Data Mining & Web Algorithm.

Date / /
Page No.

Tut - 2.

Patil Amit
19104004
B11

Q.1. a) Binary, Qualitative, Nominal.
b) Continuous, Qualitative, Ratio.
c) Discrete, Quantitative, Ordinal.
d) Continuous, Qualitative, ratio.
e) Discrete, Qualitative, Ordinal.
f) Continuous, Quantitative, Ratio.
g) Discrete, Quantitative, Ratio.
h) Discrete, Qualitative, Nominal.
i) Discrete, Qualitative, Ordinal.

Q.2. Manhattan (L_1)

L_1	A	B	C	D	E	F
A	•					
B	20	0				
C	63	53	0			
D	77	57	56	0		
E	98	88	35	31	0	
F	94	84	9	27	9	0

$d(B,A) = |54-49| + |24-9| = 20$
$d(C,A) = |69-23| + |51-9| = 63$
$d(C,B) = |28-54| + |51-24| = 53$
$d(D,A) = |54-49| + |81-9| = 77$

$d(D,B) = |54-54| + |81-24| = 57$

$d(E,A) = |28-49| + |81-9| = 98$

$d(E,B) = |28-54| + |81-24| = 88$

$d(E,C) = |28-28| + |81-51| = 35$

$d(E,D) = |28-54| + |81-81| = 31$

$d(F,A) = |32-49| + |86-91| = 94$

$d(F,B) = |32-54| + |86-24| = 84$

$d(F,C) = |32-28| + |36-81| = 9$

$d(F,D) = |32-54| + |56-21| = 27$

$d(F,F) = |32-23| + |56-86| = 9$

Euclidian (L_2):

L_2	A	B	C	D	E	F
A	0					
B		0				
C			0			
D				0		
E					0	
F						0

$d(B,A) = (5^2 + 15^2)^{1/2}$

$d(C,A) = (21^2 + 42^2)^{1/2}$

$d(C,B) = (26^2 + 27^2)^{1/2}$

$d(D,A) = (5^2 + 72^2)^{1/2}$

$d(D,B) = (57^2)^{1/2}$

$d(D,C) = (26^2 + 30^2)^{1/2}$

$d(D,A) = (26^2 + 72^2)^{1/2}$

$d(D,A) = (26^2 + 72^2)^{1/2}$

$d(E,B) = (31^2 + 57^2)^{1/2}$

$d(E,C) = (5^2 + 30^2)^{1/2}$

$d(E,D) = (31^2 + 0)^{1/2}$

$d(F,A) = (17^2 + 77^2)^{1/2}$

$d(F,B) = (22^2 + 62^2)^{1/2}$

$d(F,C) = (22^2 + 5^2)^{1/2}$

$d(F,D) = (22^2 + 5^2)^{1/2}$

$d(F,E) = (9^2 + 0)^{1/2}$.

b)

Q.7. a)

b)

Q. 3.

	DOC 1	DOC 2	DOC 3	DOC 4
clothes	0	0	0	1
breasts	1	0	0	0
home	1	1	1	1
in	0	1	2	1
increase	0	0	1	0
July	0	1	1	2
New	1	0	0	2
sales	0	1	0	1
TOP	1	1	1	1
	1	0	0	0

Q.6. a) Sorted array.

1, 5, 12, 16, 18, 20, 23, 25, 28, 33, 35, 40

median $= \frac{20+23}{2} = 21.5$.

$Q_1 = \frac{12+16}{2} = 14$ $\quad Q_3 = \frac{28+33}{2} = 30.5$.

b) Sorted Array.

1, 5, 12, 16, 18, 20, 22, 13, 25, 28, 33, 35
40

median = 22

$Q_1 = 14$ $\quad Q_3 = 30.5$

Q.7. a) $n = 50$ $\quad \Sigma x = 60 + \dots + 78 = 3746$

$$\text{mean} = \frac{3746}{50}$$

b) Modes = { 70, 72, 71 }

This get in trimodal.

DOC 4

1
0
1
1
0
2
2
1
1
0

40