2 * \frac{1}{6} = 0.333$)	2 ($2 * \frac{5}{6} = 1.67$)
No	1 ($4 * \frac{1}{6} = 0.667$)	3 ($4 * \frac{5}{6} = 3.33$)

$$\begin{aligned} \chi^2 &= \frac{(0 - 0.333)^2}{0.333} + \frac{(2 - 1.67)^2}{1.67} + \frac{(1 - 0.667)^2}{0.667} + \frac{(3 - 3.33)^2}{3.33} \\ &= 0.5971 \end{aligned}$$

Spots :

Spots	Positive [Observed (Expected)]	Negative [Observed (Expected)]
Yes	1 ($3 * \frac{1}{6} = 0.5$)	2 ($3 * \frac{5}{6} = 2.5$)
No	0 ($3 * \frac{1}{6} = 0.5$)	3 ($3 * \frac{5}{6} = 2.5$)

$$\begin{aligned} \chi^2 &= \frac{(1 - 0.5)^2}{0.5} + \frac{(2 - 2.5)^2}{2.5} + \frac{(0 - 0.5)^2}{0.5} + \frac{(3 - 2.5)^2}{2.5} \\ &= 1.2 \end{aligned}$$

* Stiff Neck– split at Spots

Stiff Neck+ -> Headache+:

Headache+ (4): 4 Positive, 0 Negative

Spots+ (2) : 2 Positive, 0 Negative; Spots- (2) : 2 Positive, 0 Negative

*** Stop at Headache+**

**** No Negative diagnosis for Headache+**

Stiff Neck+ -> Headache- Split:

Headache- (4): 2 Positive, 2 Negative

Spots+ (3) : 2 Positive, 1 Negative ; Spots- (1) : 0 Positive, 1 Negative

*** Split at Spots and Stop**

**** No more attributes**

Stiff Neck- -> Spots+:

Spots+ (3) : 1 Positive, 2 Negative

Headache- (3): 1 Positive, 2 Negative

$$I(\text{Spots+}) = 1 - (1/3)^2 - (2/3)^2 \\ = 4/9$$

$$I(\text{Headache-}) = 1 - (1/3)^2 - (2/3)^2 \\ = 4/9$$

*** Split at Headache and Stop**

**** No more attributes**

Stiff Neck- -> Spots-:

Spots- (3) : 0 Positive, 3 Negative

Headache+ (2): 0 Positive, 2 Negative

Headache- (1): 0 Positive, 1 Negative

*** Stop at Spots-**

**** No Positive diagnosis for Spots-**

Chi-squared Statistic

