CS 383 - Machine Learning

Assignment 4 - Classification

1 Theory

1. Consider the following set of training examples for an unknown target function: $(x_1, x_2) \rightarrow y$:

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) What is the sample entropy, $H(Y)$ from this training data (using log base 2) (5pts)?

$$H(Y) = -\sum_{i=1}^{K} P(Y = y_i) \log_2 P(Y = y_i)$$

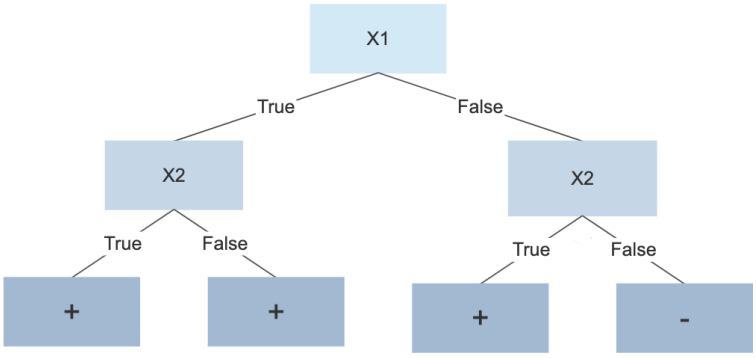
$$= -\frac{4}{7}\log_2\frac{4}{7} - \frac{3}{7}\log_2\frac{3}{7} = 0.9852$$

(b) What are the weighted average entropies of the class labels of the subsets created by variables x_1 and x_2 (5pts)?

$$H(X_1) = [\frac{8}{21} * (-\frac{7}{8}\log_2\frac{7}{8} - -\frac{1}{8}\log_2\frac{1}{8})] + [\frac{9}{21} * (-\frac{5}{13}\log_2\frac{5}{13} - -\frac{8}{13}\log_2\frac{8}{13})] = 0.80176$$

$$H(X_2) = [\frac{10}{21} * (-\frac{7}{10}\log_2\frac{7}{10} - -\frac{3}{10}\log_2\frac{3}{10})] + [\frac{11}{21} * (-\frac{5}{11}\log_2\frac{5}{11} - -\frac{6}{11}\log_2\frac{6}{11})] = 0.94019$$

(c) Draw the decision tree that would be learned by the ID3 algorithm without pruning from this training data. All leaf nodes should have a single class choice at them. If necessary use the mean class or, in the case of a tie, choose one at random.(10pts)?



2. We decided that maybe we can use the number of characters and the average word length an essay to determine if the student should get an A in a class or not. Below are five samples of this data:

# of Chars	Average Word Length	Give an A
216	5.68	Yes
69	4.78	Yes
302	2.31	No
60	3.16	Yes
393	4.2	No

(a) What are the class priors, $P(A = Yes), P(A = No)$? (5pts)

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

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

(b) Find the parameters of the Gaussians necessary to do Gaussian Naive Bayes classification on this decision to give an A or not. Standardize the features first over all the data together so that there is no unfair bias towards the features of different scales (5pts).

Standardized Data for number of chars:

$$Z_{\#} = \frac{X - mean}{\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2}} = \frac{\begin{bmatrix} 216 \\ 69 \\ 302 \\ 60 \\ 393 \end{bmatrix} - 208}{\sqrt{\frac{1}{5-1}\sum_{i=1}^{5}(x_i - 208)^2}} = \begin{bmatrix} 0.055 \\ -0.957 \\ 0.647 \\ -1.019 \\ 1.273 \end{bmatrix}$$

Standardized Data for Average Word Length:

$$Z_{avg} = \frac{X - mean}{\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2}} = \frac{\begin{bmatrix}5.68\\4.78\\2.31\\3.16\\4.2\end{bmatrix} - 4.026}{\sqrt{\frac{1}{5-1}\sum_{i=1}^{5}(x_i - 4.026)^2}} = \begin{bmatrix}1.247\\0.567\\-1.294\\-0.653\\0.131\end{bmatrix}$$

Parameters for Gaussian:

Mean of # of Chars:

When (A = Yes): $\frac{1}{3}sum(\begin{bmatrix}0.055\\-0.957\\-1.019\end{bmatrix}) =$ -0.6404281318437736

When (A = No): $\frac{1}{2}sum(\begin{bmatrix}0.64731446\\1.27396994\end{bmatrix}) = 0.9606421977656603$

Mean of Avg. Word Length:

When (A = Yes): $\frac{1}{3}sum(\begin{bmatrix}1.2477\\0.5687\\-0.6532\end{bmatrix}) = 0.3877418123068708$

When (A = No): $\frac{1}{2}sum(\begin{bmatrix}-1.2944\\0.1312\end{bmatrix}) =$ -0.5816127184603074

Standard Deviation of # of Chars:

When (A = Yes): $\sigma = \sqrt{\frac{1}{3-1}\sum_{i=1}^{3}(x_i - (-0.64042813))^2} = 0.6031334887267573$

When (A = No): $\sigma = \sqrt{\frac{1}{2-1}\sum_{i=1}^{2}(x_i - (0.96064219))^2} = 0.4431123393757109$

Standard Deviation of Avg. Word Length:

When (A = Yes): $\sigma = \sqrt{\frac{1}{3-1}\sum_{i=1}^{3}(x_i - (-0.0.38774181))^2} = 0.9633406106304947$

When (A = No): $\sigma = \sqrt{\frac{1}{2-1}\sum_{i=1}^{2}(x_i - (-0.58161271))^2} = 1.0081527132348793$

(c) Using your response from the prior question, determine if an essay with 242 characters and an average word length of 4.56 should get an A or not. Show the math to support your decision (10pts).

$$\text{P(y=A } \|f = x) \propto P(y = A) * P(f_1 = \#ofchars.\|y = A) * P(f_2 = Avg.Length\|y = A)$$
$$= \frac{3}{5} * 0.231166072 * 0.41407302 = 0.05743178$$

$$\text{P(y=NotA } \|f = x) \propto P(y = NotA) * P(f_1 = \#chars.\|y = NotA) * P(f_2 = Avg.Length\|y = NotA)$$
$$= \frac{2}{5} * 0.23478680 * 0.245657850 = 0.0230708891$$

A>NotA

Therefore, this essay would get an A

2 Naive Bayes Classifier

Classification Stats:
Accuracy:80.4957599478%
Precision:67.6883780332%
Recall: 92.0138888889%
f-Measure: 77.9985283297%

3 Logistic Regression

Accuracy: 90.21526418786692%
Precision: 86.10169491525423%
Recall: 88.19444444444444%
f-Measure: 87.13550600343053%