Decision Tree – Entropy/Information Gain

Necessary Formulas:

1. Entropy, $E = -\sum p_i \log_2 p_i$; i = 1 to k, where k = number of classes.
2. Average Entropy, $E_{New} = \sum w_i E_i$; i = 1 to n, where n = number of unique values for an attribute.
3. Information Gain, $I_g = E_{Start} - E_{New}$

Iteration 1 (For Selecting the Root Node)

We have 3 classes. So, The Value of Initial Entropy, E_{Start} will be:

$$E_{Start} = -p_1 \log_2 p_1 - p_2 \log_2 p_2 - p_3 \log_2 p_3$$

There are 4 instances with classification 1, 5 instances with classification 2 and 15 instances with classification 3. So, $p_1 = (4/24)$, $p_2 = (5/24)$ and $p_3 = (15/24)$.

$$\begin{aligned} E_{Start} &= -(4/24) \log_2 (4/24) - (5/24) \log_2 (5/24) - (15/24) \log_2 (15/24) \\ &= 0.4308 + 0.4715 + 0.4238 \\ &= 1.3261 \text{ bits} \end{aligned}$$

Now, we need to calculate E_{New} for each of the attributes.

Calculating Entropy for different Values of Age

For Age = 1,

$$\begin{aligned} E_1 &= -(2/8) \log_2 (2/8) - (2/8) \log_2 (2/8) - (4/8) \log_2 (4/8) \\ &= 0.5 + 0.5 + 0.5 \\ &= 1.5 \end{aligned}$$

For Age = 2,

$$\begin{aligned} E_2 &= -(1/8) \log_2 (1/8) - (2/8) \log_2 (2/8) - (5/8) \log_2 (5/8) \\ &= 0.375 + 0.5 + 0.4238 \\ &= 1.2988 \end{aligned}$$

For Age = 3,

$$\begin{aligned} E_3 &= -(1/8) \log_2 (1/8) - (1/8) \log_2 (1/8) - (6/8) \log_2 (6/8) \\ &= 0.375 + 0.375 + 0.3113 \\ &= 1.0613 \end{aligned}$$

E_{New} (Age) = $(8/24)\, E_1 + (8/24)\, E_2 + (8/24)\, E_3$ = 1.2867 bits

Information Gain, I_g (Age) = $E_{Start} - E_{New}$ (Age) = 1.3261 – 1.2867 = 0.0394 bits

Calculating Entropy for different Values of SpecRx

For SpecRx = 1,

$$\begin{aligned} E_1 &= -(3/12) \log_2 (3/12) - (2/12) \log_2 (2/12) - (7/12) \log_2 (7/12) \\ &= 0.5 + 0.4308 + 0.4536 \\ &= 1.3844 \end{aligned}$$

For SpecRx = 2,

$$\begin{aligned} E_2 &= -(1/12) \log_2 (1/12) - (3/12) \log_2 (3/12) - (8/12) \log_2 (8/12) \\ &= 0.2988 + 0.5 + 0.3900 \\ &= 1.1887 \end{aligned}$$

E_{New} (SpecRx) = $(12/24)\, E_1 + (12/24)\, E_2$ = 1.2866 bits

Information Gain, I_g (SpecRx) = $E_{Start} - E_{New}$ (SpecRx) = 0.0395 bits

Calculating Entropy for different Values of Astig

For Astig = 1,

$$\begin{aligned} E_1 &= -(0/12) \log_2 (0/12) - (5/12) \log_2 (5/12) - (7/12) \log_2 (7/12) \\ &= 0 + 0.5263 + 0.4536 \\ &= 0.9799 \end{aligned}$$

For Astig = 2,

$$\begin{aligned} E_2 &= -(4/12) \log_2 (4/12) - (0/12) \log_2 (0/12) - (8/12) \log_2 (8/12) \\ &= 0.5283 + 0 + 0.3900 \\ &= 0.9183 \end{aligned}$$

E_{New} (Astig) = (12/24) E_1 + (12/24) E_2 = 0.9491 bits
Information Gain, Ig (Astig) = E_{Start} - E_{New} (Astig) = 0.377 bits

Calculating Entropy for different Values of Tears

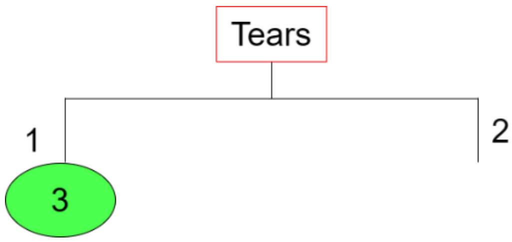
For Tears = 1,

$$\begin{aligned} E_1 &= -(0/12)\log_2(0/12) \\ &\quad -(0/12)\log_2(5/12) \\ &\quad -(12/12)\log_2(12/12) \\ &= 0 + 0 + 0 \\ &= 0 \end{aligned}$$

For Tears = 2,

$$\begin{aligned} E_2 &= -(4/12)\log_2(4/12) \\ &\quad -(5/12)\log_2(5/12) \\ &\quad -(3/12)\log_2(3/12) \\ &= 0.5283 + 0.5263 + 0.5 \\ &= 1.5546 \end{aligned}$$

E_{New} (Tears) = (12/24) E_1 + (12/24) E_2 = 0.7773 bits
Information Gain, Ig (Tears) = E_{Start} - E_{New} (Tears) = 0.5488 bits



Iteration 2 (For Branch Tears = 2)

There are 4 instances with classification 1, 5 instances with classification 2 and 3 instances with classification 3. So, p_1 = (4/12), p_2 = (5/12) and p_3 = (3/12).

$$\begin{aligned} E_{Start} &= -(4/12)\log_2(4/12) - (5/12)\log_2(5/12) - (3/12)\log_2(3/12) \\ &= 0.5283 + 0.5263 + 0.5 \\ &= 1.5546 \text{ bits} \end{aligned}$$

Now, we need to calculate E_{New} for each of the attributes.

Calculating Entropy for different Values of Age

For Age = 1,

$$\begin{aligned} E_1 &= -(2/4)\log_2(2/4) \\ &\quad -(2/4)\log_2(2/4) \\ &\quad -(0/4)\log_2(0/4) \\ &= 0.5 + 0.5 + 0 \\ &= 1.0 \end{aligned}$$

For Age = 2,

$$\begin{aligned} E_2 &= -(1/4)\log_2(1/4) \\ &\quad -(2/4)\log_2(2/4) \\ &\quad -(1/4)\log_2(1/4) \\ &= 0.5 + 0.5 + 0.5 \\ &= 1.5 \end{aligned}$$

For Age = 3,

$$\begin{aligned} E_3 &= -(1/4)\log_2(1/4) \\ &\quad -(1/4)\log_2(1/4) \\ &\quad -(2/4)\log_2(2/4) \\ &= 0.5 + 0.5 + 0.5 \\ &= 1.5 \end{aligned}$$

E_{New} (Age) = (4/12) E_1 + (4/12) E_2 + (4/12) E_3 = 1.2867 bits
Information Gain, Ig (Age) = E_{Start} - E_{New} (Age) = 1.5546 – 1.3333 = 0.2213 bits

Calculating Entropy for different Values of SpecRx

For SpecRx = 1,

$$\begin{aligned} E_1 &= -(3/6)\log_2(3/6) \\ &\quad -(2/6)\log_2(2/6) \\ &\quad -(1/6)\log_2(1/6) \\ &= 0.5 + 0.5283 + 0.4308 \\ &= 1.4591 \end{aligned}$$

For SpecRx = 2,

$$\begin{aligned} E_2 &= -(1/6)\log_2(1/6) \\ &\quad -(3/6)\log_2(3/6) \\ &\quad -(2/6)\log_2(2/6) \\ &= 0.4308 + 0.5 + 0.5283 \\ &= 1.4591 \end{aligned}$$

E_{New} (SpecRx) = (6/12) E_1 + (6/12) E_2 = 1.4591 bits
Information Gain, Ig (SpecRx) = E_{Start} - E_{New} (SpecRx) = 0.096 bits

Calculating Entropy for different Values of Astig

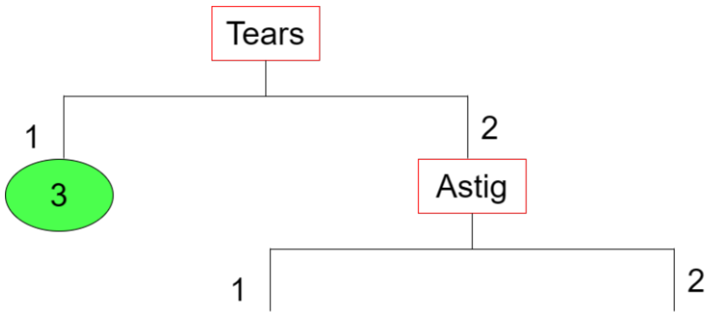
For Astig = 1,

$$\begin{aligned} E_1 &= -(0/6)\log_2(0/6) - (5/6)\log_2(5/6) - (1/6)\log_2(1/6) \\ &= 0 + 0.2192 + 0.4308 \\ &= 0.6500 \end{aligned}$$

For Astig = 2,

$$\begin{aligned} E_2 &= -(4/6)\log_2(4/6) - (0/6)\log_2(0/6) - (2/6)\log_2(2/6) \\ &= 0.3900 + 0 + 0.5283 \\ &= 0.9183 \end{aligned}$$

E_{New} (Astig) = (6/12) E_1 + (6/12) E_2 = 0.7842 bits
Information Gain, Ig (Astig) = E_{Start} - E_{New} (Astig) = 0.7704 bits



Iteration 3 (For Branch Astig = 1)

There are 5 instances with classification 2 and 1 instance with classification 3. So, $p_1 = (5/6)$ and $p_2 = (1/6)$.

$$\begin{aligned} E_{Start} &= -(5/6)\log_2(5/6) - (1/6)\log_2(1/6) \\ &= 0.2192 + 0.4308 \\ &= 0.65 \text{ bits} \end{aligned}$$

Now, we need to calculate E_{New} for each of the attributes.

Calculating Entropy for different Values of Age

For Age = 1,

$$\begin{aligned} E_1 &= -(0/2)\log_2(0/2) - (2/2)\log_2(2/2) - (0/2)\log_2(0/2) \\ &= 0 + 0 + 0 \\ &= 0 \end{aligned}$$

For Age = 2,

$$\begin{aligned} E_2 &= -(0/2)\log_2(0/2) - (2/2)\log_2(2/2) - (0/2)\log_2(0/2) \\ &= 0 + 0 + 0 \\ &= 0 \end{aligned}$$

For Age = 3,

$$\begin{aligned} E_3 &= -(0/2)\log_2(0/2) - (1/2)\log_2(1/2) - (1/2)\log_2(1/2) \\ &= 0 + 0.5 + 0.5 \\ &= 1 \end{aligned}$$

E_{New} (Age) = (2/6) E_1 + (2/6) E_2 + (2/6) E_3 = 0.3333 bits
Information Gain, Ig (Age) = E_{Start} - E_{New} (Age) = 0.6500 – 0.3333 = 0.3167 bits

Calculating Entropy for different Values of SpecRx

For SpecRx = 1,

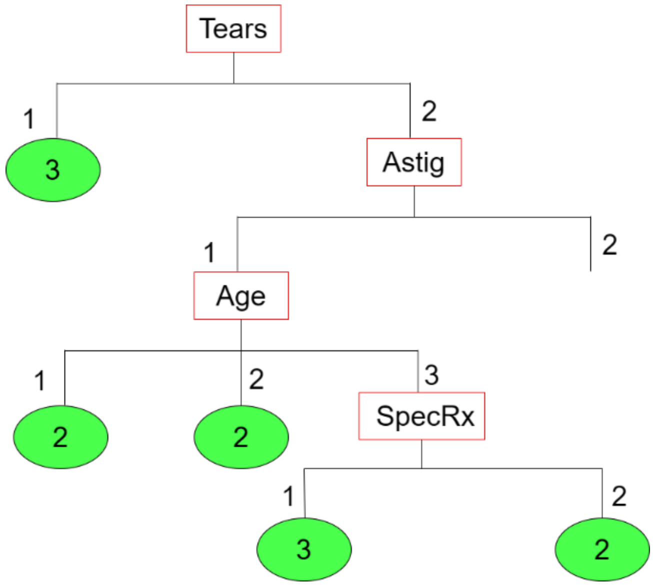
$$\begin{aligned} E_1 &= -(0/3)\log_2(0/3) - (2/3)\log_2(2/3) - (1/3)\log_2(1/3) \\ &= 0 + 0.3900 + 0.5283 \\ &= 0.9183 \end{aligned}$$

For SpecRx = 2,

$$\begin{aligned} E_2 &= -(0/3)\log_2(0/3) - (3/3)\log_2(3/3) - (0/3)\log_2(0/3) \\ &= 0 + 0 + 0 \\ &= 0 \end{aligned}$$

E_{New} (SpecRx) = (3/6) E_1 + (3/6) E_2 = 0.4592 bits

Information Gain, Ig (SpecRx) = E_{Start} - E_{New} (SpecRx) = 0.1908 bits



Iteration 4 (For Branch Astig = 2)

There are 4 instances with classification 1 and 2 instance with classification 3. So, $p_1 = (4/6)$ and $p_2 = (2/6)$.

$$
\begin{aligned}
E_{Start} &= -(4/6)\log_2(4/6) - (2/6)\log_2(2/6) \\
&= 0.3900 + 0.5283 \\
&= 0.9183 \text{ bits}
\end{aligned}
$$

Now, we need to calculate E_{New} for each of the attributes.

Calculating Entropy for different Values of Age

For Age = 1,

$$
\begin{aligned}
E_1 &= -(2/2)\log_2(2/2) - (0/2)\log_2(0/2) - (0/2)\log_2(0/2) \\
&= 0 + 0 + 0 \\
&= 0
\end{aligned}
$$

For Age = 2,

$$
\begin{aligned}
E_2 &= -(1/2)\log_2(1/2) - (0/2)\log_2(0/2) - (1/2)\log_2(1/2) \\
&= 0.5 + 0 + 0.5 \\
&= 1
\end{aligned}
$$

For Age = 3,

$$
\begin{aligned}
E_3 &= -(1/2)\log_2(1/2) - (0/2)\log_2(0/2) - (1/2)\log_2(1/2) \\
&= 0.5 + 0 + 0.5 \\
&= 1
\end{aligned}
$$

E_{New} (Age) = (2/6) E_1 + (2/6) E_2 + (2/6) E_3 = 0.6667 bits

Information Gain, Ig (Age) = E_{Start} - E_{New} (Age) = 0.9183 – 0.6667 = 0.2516 bits

Calculating Entropy for different Values of SpecRx

For SpecRx = 1,

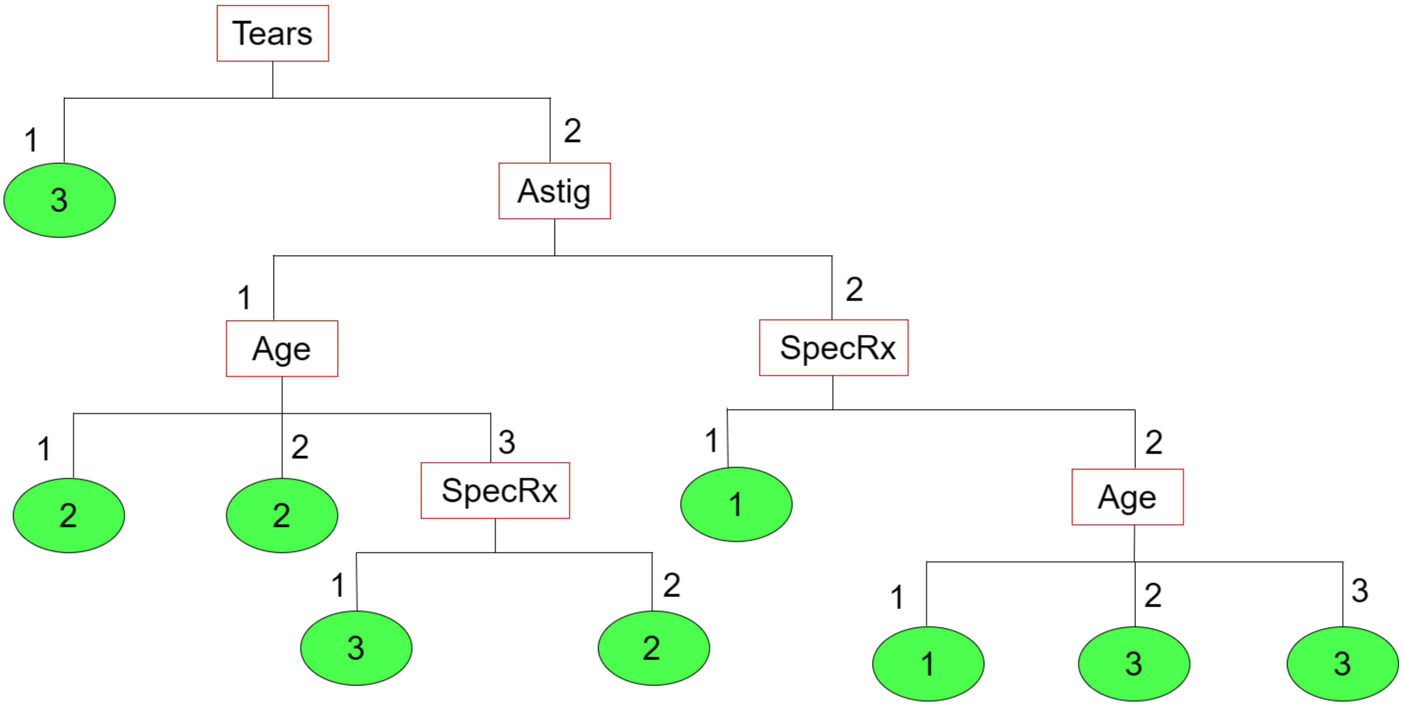
$$
\begin{aligned}
E_1 &= -(3/3)\log_2(3/3) - (0/3)\log_2(0/3) - (0/3)\log_2(0/3) \\
&= 0 + 0 + 0 \\
&= 0
\end{aligned}
$$

For SpecRx = 2,

$$
\begin{aligned}
E_2 &= -(1/3)\log_2(1/3) - (0/3)\log_2(0/3) - (2/3)\log_2(2/3) \\
&= 0.5283 + 0 + 0.3900 \\
&= 0.9183
\end{aligned}
$$

E_{New} (SpecRx) = (3/6) E_1 + (3/6) E_2 = 0.4592 bits

Information Gain, Ig (SpecRx) = E_{Start} - E_{New} (SpecRx) = 0.4591 bits



Tears
1
2
3
Astig
1
2
Age
SpecRx
1
2
3
2
2
SpecRx
1
2
1
Age
1
2
3
2
1
2
3
1
3
3