CS383 Assignment4

Himanshu Gupta

March 2020

1 Question 1

1.1

To calculate the Entropy of the initial target class H(Y), we need total number of the observations (N) = 21, total number of observations of label (n(+)) = 12, and total number of observations of label(n(-)) = 9. Using the formula of Entropy H(Y):

$$H(Y) = -\left(\frac{n(+)}{N} \times \log_2(\frac{n(+)}{N})\right) - \left(\frac{n(-)}{N} \times \log_2(\frac{n(-)}{N}))\right)$$

$$H(Y) = -\left(\frac{12}{21} \times \log_2(\frac{12}{21})\right) - \left(\frac{9}{21} \times \log_2(\frac{9}{21})\right)$$
$$= 0.5239 + 0.4613 = 0.9852$$

1.2

To the calculate the weighted average entropies for x1 and x2, we need total number of the observations (N) = 21, total number of observations with x1 attribute as true (p) = 8, total number of observations with x1 attribute as false (n) = 13, total number of (+) positively labeled observations with x1 attribute as true (p') = 7, total number of (-) negatively labeled observations with x1 attribute as true (n') = 1, total number of (+) positively labeled observations with x1 attribute as false (p") = 5, and total number of (-) positively labeled observations with x1 attribute as false (n") = 8.

Using the formula of Weighted Average Entropy for x1,

$$= -\left((\frac{p}{N}) * (\frac{p'}{p} \times \log_2(\frac{p'}{p}) + \frac{n'}{p} \times \log_2(\frac{n'}{p})\right) - \left((\frac{n}{N}) * (\frac{p''}{n} \times \log_2(\frac{p''}{n}) + \frac{n''}{n} \times \log_2(\frac{n''}{n})\right)$$
$$= -\left((\frac{8}{21}) * (\frac{7}{8} \times \log_2(\frac{7}{8}) + \frac{1}{8} \times \log_2(\frac{1}{8})\right) - \left((\frac{13}{21}) * (\frac{5}{13} \times \log_2(\frac{5}{13}) + \frac{8}{13} \times \log_2(\frac{8}{13})\right)$$
$$= 0.2071 + 0.5951 = 0.8021$$

Similarly in case of x2, we need total number of the observations (N) = 21, total number of observations with x2 attribute as true (p) = 10, total number of observations with x2 attribute as false (n) = 11, total number of (+) positively labeled observations with x2 attribute as true (p') = 7, total number of (-) negatively labeled observations with x2 attribute as true (n') = 3, total number of (+) positively labeled observations with x2 attribute as false (p") = 5, and total number of (-) positively labeled observations with x2 attribute as false (n") = 6.

Using the formula of Weighted Average Entropy for x2,

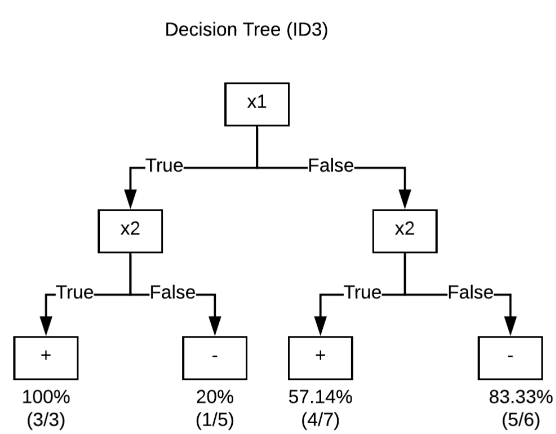
$$= -\left((\frac{p}{N}) * (\frac{p'}{p} \times \log_2(\frac{p'}{p}) + \frac{n'}{p} \times \log_2(\frac{n'}{p}) \right) - \left((\frac{n}{N}) * (\frac{p''}{n} \times \log_2(\frac{p''}{n}) + \frac{n''}{n} \times \log_2(\frac{n''}{n}) \right)$$

$$= -\left((\frac{10}{21}) * (\frac{7}{10} \times \log_2(\frac{7}{10}) + \frac{3}{10} \times \log_2(\frac{3}{10}) \right) - \left((\frac{11}{21}) * (\frac{5}{11} \times \log_2(\frac{5}{11}) + \frac{6}{11} \times \log_2(\frac{6}{11}) \right)$$

$$= 0.4197 + 0.5207 = 0.9403$$

1.3

The Decision Tree is built using the ID3 algorithm which essentially uses change in entropy for every feature to get the best feature out. For x1, change in entropy would be (0.9852 - 0.8021) = 0.1831. For x2, change in entropy would be (0.9852 - 0.9403) =0.0449. The more the change in entropy the better the feature is so x1 is better feature than x2.



2

2.1

$$\text{P(A = Yes)} = \frac{3}{5} = 0.6$$

$$\text{P(A = No)} = \frac{2}{5} = 0.4$$

2.2

$$\text{Data Matrix (D)} = \begin{bmatrix} 216 & 5.68 \\ 69 & 4.78 \\ 302 & 2.31 \\ 60 & 3.16 \\ 393 & 4.2 \end{bmatrix}$$

$$\text{Data Labels (L)} = \begin{bmatrix} Yes \\ Yes \\ No \\ Yes \\ No \end{bmatrix}$$

$$mean(D) = \begin{bmatrix} 208.00 & 4.0260 \end{bmatrix}$$

$$std(D) = \begin{bmatrix} 145.2154 & 1.3256 \end{bmatrix}$$

$$StandardizedDataMatrix = \frac{D - mean(D)}{std(D)} = \frac{\begin{bmatrix} 216 & 5.68 \\ 69 & 4.78 \\ 302 & 2.31 \\ 60 & 3.16 \\ 393 & 4.2 \end{bmatrix} - \begin{bmatrix} 208.00 & 4.0260 \\ 208.00 & 4.0260 \\ 208.00 & 4.0260 \\ 208.00 & 4.0260 \\ 208.00 & 4.0260 \end{bmatrix}}{\begin{bmatrix} 145.2154 & 1.3256 \\ 145.2154 & 1.3256 \\ 145.2154 & 1.3256 \\ 145.2154 & 1.3256 \\ 145.2154 & 1.3256 \end{bmatrix}}$$

$$= \frac{\begin{bmatrix} 8.0000 & 1.6540 \\ -139.0000 & 0.7540 \\ 94.0000 & -1.7160 \\ -148.0000 & -0.8660 \\ 185.0000 & 0.1740 \end{bmatrix}}{\begin{bmatrix} 145.2154 & 1.3256 \\ 145.2154 & 1.3256 \\ 145.2154 & 1.3256 \\ 145.2154 & 1.3256 \\ 145.2154 & 1.3256 \end{bmatrix}} = \begin{bmatrix} 0.0551 & 1.2477 \\ -0.9572 & 0.5688 \\ 0.6473 & -1.2945 \\ -1.0192 & -0.6533 \\ 1.2740 & 0.1313 \end{bmatrix}$$

Now observations with the target attribute value "Yes" is Matrix(Yes)=

$$\begin{bmatrix} 0.0551 & 1.2477 \\ -0.9572 & 0.5688 \\ -1.0192 & -0.6533 \end{bmatrix}$$

Mean and Standard Deviation of this Matrix(Yes) are [-0.6404 , 0.3877] and [0.6031 , 0.9633] respectively.
Similarly observations with the target attribute value "No" is Matrix(No) =

$$\begin{bmatrix} 0.6473 & -1.2945 \\ 1.2740 & 0.1313 \end{bmatrix}$$

Mean and Standard Deviation of this Matrix(No) are [0.9606 , -0.5816] and [0.4431 , 1.0082] respectively.

2.3

Given number of character to be 242 character, and average word length to be 4.56. After standardization new data would be become : [0.234, 0.403]

$$G(x_k, m_i, std_i) = \frac{e^{\frac{-(x_k - m_i)^2}{2*std_i^2}}}{std_i * (2\pi)^{1/2}}$$

$$P(y = "Yes"|f = x) \propto P(Yes) \times G(0.234, -0.6404, 0.6031) \times G(0.403, 0.3877, 0.9633)$$
$$\cong 0.6 * 0.231 * 0.414 \cong 0.0574$$

$$P(y = "No"|f = x) \propto P(No) \times G(0.234, 0.9606, 0.4431) \times G(0.403, -0.5816, 1.0082)$$
$$\cong 0.4 * 0.234 * 0.245 \cong 0.0228$$

As the P(y= "Yes" given f =x) is more than P(y= "No" given f =x), the student would likely to get an A.

3

$$Accuracy : 8.402868e + 01 \cong 84\%$$

$$Precision : 7.461431e + 01$$

$$Recall : 8.926174e + 01$$

$$FMeasure : 8.128342e + 01$$

4

$$Accuracy : 9.322034e + 01 \cong 93\%$$

$$Precision : 9.143836e + 01$$

$$Recall : 9.081633e + 01$$

$$FMeasure : 9.112628e + 01$$