

Linear Regression

Ahnaf An Nafee

March 2021

1 Theory

1. Target Function:

(a) Sample Entropy:

$$Total = 21$$

$$PositiveCount = 12$$

$$NegativeCount = 9$$

$$H(Y) = -(\frac{12}{21} \cdot \log_2(\frac{12}{21}) + \frac{9}{21} \cdot \log_2(\frac{9}{21})) \quad (1)$$

$$H(Y) = 0.985228$$

(b) Weighed Average Entropy:

$$p_0 = 5 \quad n_0 = 8$$

$$p_1 = 7 \quad n_1 = 1$$

$$E(H(1)) = \frac{5+8}{21} \times (\frac{-5}{13} \cdot \log_2 \frac{5}{13} + \frac{-8}{13} \cdot \log_2 \frac{8}{13}) + \frac{7+1}{21} \times (\frac{-7}{8} \cdot \log_2 \frac{7}{8} + \frac{-1}{8} \cdot \log_2 \frac{1}{8}) \quad (2)$$

$$E(H(1)) = 0.802123 \quad (3)$$

$$p_0 = 5 \quad n_0 = 6$$

$$p_1 = 7 \quad n_1 = 3$$

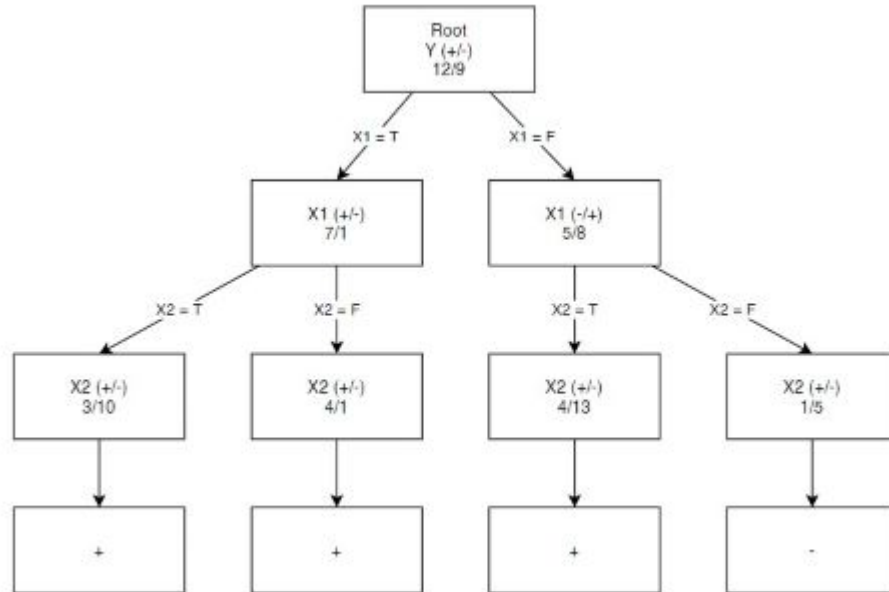
$$E(H(2)) = \frac{5+6}{21} \times (\frac{-5}{11} \cdot \log_2 \frac{5}{11} + \frac{-6}{11} \cdot \log_2 \frac{6}{11}) + \frac{7+3}{21} \times (\frac{-7}{10} \cdot \log_2 \frac{7}{10} + \frac{-3}{10} \cdot \log_2 \frac{3}{10}) \quad (4)$$

$$E(H(2)) = 0.940344 \quad (5)$$

(c) Decision Tree:

$$IG(X_1) = 0.985228 - 0.802123 = 0.183107 \quad (6)$$

$$IG(X_2) = 0.985228 - 0.940344 = 0.044885 \quad (7)$$



2. Essay Data:

(a) Class Priors:

$$P(A = Yes) = 3/5 = 0.6 \quad (8)$$

$$P(A = No) = 2/5 = 0.4 \quad (9)$$

(b) Parameters of the Gaussians:

(c) Determine:

2 Naive Bayes Classifier

1. Classification Statistics:

Precision: 94.54545454545455%
Recall: 69.84126984126983%
F-measure: 80.33707865168539%
Accuracy: 81.7351598173516%

3 Logistic Regression

1. Classification Statistics:

Precision: 37.37864077669903%
Recall: 77.77777777777779%
F-measure: 50.49180327868853%
Accuracy: 70.45009784735812%