# CS 613 - Assignment 3

## Name - Shaoshu Xu

**PART-1 Theory-1**

(a). Sample Entropy

$$H(Y) = -\frac{12}{21} * \log_2 \frac{12}{21} - \frac{9}{21} * \log_2 \frac{9}{21} = 0.9852$$

(b). Information Gains

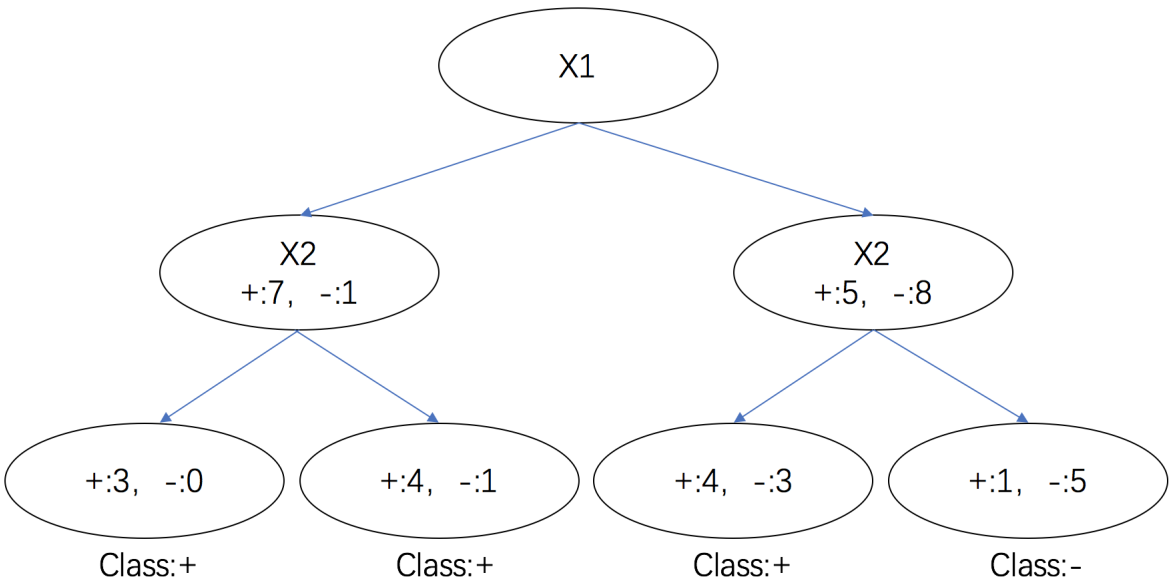
$$H(x_1) = \frac{8}{21}(-\frac{7}{8} * \log_2 \frac{7}{8} - \frac{1}{8} * \log_2 \frac{1}{8}) + \frac{13}{21}(-\frac{5}{13} * \log_2 \frac{5}{13} - \frac{8}{13} * \log_2 \frac{8}{13}) = 0.802$$

$$Information - Gain(x_1) = H(Y) - H(x_1) = 0.1832$$

$$H(x_2) = \frac{10}{21}(-\frac{7}{10} * \log_2 \frac{7}{10} - \frac{3}{7} * \log_2 \frac{3}{7}) + \frac{11}{21}(-\frac{5}{11} * \log_2 \frac{5}{11} - \frac{6}{11} * \log_2 \frac{6}{11}) = 0.9417$$

$$Information - Gain(x_2) = H(Y) - H(x_2) = 0.0448$$

(c). Decision Tree



**PART-1 Theory-2**

(a). Class Priors

$$P(A = Yes) = 3/5 = 0.6, P(A = No) = 2/5 = 0.4$$

(b). Gaussian Parameters:
X =

$$\begin{bmatrix} 216 & 5.68 & yes \\ 69 & 4.78 & yes \\ 302 & 2.31 & no \\ 60 & 3.16 & yes \\ 393 & 4.2 & no \end{bmatrix}$$

X-standardized = (X - mean(X)) ./ std(X) =

$$\begin{bmatrix} 0.0551 & 1.2477 & yes \\ -0.9572 & 0.5688 & yes \\ 0.6473 & -1.2945 & no \\ -1.0192 & -0.6533 & yes \\ 1.2740 & 0.1313 & no \end{bmatrix}$$

$$\mu_{yes1} = -0.6404, \sigma_{yes1} = 0.6031 ===>>>$$

**Model for Yes: Number of Chars - Gaussian N(-0.6404, 0.6031)**

$$\mu_{yes2} = 0.3877, \sigma_{yes2} = 0.9633 ===>>>$$

**Model for Yes: Average World Length - Gaussian N(0.3877, 0.9633)**

$$\mu_{no1} = 0.9606, \sigma_{no1} = 0.4431 ===>>>$$

**Model for No: Number of Chars - Gaussian N(0.9606, 0.4431)**

$$\mu_{no2} = -0.5816, \sigma_{no2} = 1.0082 ===>>>$$

**Model for No: Average World Length - Gaussian N(-0.5816, 1.0082)**

(c). Predict Classification

**Standardized the testing data [242, 4.56]: [0.2341, 0.4028]**

$$P(A = yes|Chars = 242, AWL = 4.56) = P(A = yes) * P(Chars = 242|N(\mu_{yes1}, \sigma_{yes1}) * P(Chars = 4.56|N(\mu_{yes2}, \sigma_{yes2})$$

$$P(A = yes|Chars = 242, AWL = 4.56) = 0.6 * 0.2312 * 0.4141 = 0.0574$$

$$P(A = no|Chars = 242, AWL = 4.56) = P(A = no) * P(Chars = 242|N(\mu_{no1}, \sigma_{no1}) * P(Chars = 4.56|N(\mu_{no2}, \sigma_{no2})$$

$$P(A = no|Chars = 242, AWL = 4.56) = 0.4 * 0.2348 * 0.2457 = 0.0231$$

**Normalization:**

$$P(A = no|Chars = 242, AWL = 4.56)_{normalized} = 1/(0.0574/0.0231 + 1) = 0.2870$$

$$P(A = yes|Chars = 242, AWL = 4.56)_{normalized} = 1 - 0.2870 = 0.7130$$

**So, the class is "Yes", and the essay should get an A.**

**PART-1 Theory-3**

(a).
We could set k = [1, 2, 3, ..., 20, ...], and iteratively using a particular value of k to evaluate the performance on validation data set. Finding the hyperparameter - k that work best on the validation set.

(b).
If we use the training data set to determine the parameter k, it will fit the training data well and give us a good accuracy. However, it can only be said that this k value has good performance on this particular training set, and it is not the true performance of this model. This k value could over-fitting our training data, but perform bad on the testing data.

(c).
If we use the testing data set to determine the parameter k, it will fit the testing data well and give us a good accuracy. However, it can only be said that this k value has good performance on this particular testing set, and it is not the true performance of this model. When we apply this model on other data set, it could perform very bad.

**PART-2 Naive Bayes Classier**

Precision = 0.682039

Recall = 0.955782

F-measure = 0.796034

Accuracy = 0.812133

**PART-3 Decision Trees**

Precision = 0.878788

Recall = 0.887755

F-measure = 0.883249

Accuracy = 0.909980

**PART-4 Additional Evaluation**

**The accuracy of Naive Bayes (Multi-class) classier:**

$$Accuracy = 0.412429$$

**The accuracy of ID3 Decision Tree (Multi-class) classier:**

$$Accuracy = 0.853107$$