Yiyun Zhang(Benny)
yz523
CS383-HW4

Theory

1.

(a) Total count = 3+4+4+1+0+1+3+5=21

Count for Y_+ = 3+4+4+1 = 12

Count for Y_- = 0+1+3+5 = 9

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

(b) P_{1T} = 3+4 = 7 N_{1T} = 0+1 = 1

P_{1F} = 4+1 = 5 N_{1F} = 3+5 = 8

P_{2T} = 3+4 = 7 N_{2T} = 0+3 = 3

P_{2F} = 4+1 = 5 N_{2F} = 1+5 = 6

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

= 0.8021

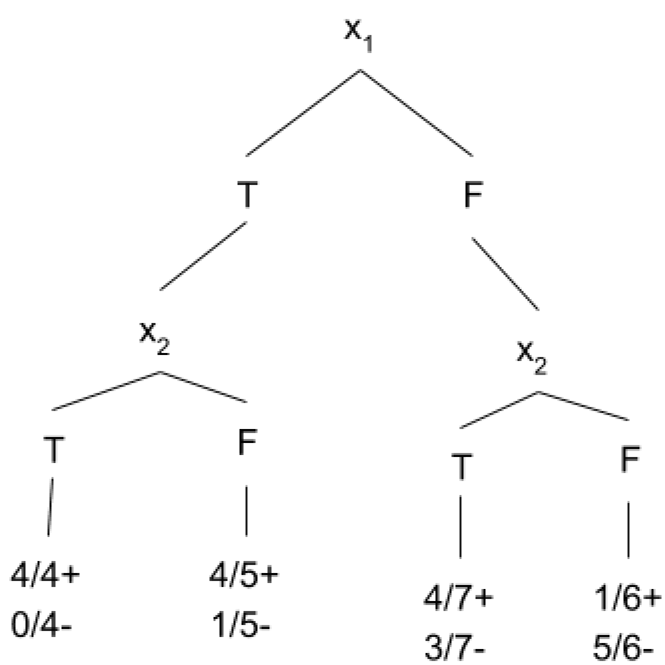
$Remainder(x_2) = \frac{7+3}{21}(-\frac{7}{7+3}log_2\frac{7}{7+3} - \frac{3}{7+3}log_2\frac{3}{7+3}) + \frac{5+6}{21}(-\frac{5}{5+6}log_2\frac{5}{5+6} - \frac{6}{5+6}log_2\frac{6}{5+6})$

= 0.9403

$IG(x_1) = H(Y) - Remainder(x_1)$ = 0.9852 - 0.8021 = 0.1831

$IG(x_2) = H(Y) - Remainder(x_2)$ = 0.9852 - 0.9403 = 0.0449

(c) $IG(x_1) > IG(x_2)$, so split x_1.



2.

(a) P(A = Yes) = $\frac{3}{5}$ = 60%   P(A = No) = $\frac{2}{5}$ = 40%

(b) $\mu_1 = \frac{(216+69+302+60+393)}{5}$ = 208 σ_1 = 145.2154

$\mu_2 = \frac{(5.68+4.78+2.31+3.16+4.2)}{5}$ = 4.0260 σ_2 = 1.3256

$$X = \begin{bmatrix} 0.0551 & 1.2477 \\ -0.9572 & 0.5688 \\ 0.6473 & -1.2945 \\ -1.0192 & -0.6533 \\ 1.2740 & 0.1313 \end{bmatrix}$$

Given an A = $\begin{bmatrix} 0.0551 & 1.2477 \\ -0.9572 & 0.5688 \\ -1.0192 & -0.6533 \end{bmatrix}$

$\mu_1 = -0.6404$ $\sigma_1 = 0.6031$

$\mu_2 = 0.3877$ $\sigma_2 = 0.9633$

Gaussian Distribution for Given an A:

$$p(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$p(\text{\# of Chars}|-0.6404,0.6031) = \frac{1}{0.6031\sqrt{2\pi}} e^{-\frac{(x+0.6404)^2}{2*0.6031^2}}$$

$$p(\text{Average Word Length}|0.3877,0.9633) = \frac{1}{0.9633\sqrt{2\pi}} e^{-\frac{(x-0.3877)^2}{2*0.9633^2}}$$

Not Given an A = $\begin{bmatrix} 0.6473 & -1.2945 \\ 1.2740 & 0.1313 \end{bmatrix}$

$\mu_1 = 0.9607$ $\sigma_1 = 0.4413$

$\mu_2 = -0.5816$ $\sigma_2 = 1.0082$

Gaussian Distribution for not Given an A:

$$p(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$p(\text{\# of Chars}|0.9607,0.4413) = \frac{1}{0.4413\sqrt{2\pi}} e^{-\frac{(x-0.9607)^2}{2*0.4413^2}}$$

$$p(\text{Average Word Length}|-0.5816,1.0082) = \frac{1}{1.0082\sqrt{2\pi}} e^{-\frac{(x+0.5816)^2}{2*1.0082^2}}$$

(c) $\mu = \frac{242-208}{145.2154} = 0.2341$ $\sigma = \frac{4.56-4.0260}{1.3256} = 0.4028$

Given A

= p(A=yes) * p(0.2341|-0.6404,0.6031) * p(0.4028|0.3877,0.9633)

$$= \frac{3}{5} * \frac{1}{0.6031\sqrt{2\pi}} e^{-\frac{(0.2341+0.6404)^2}{2*0.6031^2}} * \frac{1}{0.9633\sqrt{2\pi}} e^{-\frac{(0.4028-0.3877)^2}{2*0.9633^2}}$$

= 0.6 * 0.2312 * 0.4141

= 0.0574

Not Given A

= p(A=no) * p(0.2341|0.9607,0.4413) * p(0.4028|-0.5816,1.0082)

$$= \frac{2}{5} * \frac{1}{0.4413\sqrt{2\pi}} e^{-\frac{(0.2341-0.9607)^2}{2*0.4413^2}} * \frac{1}{1.0082\sqrt{2\pi}} e^{-\frac{(0.4028+0.5816)^2}{2*1.0082^2}}$$

= 0.4 * 0.2331 * 0.2457

= 0.2291 < 0.0574, So should get an A.

Programming

1.

```
Precision 0.92523
Recall 0.84184
F-measure 0.88157
Accuracy 0.91324
```