Expected encoding for information in root of the tree:

$$\sum_{k=1}^{2} P(C_k) * \ (-\log_2 P(C_k) = \frac{1}{2} * \left(-\log_2 \frac{1}{2}\right) + \ \frac{1}{2} * (-\log_2 \frac{1}{2}) = 1$$

Level 1:

A) Use Attribute Color:

Color: red: 2+, 1- | $C_1 : \frac{2}{3}$, $C_2 : \frac{1}{3}$

blue: 1+ | $C_1 : 1$, $C_2 : 0$

green: 0+, 2- | $C_1 : 0$, $C_2 : 1$

Expected encoding for:

$$P(A_C = red) * \sum P(C_k|) * (-\log_2 P(C_k)) = \frac{1}{2}\left[-\frac{2}{3} * \left(-\log_2 \frac{2}{3}\right) + \frac{1}{3}\left(-\log_2 \frac{1}{3}\right)\right]$$

$$= \frac{1}{2}\left[-\frac{2}{3} * (\ 1 - \log_2 3\ ) - \frac{1}{3}(\ 0 - \log_2 3\ )\ \right]$$

$$= \frac{1}{2}\left[\ -\frac{2}{3}(\ 1 - 1.58) - \frac{1}{3}(\ 0 - 1.58)\right]$$

$$= \frac{1}{2} * 0.914$$

$$= 0.457$$

$$P(A_C = blue) * \ \sum P(C_k) * (-\log_2 P(C_k)) = \frac{1}{6} * [1 * (0) + 0] = 0$$

$$P(A_C = green) * \ \sum P(C_k) * (-\log_2 P(C_k)) = \frac{2}{6} * [0 * (0) + \ 1 * (0)] = 0$$

Total expected encoding when attribute color is used:

0.457 + 0 + 0 = 0.457

1 – 0.457 = 0.543

Level 1:

B) Use Attribute Shape:

Shape: square: 2+, 2- | $C_1 : \frac{1}{2}$, $C_2 : \frac{1}{2}$

round: 1+, 1- | $C_1 : \frac{1}{2}$, $C_2 : \frac{1}{2}$

Total expected encoding when shape is used:

$$\sum_{j=1}^{j=2} P(A_i = V_{ij}) * \sum_{k=1}^{k=2} P(C_k)) * (-\log_2 P(C_k))$$

$$= \frac{4}{6} * \left[\frac{2}{4} * \left(-\log_2 \frac{1}{2}\right) + \frac{2}{4}\left(-\log_2 \frac{1}{2}\right)\right] + \frac{2}{6} * \left[\frac{1}{2} * \left(-\log_2 \frac{1}{2}\right) + \frac{1}{2}\left(-\log_2 \frac{1}{2}\right)\right]$$

$$= \frac{2}{3} * \left[\frac{1}{2} * 1 + \frac{1}{2} * 1\right] + \frac{1}{3} * \left[\frac{1}{2} + \frac{1}{2}\right] = 1.0$$

Gain: 1.0 – 1.0 = 0