THE HONG KONG POLYTECHNIC UNIVERSITY

HONG KONG COMMUNITY COLLEGE

Subject Title : Introduction to Data Mining **Subject Code** : CCN3164

Session : Semester Two, 2017/18

Numerical Answers

Question A1

(c)(i) 30

(c)(ii) $\sqrt{350}$

(c)(iii) 15

Question A2

(c) $\text{RMF}(R_1, R_2) = \frac{2}{3}$; $\text{RMF}(R_1, R_3) = \frac{1}{3}$; $\text{RMF}(R_1, R_4) = \frac{2}{3}$;
$\text{RMF}(R_2, R_3) = \frac{1}{3}$; $\text{RMF}(R_2, R_4) = \frac{2}{3}$; $\text{RMF}(R_3, R_4) = 1$;

(d)

RMF	R1	R2	R3	R4
R1	0			
R2	2/3	0		
R3	1/3	1/3	0	
R4	2/3	2/3	1	0

Question A3

(b) $\text{JC}(W_1, W_2) = \frac{3}{7}$; $\text{JC}(W_1, W_3) = \frac{3}{5}$; $\text{JC}(W_2, W_3) = \frac{4}{6} = \frac{2}{3}$

Question A4

(b)

Distance	R1	R2/R3
R2/R3	48.37	
R4	26.48	14.14

Question A5

(b) $wc(C_1) = \frac{16.2^2 + 18.6^2}{2} = 304.2$ $wc(C_2) = \frac{15.8^2 + 14.1^2}{2} = 224.225$

(d) $WC = wc(C_1) + wc(C_2) = 528.425$

Question A6

(a) $\text{Error Rate} = \frac{2+2}{9+2+2+5} = \frac{4}{18} = \frac{2}{9}$

(b) $\text{Accuracy} = 1 - \frac{2}{9} = \frac{7}{9}$

Question B2

(h) 1.33

Question B3

(a) $P(\text{Class} = \text{Y}) = \frac{2}{5}$; $P(\text{Class} = \text{N}) = \frac{3}{5}$

(b) $I(\text{Class} = \text{Y}) = -\log_2 \frac{2}{5}$; $I(\text{Class} = \text{N}) = -\log_2 \frac{3}{5}$

(c) 0.970951

(d) 0.4; 0; 0.2; 0.4

(e) $I(\text{Class=Y} \mid \text{Body Weight=Heavy}) = -\log_2 \frac{2}{3}$; $I(\text{Class=Y} \mid \text{Body Weight=Low}) = 0$

$I(\text{Class=N} \mid \text{Body Weight=Heavy}) = -\log_2 \frac{1}{3}$; $I(\text{Class=N} \mid \text{Body Weight=Low}) = 0$

(f) 0.550978

(g) 0.42

(h) 0.57