CS 383 - Machine Learning

Assignment 4 - Naive Bayes, Decision Trees and Nearest Neighbors
Summer 2017
Amir Omidi

1 Theory

1.1 Question 1

1.1.1 Part A

As we recall from homework 1, entropy of a two feature data set is calculated with the following formula:

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

So for our data it would be:

$$\begin{aligned}
H &= -\frac{12}{21} \times \log_2 \frac{12}{21} - \frac{9}{21} \times \log_2 \frac{9}{21} \\
H &= 0.98522
\end{aligned}$$

1.1.2 Part B

To calculate the information gain, we first specify the true and false count of each feature:

Feature 1

$$\begin{aligned}
P_T &= 7 \qquad N_T = 1 \\
P_F &= 5 \qquad N_F = 8
\end{aligned}$$

Feature 2

$$\begin{aligned}
P_T &= 7 \qquad N_T = 3 \\
P_F &= 5 \qquad N_F = 6
\end{aligned}$$

Now we can calculate the remainder of each feature:

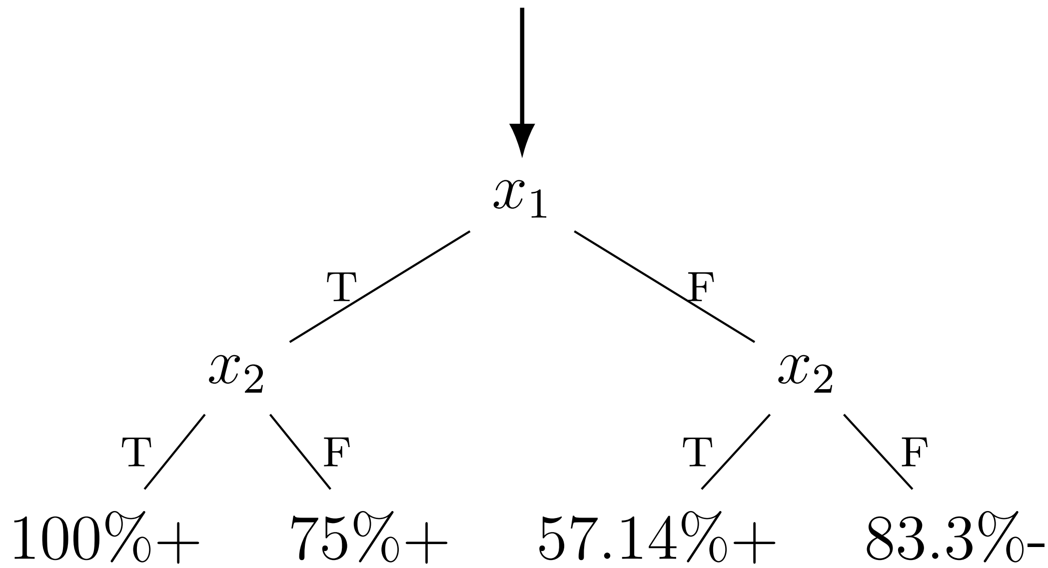
$$\begin{aligned}
Remainder(1) &= \left(\frac{8}{21} \times \left(\frac{-7}{8}\log_2\frac{7}{8} + \frac{-1}{8}\log_2\frac{1}{8}\right)\right) + \left(\frac{13}{21} \times \left(\frac{-5}{13}\log_2\frac{5}{13} + \frac{-8}{13}\log_2\frac{8}{13}\right)\right) \\
Remainder(1) &= 0.8021 \\
Remainder(2) &= \left(\frac{10}{21} \times \left(\frac{-7}{10}\log_2\frac{7}{10} + \frac{-3}{10}\log_2\frac{3}{10}\right)\right) + \left(\frac{11}{21} \times \left(\frac{-5}{11}\log_2\frac{5}{11} + \frac{-6}{11}\log_2\frac{6}{11}\right)\right) \\
Remainder(2) &= 0.9403
\end{aligned}$$

And finally, the information gain from each feature is:

$$\begin{aligned}
IG(1) &= 0.98522 - 0.8021 = 0.18312 \\
IG(2) &= 0.98522 - 0.9403 = 0.04492
\end{aligned}$$

1.1.3 Part C

The informationg gain from the first feature is larger than the second feature, therefor we will split on the first feature:



1.2 Question 2

1.2.1 Part A

The priors are:

$$
\begin{aligned}
P(A) &= \frac{3}{5} \\
P(\neg A) &= \frac{2}{5}
\end{aligned}
$$

1.2.2 Part B

The first step in this question is to standardize our data. Since this standardization step has been done before (Homework 1). The steps taken to calculate the standard deviation and means won't be shown.

$$
\begin{aligned}
\mu_{f_1} &= 208 & \mu_{f_2} &= 4.03 \\
\sigma_{f_1} &= 145.22 & \sigma_{f_2} &= 1.33
\end{aligned}
$$

$$
\begin{bmatrix}
0.0551 & 1.2477 & Yes \\
-0.9572 & 0.5688 & Yes \\
0.6472 & -1.2945 & No \\
-1.0192 & -0.6533 & Yes \\
1.2740 & 0.1313 & No
\end{bmatrix}
$$

By standardizing the data, each feature has a mean value of zero and a standard deviation of 1. We can now calculate the parameters of the Gaussians.

Model:

$$
\begin{aligned}
P(x|\mu, \sigma) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\
P(chars|0, 1) &= \frac{1}{2.51} e^{-\frac{(x)^2}{2}} \\
P(AWL|0, 1) &= \frac{1}{2.51} e^{-\frac{(x)^2}{2}}
\end{aligned}
$$

1.2.3 Part C

The number of characters is 242 with an average word length of 4.56. Let us standardize this data first:

$$
\begin{aligned}
Standardize(x, \mu, \sigma) &= Value \\
Standardize(242, 208, 145.22) &= 0.2341 \\
Standardize(4.56, 4.03, 1.33) &= 0.3985
\end{aligned}
$$

We can now calculate the probability of getting an A:

$$P(A|x_1 = 0.2341, x_2 = 0.3985) = \frac{P(0.2341|A) \times P(0.3985|A) \times P(A)}{P(0.2341|A) \times P(0.3985|A) + P(0.2341|\neg A) \times P(0.3985|\neg A) \times P(\neg A)}$$
$$P(A|x_1 = 0.2341, x_2 = 0.3985) = 0.75$$

We see the the probability is 75%. Therefor, the expected score is A.

2 Programming

2.1 k-Nearest Neighbours

Statistical analysis of KNN with $k = 5$.

$$\begin{aligned} Precision: &\quad 92.52\% \\ Recall: &\quad 84.18\% \\ F - Measure: &\quad 88.16\% \\ Accuracy: &\quad 91.32\% \end{aligned}$$