

2022/3/14 12:28

COMP192 Answer Sheet

COMP192 Exploring and Visualizing Data (Spring Semester 2011)
Midterm Examination (Answer Sheet)
Date: 25 March, 2011 (Fri)
Time: 13:35-14:35
Duration: 1 hour

Student ID: _____ Student Name: _____

Seat No. : _____

- Instructions:
- (1) Please answer **all** questions in this paper.
 - (2) The total marks you can obtain in this exam are 100 only.
 - (3) You can use a calculator.

Answer Sheet

Question	Full Mark	Mark
Q1	20	
Q2	20	
Q3	20	
Q4	20	
Q5	20	
Total	100	

COMP192 Answer Sheet

Q1 (20 Marks)

(a) (i) support = 2

(ii) confidence = $2/3 = 66.7\%$

(iii) expected confidence of the consequent of the rule = $3/5$
lift ratio of the rule = $66.7/60 = 1.11$

(iv) freq. itemsets
= { {A}, {C}, {D}, {E},
{A, C}, {A, D}, {A, E}, {C, E}, {D, E},
{A, C, E}, {A, D, E} }

(b)

A	B	C	D	E
1	1	1	0	0
1	1	1	1	0
1	1	0	1	0

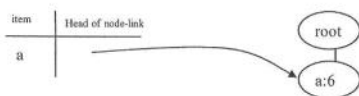
COMP192 Answer Sheet

Q2 (20 Marks)

(a) (i)



(ii)



3/8

COMP192 Answer Sheet

Q2 (Continued)

(a) (iii)



(iv) freq. itemsets
= { {a}, {b}, {c}, {a, c}, {a, b} }

(b)

The corresponding transactions are:

TID	Items
1	a, b, c
2	a, b
3	a, c
4	a, c
5	a
6	a

4/8

COMP192 Answer Sheet

Q3 (20 Marks)

(a)

Cluster 1: a, b
Cluster 2: c, d, e

(b)(i)

$$\begin{matrix} x_1 & x_2 & x_3 & x_4 \\ x_1 & \begin{pmatrix} 0 & 1 & 4 & 5.1 \end{pmatrix} \\ x_2 & \begin{pmatrix} 1 & 0 & 3 & 4.12 \end{pmatrix} \\ x_3 & \begin{pmatrix} 4 & 3 & 0 & 1.41 \end{pmatrix} \\ x_4 & \begin{pmatrix} 5.1 & 4.12 & 1.41 & 0 \end{pmatrix} \end{matrix}$$

(ii)

$$\begin{matrix} x_1 & x_2 & x_3 & x_4 \\ x_1 & \begin{pmatrix} 0 & 1 & 4 & 5.1 \end{pmatrix} \\ x_2 & \begin{pmatrix} 1 & 0 & 3 & 4.12 \end{pmatrix} \\ x_3 & \begin{pmatrix} 4 & 3 & 0 & 1.41 \end{pmatrix} \\ x_4 & \begin{pmatrix} 5.1 & 4.12 & 1.41 & 0 \end{pmatrix} \end{matrix}$$

$$\begin{matrix} x_1: (1, 2) \\ x_2: (2, 2) \\ x_3: (5, 2) \\ x_4: (6, 1) \end{matrix}$$

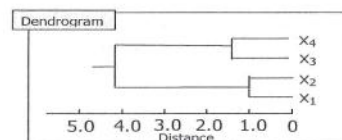
$$\begin{matrix} (x_1, x_2) & \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ 0 & 3.5 & 0 & 4.61 \end{pmatrix} \\ x_3 & \\ x_4 & \end{matrix}$$

$$\begin{matrix} (x_1, x_2): (1.5, 2) \\ x_3: (5, 2) \\ x_4: (6, 1) \end{matrix}$$

5/8

COMP192 Answer Sheet

Q3 (Continued)



6/8

COMP192 Answer Sheet

Q4 (20 Marks)

(a)

Make initial guesses of the means m_1, m_2, \dots, m_k Set the counts n_1, n_2, \dots, n_k to zero

Until interrupted

Acquire the next example x If m_i is closest to x Increment n_i Replace m_i by $m_i + 1/n_i (x - m_i)$

(b)

 x_j : the j -th example in cluster i $m_i(t)$: the mean vector of cluster i containing t examplesConsider that x is the t -th example in cluster i

$$m_i(t-1) = \frac{x_1 + x_2 + \dots + x_{t-1}}{t-1}$$

$$m_i(t) = \frac{x_1 + x_2 + \dots + x_{t-1} + x_t}{t}$$

$$= \frac{m_i(t-1) \times (t-1) + x_t}{t}$$

$$= \frac{t \times m_i(t-1) + x_t - m_i(t-1)}{t}$$

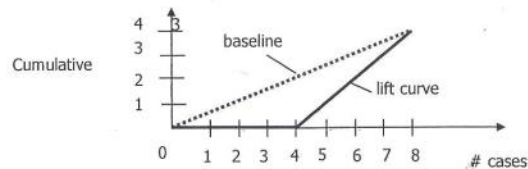
$$= m_i(t-1) + \frac{1}{t}(x_t - m_i(t-1))$$

COMP192 Answer Sheet

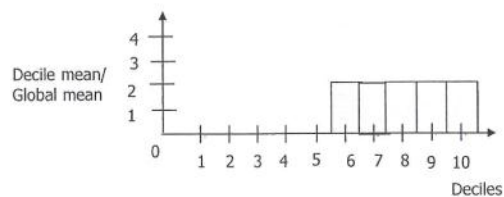
Q5 (20 Marks)

(a) We need to transform the first three categorical attributes to the other corresponding three numeric attributes.

(b)(i)



(ii)



(iii) No. This is because the non-zero bars usually appear at large values along the x-axis instead of small values.

End of Answer Sheet