DISCLAIMER

I have only answered 7 out of 9 questions because Lalitha ma'am didn't teach the concept of that sum. The answers provided may be incorrect. All users are advised to thoroughly review the solutions before using them. I will not be held responsible for any errors.

If you have any objections to the structure of the content, please feel free to share your answers with this group.

THANK YOU

Unit-3 & 4 sums

1. Consider that the minimum and maximum values for the attribute "salary" are 12000 and 98000 respectively and the mapping range of salary is [0.0, 1.0]. Find the transformation for the salary 73600 using min-max normalization.

minimum = 12000

maximum = 98000

range = [0.0, 1.0]

V = 73600

$$v' = \frac{V - min(A)}{max - min} \times (new\ max(A) - new\ min(A)) + new\ min$$

$$v' = \frac{73600 - 12000}{98000 - 12000} \times (1.0 - 0.0) + 0.0$$

$$v' = \frac{61600}{86000}$$

$$v' \approx 0.7163$$

Thus, the normalized salary value is approximately 0.7163

2. consider the following set of data X = {15, 27, 62, 35, 39, 50, 44, 44, 22, 98} Do preprocessing using smoothing by bin means and bin boundary to smooth the data, using a bin of depth 3 Evaluate it.

Data = {15, 27, 62, 35, 39, 50, 44, 44, 22, 98

depth = 3

sorted data = {15, 22, 27, 35, 39, 42, 44, 50, 62, 98}

equi width

bin1: 15, 22, 27

bin2: 35, 39, 44

bin3: 46, 50, 62

bin4: 98

Smoothing by bin means:

bin1: 21 21 21

bin2: 39 39 39

bin3: 52 52 52

bin4: 98

Smoothing by boundaries

bin1: 15 15 27

bin2: 35 35 44

bin3: 44 44 62

bin4: 98

Smoothing by bin Means:

{21, 21, 21, 39, 39, 39, 52, 52, 52, 98}

Smoothing by boundaries

{15, 15, 27, 35, 35, 44, 46, 44, 62, 98}

3. Suppose that the data for analysis include the attribute age. The age values for the data tuples are 13, 15, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.

(i) use smoothing by bin means to smooth the above data using a bin depth of 3. Illustrate your steps.

data: 13, 15, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.

bin depth: 3

equi width:

bin 1: 13, 15, 16

bin 2: 19, 20, 20

bin 3: 21, 22, 22

bin 4: 25, 25, 25

bin 5: 25, 30, 33

bin 6: 33, 35, 35

bin 7: 35, 35, 36

bin 8: 40, 45, 46

bin 9: 52, 70

Smoothing by bin means:

bin 1: 15 15 15

bin 2: 20 20 20

bin 3: 22 22 22

bin 4: 25 25 25

bin 5: 29 29 29

bin 6: 34 34 34

bin 7: 35 35 35

bin 8 44 44 44

bin 9: 61 61

Smoothing by median:

bin 1: 15 15 15

bin 2: 20 20 20

bin 3: 22 22 22

bin 4: 25 25 25

bin 5: 30 30 30

bin 6: 35 35 35

bin 7: 35 35 35

bin 8: 45 45 45

bin 9: 61 61

Smoothing by boundary:

bin 1: 13 16 16

bin 2: 19 20 20

bin 3: 21 22 22

bin 4: 25 25 25

bin 5: 25 33 33

bin 6: 33 35 35

bin 7: 35 35 36

bin 8: 40 46 46

bin 9: 52 70

Smoothing by bin means:

{15, 15, 15, 20, 20, 20, 22, 22, 22, 25, 25, 25, 29, 29, 29, 34, 34, 34, 35, 35, 35, 44, 44, 44, 61, 61}

Smoothing by median:

{15, 15, 15, 20, 20, 20, 22, 22, 22, 25, 25, 25, 30, 30, 30, 35, 35, 35, 35, 35, 35, 45, 45, 45, 61, 61}

-Smoothing by boundary

{13, 8, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 33, 33, 33, 35, 35, 35, 35, 36, 40, 46, 46, 52, 70}.

4. Consider the following data for the attribute AGE: 4, 8, 21, 5, 21, 24, 34, 28, 25, perform smoothing by bin means and bin boundary using a bin depth of 3.

given data to be sorted in ascending order

{4, 5, 8, 21, 21, 24, 25, 28, 34}

depth=3

equi width

bin1: 4 5 8

bin2: 21 21 24

bin3: 25 28 34

Smoothing by Bin Boundaries

bin1: 4 4 8

bin2: 21 21 24

bin3: 25 25 34

Smoothing by Bin means:

bin1: 6 6 6

bin2: 22 22 22

bin3: 29 29 29

Smoothing by Bin means

{6, 6, 6, 22, 22, 22, 29, 29, 29}

Smoothing by Bin Boundary

{4, 4, 8, 21, 21, 24, 25, 25, 34}

5. Analyze using equi-depth binning method, partition the data given below into 4 bins and perform smoothing according to the following methods.

(i) smoothing by bin means

(ii) smoothing by bin median

(iii) smoothing by bin boundaries

24, 25, 26, 27, 28, 56, 67, 70, 70, 75, 78, 89, 89, 90, 91, 94, 95, 96, 100, 102, 103, 107, 109, 112

depth = 4 bins

equi-depth

bin1: 24, 25, 26, 27, 28, 56

bin2: 67, 70, 70, 75, 78, 89

bin3: 89, 90, 91, 94, 95, 96

bin4: 100, 102, 103, 107, 109, 112

median

bin1: 26, 26, 26, 26, 26, 26

bin2: 73, 73, 73, 73, 73, 73

bin3: 93, 93, 93, 93, 93, 93

bin4: 105, 105, 105, 105, 105, 105

mean

bin1: 31, 31, 31, 31, 31, 31

bin2: 75, 75, 75, 75, 75, 75

bin3: 93, 93, 93, 93, 93, 93

bin4: 106, 106, 106, 106, 106, 106

boundary

bin1: 24, 24, 24, 24, 24, 56

bin2: 67, 67, 67, 67, 89, 89

bin3: 89, 89, 89, 96, 96, 96

bin4: 100, 100, 100, 112, 112, 112

6. Explain and apply apriori algorithm for discovering frequent items sets of the analyze table

Trans120	Items purchased
101	Milk, bread, eggs
102	Milk, juice
103	juice, butter
104	Milk, bread, eggs
105	coffee, eggs
106	coffee
107	coffee, juice
108	Milk, bread, cookies, eggs
109	cookies, butter
110	Milk, bread

use 0.3 for the minimum support value.

0.3 for minimum support use either decimal value 0.3 or round off no. 3

Step 1: generate frequent 1-itemsets

	Milk	5
	bread	4
	eggs	4
	juice	3
x	butter	2
	coffee	3
x	cookies	2

frequent 1-itemset (support ≥ 3)

{Milk, bread, eggs, juice, coffee}

Step2. Generate Frequent 2-Itemsets

	Milk, bread	4
	milk, eggs	3
x	milk, juice	1
x	milk, coffee	0
	bread, eggs	3
x	bread, juice	0
x	bread, coffee	0
x	eggs, juice	0
x	eggs, coffee	1
x	juice, coffee	1

Frequent 2-Itemsets (support ≥ 3):

{Milk, bread}, {milk, eggs}, {bread, eggs}

Step3 Generate Frequent 3-Itemsets

Milk, bread, eggs	3

Final Frequent Itemsets:

1. Frequent 1-Itemsets

{milk, bread, Eggs, juice, coffee}

2. Frequent 2-Itemsets:

{Milk, bread}, {Milk, Eggs}, {Bread, Eggs}

3 Frequent 3-Itemsets.

{Milk, bread, Eggs}

7. Find all frequent item sets for the given training set using Apriori and FP growth respectively. compare the efficiency of the two mining processes.

TID	ITEMS BOUGHT
T100	{M, O, N, K, E, Y}
T200	{D, O, N, K, E, Y}
T300	{M, A, K, E}
T400	{M, U, C, K, Y}
T500	{C, O, O, K, I, E}

Step1: find the distinct count

X	A	1
X	C	1
X	D	1
	E	4
X	I	1
	K	5
	M	3
X	N	2
	O	4
X	U	1
	Y	3

after step 1:

E	4
K	5
M	3
O	4
Y	3

Step 2: Descending order of values

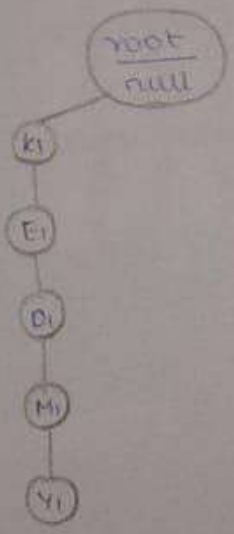
K	5
E	4
O	4
M	3
Y	3

Step 3: redefine the given set:

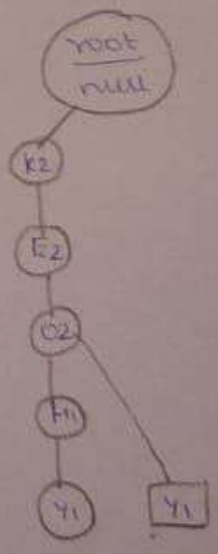
1	K E O M Y
2	K E O Y
3	K E M
4	K M Y
5	K E O

FP Tree:

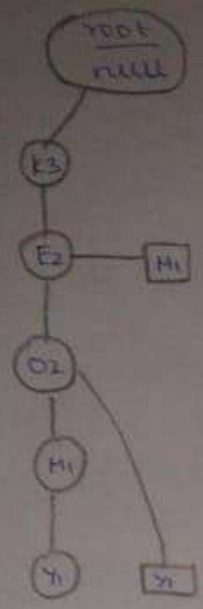
Step 1:



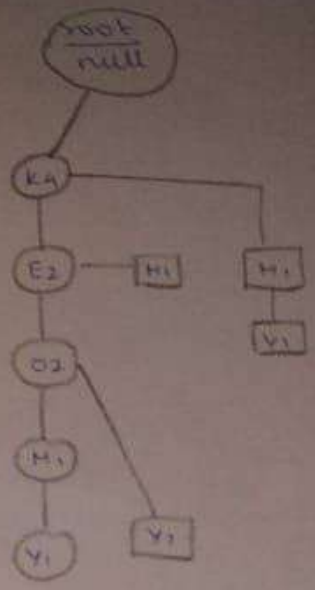
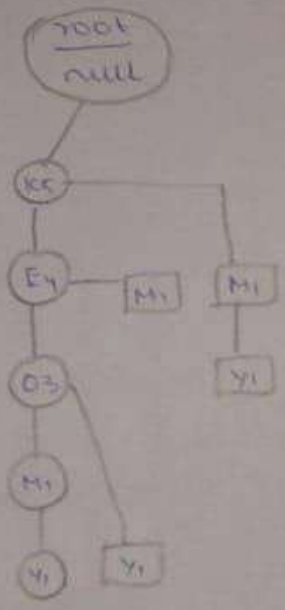
Step 2:



Step 3:



Step 4:



Step 5: