



NATIONAL INSTITUTE OF TECHNOLOGY  
WARANGAL - 506 004  
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING  
II MCA, II Semester  
I Minor Examination, February 2018

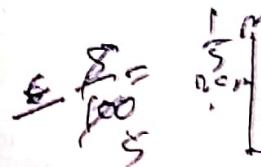
Multi Image Processing  
Time : 1 Hr.

Date: 15-02-2018  
Max. Marks: 10

$\frac{1}{2} = \frac{1}{2} = \frac{1}{2}$

- 1 Out of the three distances: Euclidian, City-block and Chessboard distances, for a given distance 'n', which one deals with more number of pixels. 1/2
- 2 If an image is formed via transmission of illumination through X-ray, then the value at  $f(x,y)$  is a product of illumination at  $(x,y)$  and ..... 1/2
- 3 If the size of the image formed in retina is 5 mm, and the original size is 20 m, what is the distance between the object and the eye in cms? 1/2
- 4 Monotonic transformation performs a (choose all the possible answer(s)) 1/2
  - (a) one-to-one mapping
  - (b) many-to-one mapping
  - (c) one-to-many mapping
  - (d) many-to-many mapping
- 5 What is the difference between Image Enhancement and Image Restoration? 1/2
- 6 Quantization refers to
  - (a) Testing the possible positions of an object in an image.
  - (b) Discretization of the values an image pixel can take..
  - (c) Inversion of the pixel values
  - (d) Discretization of the spatial image domain1/2
- 7 Consider the image segment shown below, Let  $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$ , Compute the lengths of the shortest 8-path between 'a' and 'b'. 1/2

1	2	3	4	5	6	7	8
2	3	4	5	6	7	8	1
3	4	5	6	7	8	1	2
4	5	6	7	8	1	2	3
5	6	7	8	1	2	3	4
6	7	8	1	2	3	4	5
7	8	1	2	3	4	5	6
8	1	2	3	4	5	6	7



$$\text{Distance} = \sqrt{h^2 + b^2} = \sqrt{5^2 + 17^2} = \sqrt{314} = 17.7 \text{ units}$$

$$f_n = \frac{1}{2e^{\pi/2}}$$

$$n = \frac{1}{2e^{\pi/2}} = 0.27$$

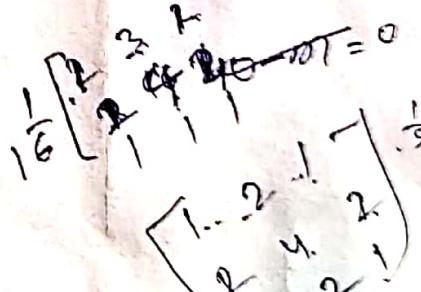
$$\frac{m}{h} = \sqrt{0.27 \times 17^2} = 5.27$$

- 8 Assume that, the image (a) is a gray image with ranges 0 to 255. The size of the image is 256x256. Give a suitable mask so that if we do the AND operation, the resultant image should have only the entire building without the statue. Assume that, the size of the statue is 40x20. 1



$$m=40$$

$$n=20$$



$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$m=256$$

$$n=256$$



NATIONAL INSTITUTE OF TECHNOLOGY  
WARANGAL - 506 004  
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING  
II MCA, II Semester  
I Minor Examination, February 2018

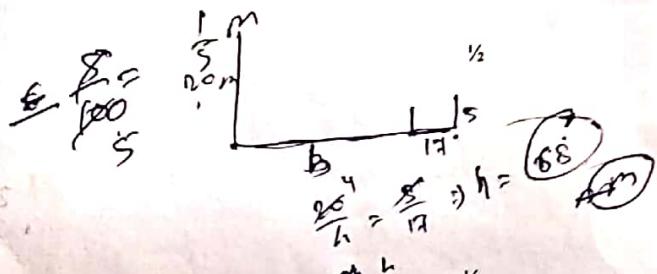
$$\frac{5m}{n} = \frac{500}{50} = 10$$

Sub : Image Processing  
Time : 1 Hr.

Date: 15-02-2018  
Max. Marks: 10

- 1 Out of the three distances: Euclidian, City-block and Chessboard distances, for a given distance 'n', which one deals with more number of pixels. ½
- 2 If an image is formed via transmission of illumination through X-ray, then the value at  $f(x,y)$  is a product of illumination at  $(x,y)$  and  $c \times r$  ½
- 3 If the size of the image formed in retina is 5 mm, and the original size is 20 m, what is the distance between the object and the eye in cms? ½

- 4 Monotonic transformation performs a  
(choose all the possible answer(s))  
(a) one-to-one mapping  
(b) many-to-one mapping  
(c) one-to-many mapping  
(d) many-to-many mapping



- 5 What is the difference between Image Enhancement and Image Restoration?

- Quantization refers to  
(a) Testing the possible positions of an object in an image.  
(b) Discretization of the values an image pixel can take.  
(c) Inversion of the pixel values.  
(d) Discretization of the spatial image domain.

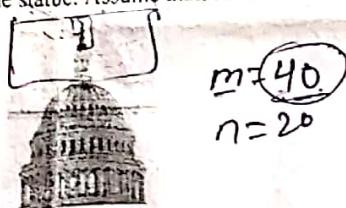
$$h = \frac{1}{20 \times 17} \text{ m} \quad \text{or} \quad h = \frac{1}{20 \times 17} \text{ cm}$$

- 7 Consider the image segment shown below. Let  $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$ . Compute the lengths of the shortest 8-path between 'a' and 'b'.

1	2	3	4	5	6	7	8
2	3	4	5	6	7	8	1
3	4	5	6	7	8	1	2
4	5	6	7	8	1	2	3
5	6	7	8	1	2	3	4
6	7	8	1	2	3	4	5
7	8	1	2	3	4	5	6
8	1	2	3	4	5	6	7

$$\frac{20}{h} = \frac{1}{20 \times 17}$$

- 8 Assume that, the image (a) is a gray image with ranges 0 to 255. The size of the image is  $256 \times 256$ . Give a suitable mask so that if we do the AND operation, the resultant image should have only the entire building without the statue. Assume that, the size of the statue is  $40 \times 20$ .



$$m=40$$

$$n=20$$

1. 1. 2. 3

4. 5. 6

7. 8. 9

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}_{m \times n}$$

$$m=256$$

$$N=2^{8 \times 8}$$

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 1 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}_{16 \times 16}$$

- 9 Suppose that a digital image is subjected to histogram equalization. What will be the effect of a second pass of histogram equalization (on the histogram-equalized image). Verify for the following data.

$r_k$	$n_k$	$p_r(r_k) = n_k/MN$
$r_0 = 0$	790	0.19 <del>0.225</del>
$r_1 = 1$	1023	0.25
$r_2 = 2$	850	0.21
$r_3 = 3$	656	0.16
$r_4 = 4$	329	0.08
$r_5 = 5$	245	0.06
$r_6 = 6$	122	0.03
$r_7 = 7$	81	0.02

- 10 Consider the sub image shown below. If each pixel is represented with 3-bits. What are the contents of each of the bit plane images.

$$\begin{bmatrix} 7 & 7 & 7 \\ 7 & 7 & 7 \\ 7 & 7 & 7 \end{bmatrix}$$

$$2^3 = 8$$

$$3 \times 3 \times 3 = 27$$

- 11 (a) Give the condition(s) under which the  $D_4$  distance between two points  $p$  and  $q$  is equal to the shortest 4-path between these points.  
 (b) Is this path Unique?

- 12 Give a single intensity transformation function for spreading the intensities of an image so the lowest intensity is 0 and the highest is  $L-1$ . Show your transformation results for the following sub image for which  $L-1$  is 15.

- 13 Sketch the set  $(A \cap B) \cup (A \cup B)^c$



12

$$\begin{bmatrix} 2 & 3 & 4 & 5 \\ 5 & 4 & 3 & 2 \\ 6 & 7 & 8 & 9 \\ 9 & 8 & 7 & 6 \end{bmatrix}$$

$17/2$



NATIONAL INSTITUTE OF TECHNOLOGY  
WARANGAL - 506 004  
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING  
II MCA, II Semester  
MID Examination, February 2018

Date: 22-02-2018  
Max. Marks: 30

Sub : Image Processing  
Time : 2 Hrs.

4M

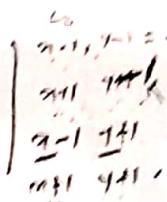
- 1 You are preparing a report and have to insert in it an image of size  $2048 \times 2048$  pixels.  
 (a) Assuming no limitations on the printer, what would be the resolution per mm have to be for the image to fit in a space of size  $5 \times 5$  cm?  
 (b) What would the resolution have to be in dpi for the image to fit in  $2 \times 2$  inches?

- 2 (a) Consider the two image subsets  $S_1$  and  $S_2$ , shown in the following figure. For  $V = \{0\}$ , determine whether these two subsets are (a) 4-adjacent, (b) 8-adjacent, or (c) m-adjacent.

6M



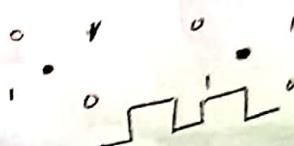
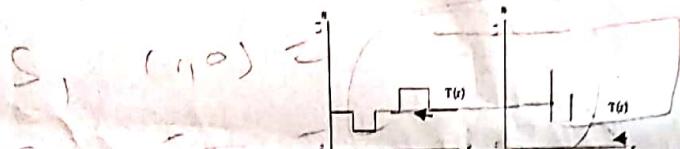
$$\begin{array}{l} n=1, \gamma=0 \\ n=1, \gamma=0 \\ n=1, \gamma=0 \\ n=1, \gamma=1 \end{array}$$



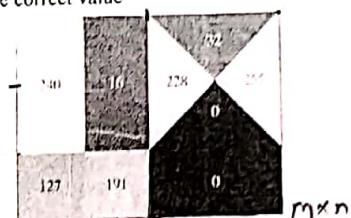
- (b) Consider the three image segments  $S_1$ ,  $S_2$  and  $S_3$ . If  $S_1$  is m-adjacent to  $S_2$  and  $S_2$  is m-adjacent to  $S_3$ . Can we say  $S_1$  is m-adjacent to  $S_3$ ? Justify, with an example.

4M

- 3 What is the difference between the following two types of Intensity-level slicing. What is the type of the images (gray/binary/...) produced by these slicing. What changes we see in the resultant images?



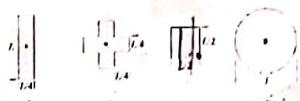
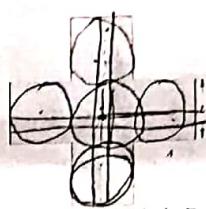
- 4 Obtain the un-normalized and the normalized histograms of the following 8-bit,  $M \times N$  image. Give your histogram either in a table or a graph, labeling clearly the value and location of each histogram component in terms of  $M$  and  $N$ . Double check your answer by making sure that the histogram components add to the correct value



6M

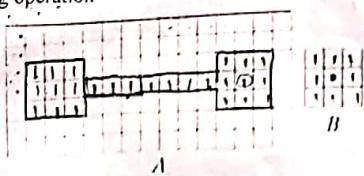
- 5 Let  $A$  denote the set shown shaded in the following figure. Refer to the structuring elements shown (the black dots denote the origin). Sketch the result of the following morphological operations:  
 (i)  $(A \oplus B_4) \ominus B_2$  (ii)  $(A \oplus B_1) \ominus B_3$ .

4M



- 6 Show all intermediate steps of your computations for the following:  
 (a) Obtain the opening of the figure below using a  $3 \times 3$  SE of 1s. Do all operations manually.  
 (b) Repeat (a) for the closing operation

6M





NATIONAL INSTITUTE OF TECHNOLOGY  
WARANGAL - 506 004  
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING  
II MCA, II Semester  
Minor-II Examination, April 2018

Sub : Image Processing  
Time : 30 mins.

Date: 20-04-2018  
Max. Marks: 10

- What are the different types of redundancies in images/videos? Explain and also give the model of image compression. 3M
- If the given image is result of i-level decomposed Haar Wavelet Transform result. Give the original image. 3M

$$\begin{array}{c} \text{+} \\ \boxed{\begin{bmatrix} 12 & 13 & 14 & 13 & 4 & -2 & 6 & 3 \\ 13 & 12 & 11 & 10 & 4 & 5 & -1 & 0 \\ 12 & 13 & 9 & 14 & 3 & 5 & 3 & 2 \\ 13 & 15 & 9 & 9 & -1 & -2 & 0 & -6 \\ -4 & 0 & 1 & 3 & 2 & -1 & 2 & \\ 2 & 4 & -1 & 5 & 5 & -2 & -1 & 3 \\ -1 & -4 & 0 & -1 & -2 & -1 & -6 & -4 \\ -3 & -4 & -1 & -5 & 0 & -2 & -2 & -1 \end{bmatrix}} = \begin{bmatrix} 10 & 9 & 14 & 14 & 7 & 0 & 5 & 5 \\ 14 & 17 & 14 & 12 & 1 & -4 & 7 & 1 \\ 15 & 16 & 10 & 15 & 9 & 3 & -2 & +3 \\ 13 & 8 & 12 & 5 & -1 & 7 & 0 & -3 \\ 18 & 9 & 9 & 13 & 1 & 4 & -3 & -2 \\ 10 & 17 & 9 & 18 & 5 & 6 & 9 & 6 \\ 16 & 11 & 8 & 18 & 5 & 6 & 9 & 6 \\ 16 & 19 & 10 & 4 & 3 & -4 & -2 & -7 \end{bmatrix} \end{array}$$

- Consider the following three layers of size  $4 \times 4$  as R, G and B of a color image and  $2 \times 2$  M. 3M  
are the layers of color watermark. Give the result after keeping (into LSBs) the watermark into the color image such that:

Original	Watermark
R.	G
G	R
B	B

Original Image	R	G	B
	$\begin{bmatrix} 7 & 9 & 5 & 8 \\ 4 & 2 & 7 & 8 \\ 4 & 5 & 7 & 6 \\ 1 & 2 & 3 & 5 \end{bmatrix}$	$\begin{bmatrix} 27 & 15 & 25 & 8 \\ 14 & 2 & 7 & 8 \\ 54 & 5 & 7 & 6 \\ 1 & 2 & 31 & 5 \end{bmatrix}$	$\begin{bmatrix} 9 & 9 & 4 & 8 \\ 4 & 2 & 12 & 8 \\ 14 & 15 & 7 & 26 \\ 1 & 2 & 13 & 5 \end{bmatrix}$
Watermark	$\begin{bmatrix} 2 & 3 \\ 1 & 3 \end{bmatrix}$	$\begin{bmatrix} 0 & 3 \\ 1 & 2 \end{bmatrix}$	$\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$
Note: Assume that 2-bits are enough to represent the pixels in watermark image			

$$0 = 0^0$$

$$3 = 1^1$$

$$1 = 2^0$$

$$2 = 1^1$$

$$9 = 1^2$$

$$4 = 2^1$$

$$1 = 1^1$$

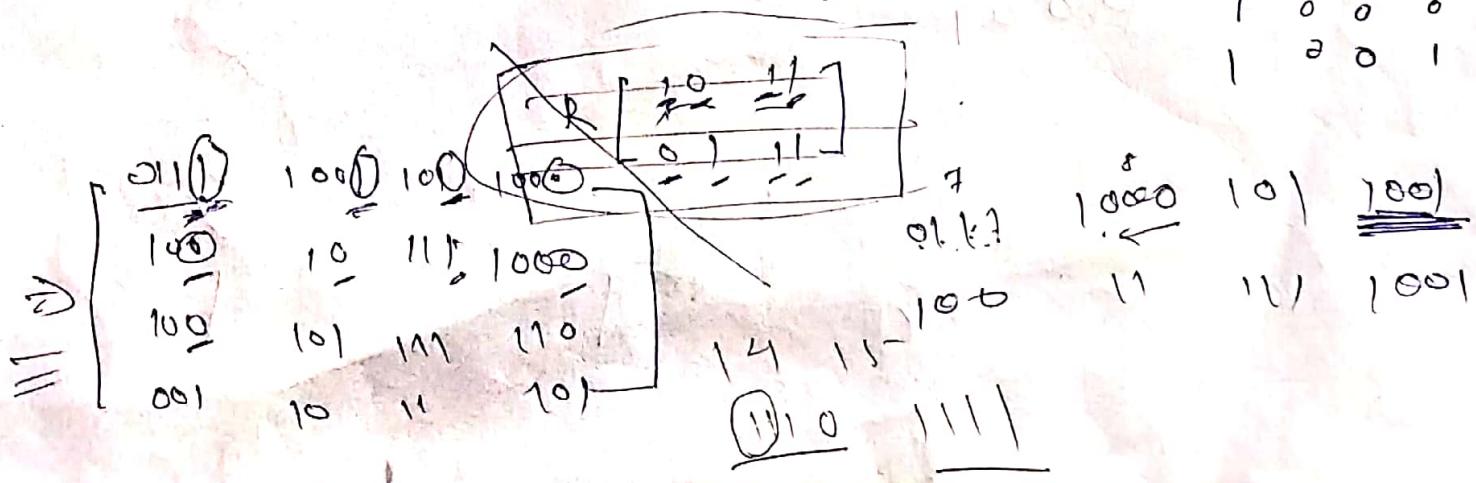
$$0 = 0^0$$

$$1 = 0^0$$

$$2 = 0^1$$

$$1 = 1^1$$

$$0 = 0^0$$





NATIONAL INSTITUTE OF TECHNOLOGY  
WARANGAL - 506 004  
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING  
II MCA, II Semester  
END Examination, May 2018

Sub : Image Processing  
Time : 3 Hrs.

Date: 04-05-2018  
Max. Marks: 50

5.30  
English

(DP)  
In  
the  
y  
of  
its 3 M

t 3 M

3 M

3 M

3 M

3 M

3 M

3 M

3 M

3 M

3 M

3 M

3 M

Scanned by CamScanner

- 1 What is Image Resolution? Write all the definitions you know and justify the definition in which context we use it?  $\text{ppi}, \text{dpi}, \text{api}$  5M
- 2 Let the following is the subimage, where each pixel is represented with 8-bits. Obtain all the bit plane images of this. 6M

$$\begin{bmatrix} 0 & 1 & 6 \\ 7 & 2 & 4 \\ 6 & 5 & 3 \end{bmatrix}$$

3-bit plane  
8-bit plane

first  
words  
are  
etc

$y_1$   
 $y_2$

$y_3$   
 $y_4$

$y_5$   
 $y_6$

$y_7$   
 $y_8$

- 3 What is morphology? What are the different morphological operations on Binary Images?
- 4 Obtain the Morphological Gradient for the following gray image.  $\text{dilation} (I \oplus T)$  6M

$$\begin{bmatrix} 2 & 1 & 3 & 6 & 7 & 2 \\ 0 & 2 & 5 & 4 & 7 & 3 \\ 6 & 2 & 5 & 4 & 1 & 6 \\ 3 & 5 & 4 & 1 & 2 & 2 \\ 5 & 4 & 1 & 3 & 6 & 5 \\ 2 & 5 & 4 & 1 & 6 & 3 \end{bmatrix}$$

$$(I \oplus T) \rightarrow \text{gradient}$$

$$V_m(n) \rightarrow V_m(n)$$

$$V_m(n) \rightarrow V_m(n)$$

$$1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7$$

- 5 Consider the following color image and intensity distributions. Get the histogram equalized image color image for this. 7M

Value	RED Band	GREEN Band	BLUE Band	cdf
0	790	656	245	1691
1	1023	329	122	1414
2	850	245	81	1176
3	656	122	790	1458
4	329	81	1023	1433
5	245	790	850	1685
6	122	1023	656	1801
7	81	850	329	1260

$$11178 \text{ cdf}$$

- 6 What is the method used in Haar forward Wavelet Transform? By using this method give the results after 2-level decomposition of the given image. 7M

$$\begin{bmatrix} 7 & 5 & 9 & 2 & 4 & 8 & 7 & 5 \\ 9 & 6 & 5 & 8 & 7 & 4 & 9 & 5 \\ 4 & 7 & 8 & 9 & 5 & 4 & 7 & 1 \\ 5 & 8 & 7 & 6 & 9 & 7 & 4 & 2 \\ 5 & 8 & 7 & 9 & 5 & 7 & 8 & 6 \\ 5 & 7 & 6 & 8 & 7 & 4 & 9 & 5 \\ 8 & 5 & 8 & 7 & 4 & 2 & 8 & 7 \\ 9 & 5 & 7 & 8 & 4 & 5 & 8 & 3 \end{bmatrix} \quad \begin{array}{ccccccccc} 6.75 & 6 & 5.675 & 6.5 & 1.25 \\ 6 & 7.5 & 5 & 3.5 & -1.5 \\ 6.25 & 7.25 & 6.75 & 7.5 & -1.25 \\ 6.75 & 7.5 & 3.75 & 6.5 & .25 \\ -.75 & -0.5 & .25 & -0.5 & -0.25 \\ -.5 & 1 & .25 & -0.5 & -0.25 \\ .25 & -.75 & -.5 & 0 & 0 \\ 0 & 0 & -.75 & -1.75 & -1.75 \end{array}$$

- 7(a) What is Digital Watermarking? What are the applications of this. 25  
(b) What is visible watermarking and invisible watermark? Give the procedure how to do these two techniques on a image.

- 8 Use LZW compression technique and give the result for the following image. 6M

$$\begin{bmatrix} 39 & 39 & 126 & 126 \\ 39 & 39 & 126 & 126 \\ 39 & 39 & 126 & 126 \\ 39 & 39 & 126 & 126 \end{bmatrix}$$

$$\begin{array}{cccccc} 1 & -0.25 & 1.5 & 0 \\ -0.5 & 2 & 2 & 0 \\ -0.75 & 0.25 & 1.5 & 0 \\ 0 & 0.25 & 1.5 & 0 \\ 2.5 & -1.75 & 0.5 & 0 \\ 0 & -1.5 & -1 & 0 \\ -0.25 & -1.25 & 0.5 & 0 \\ -0.5 & -0.75 & 1 & 0 \end{array}$$

**Department of Computer Science & Engineering  
National Institute of Technology, Warangal**

Minor Test - II

CS 6351 Principles of Data Warehousing and Data Mining  
II year - II semester MCA  
Answer ALL Questions

15

Marks: 20

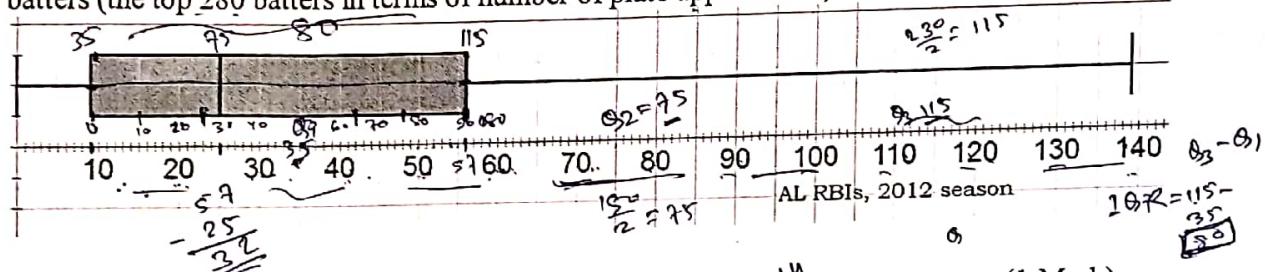
Time: 1 Hour.

**Data Warehousing Part**

280

22

The following boxplot shows the 2012 season runs batted in (RBIs) of 280 American League batters (the top 280 batters in terms of number of plate appearances).



Q1) What is the size of the IQR of this distribution?

14

(1 Mark)

Q2) How many AL hitters hit more than 25 RBIs in 2012?

280x 75  
Top

(2 Marks)

120

Q3) B. J. Upton of the Tampa Bay Rays hit 78 RBIs in 2012; this is the 90th percentile value on this chart. How many players hit between 56 and 78 RBIs?

(3 Marks)

Q4) What issues are to be considered into account in data integration phase of data warehouse? How do you resolve them? (Explain each technique in detail)

(2 Marks)

Q5) Determine least squares linear prediction equation for the data:

(2 Marks)

x	0	1	2	3	4
y	-5	-1	0	2	4

$$\frac{61}{280 \times 22} = 70.2$$

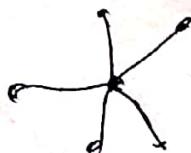
$$\begin{aligned} & 280 \times 78 = 960 \\ & 280 \times 22 = 6100 \end{aligned}$$

**Data Mining Part**

22X

A) Discuss the issues in constructing Decision tree for a given training data and write an algorithm that produces Decision tree. 3 + 4 marks

B) Write Bayes theorem? Explain how it is used in constructing classifier. 3 marks



$$T_s \leftarrow 4.2 + 2.1x$$



Answer all the Questions.

$x \subseteq Y$

- 1 Let  $Y$  be an itemset and  $X$  is a subset of  $Y$ . Prove the statement "If a rule  $X \rightarrow (Y - X)$  does not satisfy the confidence threshold, then any rule  $Z \rightarrow (Y - Z)$ , where  $Z$  is a subset of  $X$ , must not satisfy the confidence threshold as well".

3

- 2 Find Frequent itemsets valid on the following dataset using FP-growth algorithm. Minimum support=25%

4

TID	List of Items						
1	E, G, H	6	B, G, I	11	B, C, G	16	A, C, D, E
2	A, B, E, G	7	C, D, E, E	12	A, D, F, H	17	B, D, F, I
3	B, C, D, E, I	8	A, D, E, I	13	A, C, D, E, I	18	A, D, G, H
4	A, G, I	9	B, Q	14	D, G, H	19	C, E, G, H, I
5	A, B, D, H	10	A, D, E, F	15	G, B	20	B, D, F

- 3 Suppose  $X$  be a candidate k-itemset. Show that if  $X$  has a subset of size less than  $k - 1$  that is infrequent, then at least one of the  $(k - 1)$ -size subsets of  $X$  is necessarily infrequent.

3

- 4 State why, for integration of multiple heterogeneous information sources, many companies prefer the update driven approach ( which constructs and uses data warehouses), rather than query drive approach ( which applies wrappers and integrators). Describe situations where the query driven approach is preferable over the update driven approach

2

- 5 Design data warehouse for wholesale furniture company. Your data warehouse should allow to analyze the company's situation at least with respect to the Furniture, Customers and Time. Moreover, the company needs to analyze:

8

- the furniture with respect to its type (chair, table, wardrobe, cabinet...), category (kitchen, living room, bedroom, bathroom, office...) and material (wood, marble...)
- the customers with respect to their spatial location, by considering at least cities, regions and states.

The company is interested in learning at least the quantity, income and discount of its sales.

a) Identify facts, dimensions and measures. (1 Mark)

b) Design snowflake schema (2 Marks)

c) Write SQL Analytical queries:

I. Find the quantity, the total income and discount with respect to each city, type of furniture and the month (3 Marks)

II. Find the average quantity, income and discount with respect to each country, furniture material and year (2 Marks)

III. Determine the 5 most sold furniture's during the November month. (2 Marks)

25/02/2021



**CS 6351 PRINCIPLES OF DATA WAREHOUSING & DATA MINING**

End Semester Examination

Date: 01-05-2018

Time: 2.30PM to 5.30PM

**II year MCA - II Semester**

Total Marks: 50 M

*Answer ALL the questions.*

1	Suppose sales price for 12 sales price records has been sorted as follows: 5, 10, 11, 13, 15, 35, 50, 55, 72, 92, 204, 215 Partition them into three bins using i) equal-frequency partitioning ii) equal width partitioning and iii) clustering	100 B	1M 1M 3M																																																																												
2	A) What are the value ranges for z-score normalization and normalization by decimal scaling. B) Write process to detect redundancy when integrating data in data warehousing? Discuss your approach with an example.		2M 3M																																																																												
3	Propose an algorithm for automatic generation of concept hierarchy for numerical data based on equal width partition rule.		5M																																																																												
4	A) How do you transform star schema into snowflake schema? Discuss with an example. B) What is a fact less fact table? Design simple star schema with a fact less fact table to keep track of patients in a hospital by diagnostic procedures and time		2M 3M																																																																												
5	A) The median is one of the most important <u>holistic measures</u> in data analysis. Propose any three methods for median approximation. Analyze their respective complexity under different parameter settings and decide to what extent the real value can be approximated. Also suggest a heuristic strategy to balance between accuracy and complexity and then apply it to all the proposed methods. B) Discuss OLAP operations with examples		3M 2M																																																																												
6	Explain any two common approaches for pruning decision trees.		2M																																																																												
7	Illustrate Classification with ANN using Back Propagation Algorithm.		4M																																																																												
8	A) Explain the working of k-means algorithm and its limitations. B) Use the K-means algorithm to cluster the following eight points into <u>three clusters</u> A(2,10), B(2,5), C(8,4), D(5,8), E(7,5), F(6,4), G(1,2), H(4,9).		4M 5M																																																																												
9	Find the nodes created at level 0 and level 1 during <u>decision tree</u> classification model to classify bank loan applications by assigning applications to one of the three risk classes. <table border="1"><thead><tr><th></th><th>Owes home</th><th>Married</th><th>Gender</th><th>Employed</th><th>Credit rating</th><th>Risk class (Class Label)</th></tr></thead><tbody><tr><td>1</td><td>Yes</td><td>Yes</td><td>Male</td><td>Yes</td><td>A</td><td>B</td></tr><tr><td>2</td><td>No</td><td>No</td><td>Female</td><td>Yes</td><td>A</td><td>A</td></tr><tr><td>3</td><td>Yes</td><td>Yes</td><td>Female</td><td>Yes</td><td>B</td><td>C</td></tr><tr><td>4</td><td>Yes</td><td>No</td><td>Male</td><td>No</td><td>B</td><td>B</td></tr><tr><td>5</td><td>No</td><td>Yes</td><td>Female</td><td>Yes</td><td>B</td><td>C</td></tr><tr><td>6</td><td>No</td><td>No</td><td>Female</td><td>Yes</td><td>B</td><td>A</td></tr><tr><td>7</td><td>No</td><td>No</td><td>Male</td><td>No</td><td>B</td><td>B</td></tr><tr><td>8</td><td>Yes</td><td>No</td><td>Female</td><td>Yes</td><td>A</td><td>A</td></tr><tr><td>9</td><td>No</td><td>Yes</td><td>Female</td><td>Yes</td><td>A</td><td>C</td></tr><tr><td>10</td><td>Yes</td><td>Yes</td><td>Female</td><td>Yes</td><td>A</td><td>C</td></tr></tbody></table>		Owes home	Married	Gender	Employed	Credit rating	Risk class (Class Label)	1	Yes	Yes	Male	Yes	A	B	2	No	No	Female	Yes	A	A	3	Yes	Yes	Female	Yes	B	C	4	Yes	No	Male	No	B	B	5	No	Yes	Female	Yes	B	C	6	No	No	Female	Yes	B	A	7	No	No	Male	No	B	B	8	Yes	No	Female	Yes	A	A	9	No	Yes	Female	Yes	A	C	10	Yes	Yes	Female	Yes	A	C	4M
	Owes home	Married	Gender	Employed	Credit rating	Risk class (Class Label)																																																																									
1	Yes	Yes	Male	Yes	A	B																																																																									
2	No	No	Female	Yes	A	A																																																																									
3	Yes	Yes	Female	Yes	B	C																																																																									
4	Yes	No	Male	No	B	B																																																																									
5	No	Yes	Female	Yes	B	C																																																																									
6	No	No	Female	Yes	B	A																																																																									
7	No	No	Male	No	B	B																																																																									
8	Yes	No	Female	Yes	A	A																																																																									
9	No	Yes	Female	Yes	A	C																																																																									
10	Yes	Yes	Female	Yes	A	C																																																																									
10	A) Explain the working of DBSCAN algorithm B) Explain the multiphase clustering technique applied in BIRCH algorithm. Why is it that BIRCH encounters difficulties in finding clusters of arbitrary shapes?		3M 3M																																																																												



Date: 21-02-2018

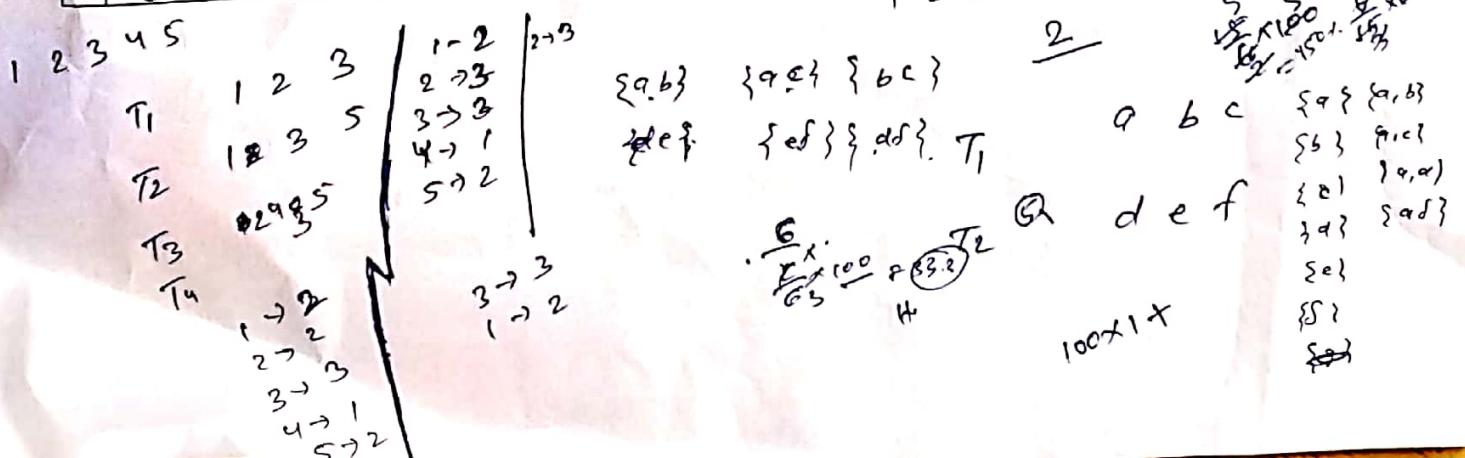
Time: 2.30PM to 4.30PM

II year – II Semester MCA

Total Marks: 30M

*Answer all the Questions.*

1	a) What is the Column capacity in ROLLUP, CUBE and GROUPING SET functions? b) Write the equivalent Group by statement for GROUP BY GROUPING SETS(x, (y), ( ))? c) Write the equivalent Group By statement for GROUP BY GROUPING SETS(x, ROLLUP(y, z))? d) Write the equivalent query using four Group by clauses for ROUP BY CUBE ((x, y), z)? e) How many groupings exists for GROUP BY GROUPING SETS (a, b), GROUPING SETS(c, d)? $\leq 4$	1.5M 1.5M 1.5M 1.5M 1.5M
2	a) How is data warehouse schema design is different from database schema design and what are its advantages? b) Why fact table is larger than dimension table c) Define Junk Dimension. In what way is it different from de-generate dimension. Provide two examples for each type	2M 1M 1M
3	a) Compare OLAP systems with OLTP with respect to <u>orientation, user, unit of work</u> , number of users accessed, priority and metric features b) Suppose that a data warehouse for Big University consists of the following: four dimensions: <u>student, course, semester, and instructor</u> , and two <u>measures count and avg_grade</u> . When at the lowest conceptual level (e.g. for a given student, course, semester, and instructor combination), the <u>avg_grade</u> measure stores the actual course grade of the student. At higher conceptual levels, <u>avg_grade</u> stores the average grade for the given combination. Draw a snowflake schema diagram for the data warehouse. Clearly mention all your assumptions.	1.5M 2M
4	Direct Hashing and Pruning (DHP) is an improvement over Apriori. Let {a, b, c} and {d, e, f} are two transactions and all 6 items are frequent. Then find the <u>percentage improvement</u> in pruning the number of candidate 2-itemsets generated in DHP over Apriori.	3M
5	a. List any two reasons that motivated the development of FP-Growth algorithm. b) Write FP-growth algorithm for discovering frequent itemsets. c. In FP-tree mining, ordering of items by decreasing support is used and it always lead to the smallest tree. Do you agree with this statement? Justify your answer	2M 4M 2M
6	Consider the following set of frequent 3-itemsets: {1, 2, 3}, {1, 2, 4}, {1, 2, 5}, {1, 3, 4}, {1, 3, 5}, {2, 3, 4}, {2, 3, 5}, {3, 4, 5}. Assume that there are only five items in the data set. (a) List all candidate 4-itemsets obtained by a candidate generation procedure using the L <sub>k-1</sub> × L <sub>1</sub> merging strategy. (b) List all candidate 4-itemsets obtained by the candidate generation procedure in <u>Apriori</u> . (c) List all candidate 4-itemsets that survive the candidate pruning step of the <u>Apriori</u> algorithm.	1.5M 1.5M 1M





NATIONAL INSTITUTE OF TECHNOLOGY, WARANGAL  
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING  
II MCA II Semester - End Semester Examinations - April/May 2018  
Algorithm Analysis and Design (CS6351)

Date: 27-04-2018

Time: 2.30 PM to 5.30 PM

Max. Marks: 50

Note: Answer all Questions. Provide a clear, neat, and complete pseudo code for your algorithm in English.

1. Formally define the asymptotic notations big-oh, big-omega and theta with suitable examples. Does there exist functions  $f(n)$  and  $g(n)$  such that neither  $f(n)=O(g(n))$  nor  $f(n)=\Omega(g(n))$ ? Justify your answer. 4 M
2. Give asymptotic upper and lower bounds for  $T(n)$  in each of the following recurrences. Assume that  $T(n)$  is constant for sufficiently small  $n$ . Make your bounds as tight as possible, and justify your answers. 4 M
  - a)  $T(n) = T(n-1) + \log n$
  - b)  $T(n) = \sqrt{n} T(\sqrt{n}) + n$
3. For each of the following algorithms, give a recurrence for the worst-case running time and represent it using O-notation. 6 M
  - a). Binary search  $T(n) = T(n/2) + 1 \Rightarrow O(\log n)$
  - b). Quicksort  $\Rightarrow T(n) = O(n^2)$
  - c). Strassen's matrix multiplication  $\Rightarrow T(n) = 8T(n/2) + O(n^2) \Rightarrow T(n) = 7T(n/2) + 18(n/2)^2 \Rightarrow n^3$
  - d). Karatsuba's algorithm for large integer multiplication  $T(n) = 3T(n/2) + O(n) \Rightarrow O(n^1.5) \approx O(n \log n)$
4. State selection problem. Give the worst-case linear time algorithm for the selection problem discussed in the class (which divides an array of size  $n$  into  $n/5$  groups, as equally sized as possible). Devise a recurrence for the running time of the algorithm and hence show that it runs in linear time. 4 M
 
$$T(n) = T(n/5) + T(2n/5) + 3D(n)$$

What are the two ingredients an optimization problem must have to apply dynamic programming technique? In dynamic programming, we derive a recurrence relation for the solution to one subproblem in terms of solutions to other subproblems. To turn this relation into a bottomup dynamic programming algorithm, we need an order to fill in the solution cells in a table, such that all needed subproblems are solved before solving a subproblem. For each of the following relations, give such a valid traversal order, or if no traversal order is possible for the given relation, briefly justify why. 5 M

  - a).  $A(i, j) = \min\{A(i-1, j), A(i, j-1), A(i-1, j-1)\}$
  - b).  $A(i, j) = \max\{A(\min\{i, j\}-1, \min\{i, j\}-1), A(\max\{i, j\}-1, \max\{i, j\}-1)\}$
  - c).  $A(i, j) = \min\{A(i-2, j-2), A(i+2, j+2)\}$
5. Define independent set of a graph and give an example. Devise a polynomial time dynamic programming solution to find the size of a maximum independent set of a tree. Analyze the running time of your algorithm. 4 M
6. Solve the following instance of 0/1 knapsack problem by applying dynamic programming algorithm. 4 M
 

Item No	Weight (Kg)	Profit (Rs)
1	2	1
2	3	2
3	4	5

Knapsack capacity = 6 Kg.

$$V(i, j) = \max \{ V(i-1, w), V(i-1, w-w_i) + p_i \}$$
7. Define a Minimum Spanning Tree (MST). With pseudo code explain Prim's algorithm to find an MST of a simple, undirected, weighted and connected graph. Analyze the running time of Prim's algorithm when the priority queue is implemented as i) a binary heap and ii) a Fibonacci heap data structure. 6 M
8. Consider the following decision problem.  
Vertex-Cover-Decison-Problem (NCDP)  
Input : A simple, undirected graph  $G(V, E)$  and a positive integer  $r \leq |V(G)|$   
Question : Does  $G$  have a vertex cover of size at most  $r$ ?  
Show that NCDP is NP-complete by reducing from the Clique decision problem. 4 M
9. Define the terms: vertex cover and the maximal matching. Give a 2-approximation algorithm for finding an optimal vertex cover of a simple, undirected and connected graph. Analyze the running time of your algorithm and show that the algorithm has an approximation ratio 2. Can the approximation guarantee of your algorithm be improved by a better analysis? Justify your answer. 6 M
10. A sequence of  $n$  operations is performed on a data structure. The  $i^{\text{th}}$  operation costs  $i$  if  $i$  is an exact power of 2, and 1 otherwise. Use the aggregate analysis to determine the amortized cost per operation. 3 M

--- "The pain you feel today is the strength you will feel tomorrow. For every challenge encountered there is opportunity for growth." ---



NATIONAL INSTITUTE OF TECHNOLOGY, WARANGAL  
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING  
II MCA II Semester - Mid Examinations - February 2018  
Algorithm Analysis and Design (CS6351)

$\leq T(n/2) +$   
 $n^{1/2} = O(n)$

Date: 19-02-2018

Time: 2.30 PM to 4.30 PM

Max. Marks: 30

Note: Answer all Questions. Provide a clear, neat, and complete pseudo code for your algorithm in English.

1. What are the two ingredients an optimization problem must have to apply dynamic programming (DP) technique? Consider a variant of the matrix-chain multiplication problem (discussed in the class) in which the goal is to parenthesize the sequence of matrices so as to maximize, rather than minimize, the number of scalar multiplications. Does this problem exhibit optimal substructure? Justify your answer. 3 M

optimal  
substructure

2. Devise an algorithm to find both the minimum and the maximum elements from a set of  $n$  elements using at most  $3 \lceil n/2 \rceil$  comparisons. 3 M

3. Consider the following modification to mergesort, called 4-way mergesort: Instead of dividing the input array into two subarrays, 4-way mergesort divides the array into four subarrays (as equally-sized as possible), sorts them recursively, and then calls a 4-way merge function, which combines four sorted arrays into one sorted array. Let  $T(n)$  represents the worst case number of comparisons used in 4-way mergesort given an array of  $n$  elements as input. Give a recurrence for  $T(n)$  and represent it using  $\Theta$ -notation. 3 M

$\Theta(n + \frac{n}{4} + \frac{n}{16} + \dots)$

4. An array  $A[1\dots n]$  is unimodal if it consists of an increasing sequence followed by a decreasing sequence. More precisely, if there is an index  $m$  in  $\{1, 2, \dots, n\}$  such that  $A[i] < A[i+1]$  for  $1 \leq i < m$  and  $A[i] > A[i+1]$  for  $m \leq i < n$ . In particular,  $A[m]$  is the maximum element, and it is the unique "locally maximum" element surrounded by smaller elements. Design an algorithm to compute the maximum element of a unimodal input array  $A[1\dots n]$  in  $O(\log n)$  time. 3 M

5. With pseudo code explain the Karatsuba's algorithm to multiply two  $n$ -digit integers efficiently. Let  $T(n)$  be the time required to multiply two  $n$ -digit integers by Karatsuba's method. Give an exact recurrence for  $T(n)$  and hence show that  $T(n) = O(n^2)$ . 3 M

6. Consider the following modifications to pivot selection in quick sort:  
 a). Suppose quicksort always pivoted on the  $(n-4)^{\text{th}}$  smallest value of the current sub-array of length  $n$  i.e., if there were 10 elements then sixth smallest element would have been chosen as pivot. Give a recurrence for the worst-case running time of quicksort in this case and represent it using  $\Theta$ -notation.  
 b). Suppose quicksort were always to pivot on the  $\lceil n/4 \rceil^{\text{th}}$  smallest value of the current sub-array of length  $n$ . Give a recurrence for the running time of quicksort in this case and represent it using  $\Theta$ -notation. 3 M

7. Use the Master method (discussed in the class) to give tight asymptotic bounds for the following recurrences. 3 M

$$\begin{aligned} i. T(n) &= 9 T(n/3) + 4 n^2 + 2 \\ ii. T(n) &= 4 T(n/2) + 4 n^2 \log n + n + 1 \\ iii. T(n) &= 16 T(n/4) + 8 n^3 + 4 \end{aligned}$$

$$\begin{aligned} &\approx [2T(n/3) + 1] + 1 \\ &2^2 T(n/3) + 2^2 + 1 \\ &2^3 T(n/3) + 2^3 + 1 \end{aligned}$$

8. A Professor wishes to develop a matrix-multiplication algorithm that is asymptotically faster than Strassen's algorithm. His/her algorithm will use the divide-and-conquer method, dividing each matrix into pieces of size  $n/8 \times n/8$ , and the divide and combine steps together will take  $\Theta(n^2 \log n)$  time. He/she needs to determine how many subproblems his/her algorithm has to create in order to beat Strassen's algorithm. If his/her algorithm creates  $p$  subproblems, then write the recurrence for the running time  $T(n)$  of Professor's algorithm. What is the largest integer value of  $p$  for which Professor's algorithm would be asymptotically faster than Strassen's algorithm? 3 M

$\log_2 8 = 3$

9. State selection problem. Consider the worst-case linear time algorithm for the selection problem discussed in the class (which divides an array of size  $n$  into  $\lceil n/5 \rceil$  groups, as equally sized as possible). Consider the following modification to the selection algorithm, instead of dividing the array of size  $n$  into  $n/5$  groups:

"Suppose we always divide  $n$  elements into  $\lceil n/3 \rceil$  groups of size 3"

- Give a recurrence relation for the running time of the modified selection algorithm. 3 M

10. Suppose we are given an array  $A$  of  $n$  distinct elements, and we want to find  $n/2$  elements in the array whose median is also the median of  $A$ . Any algorithm that does this must take  $\Omega(n \log n)$  time in the worst-case. 3 M

Indicate whether the above statement is true or false, and briefly explain why?

---- Life is 10% what happens to you and 90% how you react to it! ----