ML-HW3

David A. Hallloran

May 3, 2022

1 Theory

1. Use the following table for subsequent examples:

Y	x_1	x_2	Count
+	T	T	3
+	T	F	4
+	F	T	4
+	F	F	1
-	T	T	0
-	T	F	1
-	F	T	3
-	F	F	5

   (a) Calculate Sample Entropy:
   We have 2 classes, + and -. So we sum the count instances for each class to give us a total of times each class is observations. This gives us 12 for positive, 9 for negative, and 21 for total observations. Calculating probabilities this gives us P(+) = .5714 and P(-) = .4286. Entropy for this example is calculated as:

$$H(P(v_1), P(v_2) = \sigma(-P(v_i)log_2 P(v_i))$$

   So plugging in our probabilities this gives us an entropy of:

$$H(+,-) = -(.5714 * log_2(.5714)) + -(.4286 * log_2(.4286))$$
$$H(+,-) = .1389 + .1578$$
$$H(+,-) = .2967$$

   (b) Weighted Entropy's:
   To weight the entropy's by observation we calculate how many times each observation appears for each class. So since x_1 and x_2 are both binary true/false values, we need the amount of x_1 = T for Y = + and Y = -, x_1 = F for Y = + and Y = -, etc.

For x_1 = T, this happens 8/21 times.
x_1 = F = 13/21
x_2 = T = 10/21
x_2 = F 11/21

In addition to these totals we also need to know how many of these features happen in both the positive and negative class. This gives us:

x_1 = T, +	7/8
x_1 = T, -	1/8
X_1 = F, +	5/13
x_1 = F, -	8/13
x_2 = T, +	7/10
x_2 = T, -	3/10
x_2 = F, +	5/11
x_2 = F, -	6/11

This gives us a weighted entropy for x_1 of:

$$H(x_1) = \frac{8}{21} * (-1 * (\frac{7}{8} * log_2(\frac{7}{8})) + (-1 * (\frac{1}{8} * log_2(\frac{1}{8}))))$$
$$+\frac{13}{21} * (-1 * (\frac{5}{13} * log_2(\frac{5}{13})) + (-1 * (\frac{8}{13} * log_2(\frac{8}{13}))))$$

$$H(x_1) = 0.20707216883794147 + 0.5950512314951137$$

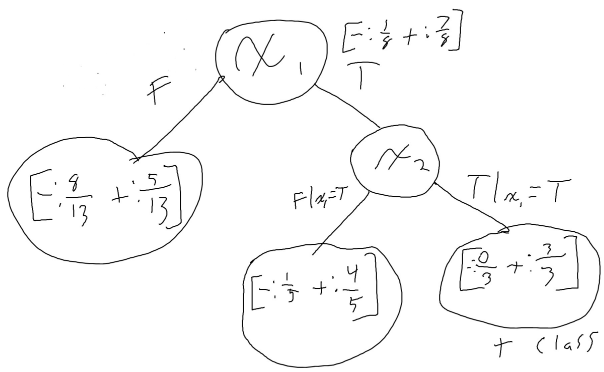
$$H(x_1) = 0.8021234003330552$$

$$H(x_2) = \frac{10}{21} * (-1 * (\frac{7}{10} * log_2(\frac{7}{10})) + (-1 * (\frac{3}{10} * log_2(\frac{3}{10}))))$$
$$+\frac{11}{21} * (-1 * (\frac{5}{11} * log_2(\frac{5}{11})) + (-1 * (\frac{6}{11} * log_2(\frac{6}{11}))))$$

$$H(x_2) = 0.4196623329669965 + 0.5206824917260249$$

$$H(x_2) = 0.9403448246930214$$

(c) Decision Tree:
Since x_1 has the lower entropy we will split on that. The sample ID3 tree is drawn below:



2. .

(a) P(A=Y) = 3/5 P(A=N) = 2/5

(b) In order to z-score, we need the mean and standard deviation of the features based on class. This gives us:

$$Mean_{c0} = 347.5|STD_{c0} = 45.5$$
$$Mean_{c1} = 115|STD_{c1} = 71.5122$$
$$Mean_{w0} = 3.255|STD_{w0} = .945$$
$$Mean_{w1} = 4.54|STD_{w1} = 1.0427$$

The z score formula is $z = (x - mean)/std$. So this gives us:

$$x_{c0} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$x_{c1} = \begin{bmatrix} 1.4123 \\ -0.6432 \\ -0.7691 \end{bmatrix}$$

$$x_{w0} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$x_{w1} = \begin{bmatrix} 1.0933 \\ 0.2302 \\ -1.3235 \end{bmatrix}$$

(c) If we have a validation observation of [242, 4.56], calculate which class this belongs to:

So first let's z-score this observation. This gives us:

$$char0 = (242 - 347.5)/45.5 = -2.3186$$
$$char1 = (242 - 115)/71.5122 = 1.7759$$
$$word0 = (4.56 - 3.255)/.945 = 1.3809$$
$$word1 = (4.56 - 4.54)/1.0427 = 0.0191$$

Now we will use the Gaussian Probability Density function to give us a probability:

$$p(x|y = A) = (3/5) * \frac{1}{(45.5) * \sqrt{2\pi}} * e^{\frac{((-2.3186)-(347.5))}{2*(45.5)^2}} *$$
$$\frac{1}{(.945) * \sqrt{2\pi}} * e^{\frac{((1.3809)-(3.255))}{2*(.945)^2}}$$
$$p(x|y = A) = .6 * 0.00805 * 0.1478$$
$$p(x|y = A) = .000715$$

$$p(x|y =!A) = (2/5) * \frac{1}{(71.5122) * \sqrt{2\pi}} * e^{\frac{((1.7759)-(115))}{2*(71.5122)^2}} *$$
$$\frac{1}{(1.0427) * \sqrt{2\pi}} * e^{\frac{((.0191)-(4.54))}{2*(1.0427^2}}$$
$$p(x|y =!A) = .4 * 0.0055 * 0.0478$$
$$p(x|y =!A) = .000105$$

For binary classification we can take the higher probability, so our model predicts this value would receive an A

2 Naive Bayes

```
Precision: 0.6273637374860956
Recall: 0.9791666666666666
F Measure: 0.7647457627118645
Accuracy: 0.7736464448793215
```

3 Decision Tree

Unfortunately, I could not get this working so there is nothing to put in here

4 Multi-class

```
PS C:\Users\bonem\Desktop\Class\homework-repo\spring_2022\machine_learning\hw3> python p4.py
Accuracy: 0.7994350282485876
```