1

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination: 2011-2012

Course Name: M. Tech. (QR & OR) 2nd YEAR

Subject Name: Advanced Statistical Methods

Date of Examination: 05.09.11 Maximum Marks: 50 Duration: 2 hours

Note:
1. This paper carries 60 marks.
2. Answer all four questions but the maximum you can score is 50.
3. All notations have their usual meanings

1) Let $\mathbf{Y} \sim N_4(\mu, \Sigma)$, where

$$\mu = \begin{bmatrix} 4 \\ 6 \\ 9 \\ 3 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} 3 & 5 & -8 & 4 \\ 5 & 7 & 0 & 0 \\ -8 & 0 & 1 & -2 \\ 4 & 0 & -2 & 9 \end{bmatrix}$$

Find the

a) distribution of $Y_1 + Y_2 - 3Y_3 + Y_4$

b) joint distribution of $Y_1 - 2Y_4 + Y_3$ and $Y_1 - 3Y_2 + 2Y_3$

c) joint distribution of Y_1, Y_3 and Y_4

d) joint distribution of Y_3, Y_4 and $(Y_1 + Y_2 + Y_3)$

State the necessary result in each case.

$(2 + 4 + 2 + 4 = 12)$

2) Two samples are drawn from $N_2(\mu, \Sigma)$.

$$\text{Sample-1}: \begin{bmatrix} 3 & 17 \\ 7 & 12 \\ 8 & 15 \\ 6 & 24 \\ 10 & 18 \end{bmatrix} \quad \text{and} \quad \text{Sample-2}: \begin{bmatrix} 12 & 19 \\ 18 & 12 \\ 17 & 14 \\ 10 & 18 \end{bmatrix}$$

Test the hypothesis H_0: $\mu_1 = \mu_{2an}$ If required test for each variable separately

[12]

P.T.O

1

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination: 2011-2012

Course Name: M. Tech. (QR & OR) 2nd YEAR

Subject Name: Advanced Statistical Methods

Date of Examination: 05.09.11 Maximum Marks: 50 Duration: 2 hours

Note:
1. This paper carries 60 marks.
2. Answer all four questions but the maximum you can score is 50.
3. All notations have their usual meanings

1) Let $\mathbf{Y} \sim N_4(\mu, \sum)$, where

$$\mu = \begin{bmatrix} 4 \\ 6 \\ 9 \\ 3 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} 3 & 5 & -8 & 4 \\ 5 & 7 & 0 & 0 \\ -8 & 0 & 1 & -2 \\ 4 & 0 & -2 & 9 \end{bmatrix}$$

Find the

a) distribution of $Y_1 + Y_2 - 3Y_3 + Y_4$

b) joint distribution of $Y_1 - 2Y_4 + Y_3$ and $Y_1 - 3Y_2 + 2Y_3$

c) joint distribution of Y_1, Y_3 and Y_4

d) joint distribution of Y_3, Y_4 and $(Y_1 + Y_2 + Y_3)$

State the necessary result in each case.

$(2 + 4 + 2 + 4 = 12)$

2) Two samples are drawn from $N_2(\mu, \sum)$.

$$\text{Sample-1:} \begin{bmatrix} 3 & 17 \\ 7 & 12 \\ 8 & 15 \\ 6 & 24 \\ 10 & 18 \end{bmatrix} \quad \text{and} \quad \text{Sample-2:} \begin{bmatrix} 12 & 19 \\ 18 & 12 \\ 17 & 14 \\ 10 & 18 \end{bmatrix}$$

Test the hypothesis H_0: $\mu_1 = \mu_{2an}$ If required test for each variable separately

[12]

P.T.O.

3) A chemical engineer wants to compare two types of protective coatings (say A and B) for resistance to corrosion. 15 pipes each were coated using coating type A and B respectively. 15 pairs of pipes (one coated with type A and the other with type B) were formed and buried at 15 different locations and left for the same length of time.

Corrosion was measured as follows:

y_1: Maximum depth of pit in thousands of an inch for type A

y_2: Number of pits for type A

x_1: Maximum depth of pit in thousands of an inch for type B

x_2: Number of pits for type B

i) State and test the appropriate hypothesis.

ii) Carry out univariate tests

iii) What are your conclusions?

Maximum depth of Pits and Number of Pits of coated pipes

Location	Coating A		Coating B		Difference	
	Depth	Number	Depth	Number	Depth	Number
	y_1	y_2	x_1	x_2	$d_1 = y_1 - x_1$	$d_2 = y_2 - x_2$
1.	73	31	51	35	22	-4
2.	43	19	41	14	2	5
3.	47	22	43	19	4	3
4.	53	26	41	29	12	-3
5.	58	36	47	34	11	2
6.	47	30	32	26	15	4
7.	52	29	24	19	28	10
8.	38	36	43	37	-5	-1
9.	61	34	53	24	8	10
10.	56	33	52	27	4	6
11.	56	19	57	14	-1	5
12.	34	19	44	19	-10	0
13.	55	26	57	30	-2	-4
14.	65	15	40	7	25	8
15.	75	18	68	13	7	5

[10 + 4 + 2 = 16]

P

4) In a study involving 2 variables 10 sets of observations each were obtained from 4 different groups. The sample means for the three groups on each of the variables were 10 and 20; 20 and 25; 15 and 20. 20 and 30. The sample covariance matrices for each group were

$$S_1 = \begin{bmatrix} 6 & 1 \\ 1 & 5 \end{bmatrix}, \quad S_2 = \begin{bmatrix} 6 & 2 \\ 2 & 7 \end{bmatrix}, \quad S_3 = \begin{bmatrix} 8 & 3 \\ 3 & 9 \end{bmatrix}, \quad S_3 = \begin{bmatrix} 7 & 4 \\ 4 & 8 \end{bmatrix}$$

The experimenter wished to test the hypothesis H_0: $\mu_1 = \mu_2 = \mu_5$ about the means of these variables:

Answer the following questions:

(a)

1) What is the value of the Wilk's lambda?
2) What is the critical value of the significance test at the α level of .05?
3) What is the decision regarding the null hypothesis at the α level of .05?

(b)

1) What is the observed value of Roy's largest root test ?
2) What is the value of Hotelling - Lawley trace?
3) What is the observed value of the Pillai' test?

(15 + 5 = 20)

P.T.O

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination: 2011-12 (First Semester)

Course Name: M-TECH (QROR) II Year.

Subject Name: Software Engineering

Date: 6-9-2011 ***Maximum Marks: 50*** ***Duration: 2½ hours***

Note: *The question paper is for 60 marks, answer as many as you can, you can get at most 50 marks.*

1. What is the need for validating the requirements? Explain any one requirement validation technique. [6]

2. What are functional and non-functional requirements in software engineering? [6]

3. What is the difference between student program and industrial software? Explain your answer in details. [6]

4. What is the role of software architecture in a software system? How it is different from design? [6]

5. Differentiate between the following terms: (b) Process model and Process (c) A software product and a software process. [6]

6. Explain the importance of project staffing and different staff structures along with their advantages. [6]

7. Write any three process models and explain one of them in details with its advantages and disadvantages. [8]

8. Explain in brief three different software cost estimation methods. [8]

9. Find the Cost, Effort, Project duration, and People required using the intermediate semi-detached type COCOMO for the problem given below.

CAD Co., Inc. wants to produce a system that will perform computer aided design for a home construction industry. The system will have seven modules and their size in lines of code is given in the brackets: (1) UICF (2300), (2) 2DDGA (5300), (3) 3DGA (6800), (4) DBM (3350), (5) CGDF (4950), (6) PCF (2100) and (7) DAM (8400). The required model parameters are $a_i = 3.0$, $b_i = 1.12$, $c_i = 2.5$ and $d_i = 0.35$. The estimated effort multipliers for different cost drivers are as follows: data base size is high (1.08), product complexity is very high (1.30), main storage is very high (1.21), execution time constraints is high (1.11), programmer capability is very high (0.70), programming language experience is very low (1.07), modern programming practices is high (0.91) and all other related characteristics are normal. [8]

---XXX---

Note: Books, note-books, computers, mobiles etc. are not allowed during examination. Only calculators can be used.

INDIAN STATISTICAL INSTITUTE

Mid Semestral Examination: (2011 - 2012)

Course Name: M. Tech. (QROR) **Year: 2nd year**

Subject Name: Database Management Systems

Date: September 7, 2011 **Maximum Marks: 50** **Duration: 2 hrs**

Answer to all the questions.

1. Illustrate and elaborate the architecture of a database management system. 20

2. Describe the disadvantages of file processing systems over database management systems. 15

3. Write short notes on $8 + 6 + 5 = 19$

   i) database users

   ii) DDL and DML

   iii) transaction manager

INDIAN STATISTICAL INSTITUTE

M. Tech. (QR & OR) 2nd YEAR

Year: 2011

MIDSEMESTER EXAMINATION

Subject: Operations Research-II

Date of Exam: 08-09-2011 **Max. Marks: 100**

Time: 3 hours.

Group-A [50 Marks]

Answer any two from 1. to 3.

Let S be a nonempty open convex set in R^n and $f : S \to R$ be differentiable on S. Prove that f is convex if and only if

$$[\nabla f(x_2) - \nabla f(x_1)]^t (x_2 - x_1) \geq 0 \text{ for each } x_1, x_2 \in S.$$

State the definition of convex, quasiconvex and pseudoconvex functions.

[6 + 4 = 10]

Define epigraph and sub-gradient of a function.
Suppose A is an $m \times n$ matrix and c is an n vector. Then, exactly one of the following two systems has a solution:

System 1 $Ax \leq 0$, $c^t x > 0$ for some $x \in R^n$

System 2 $A^t y = c$ and $y \geq 0$ for some $y \in R^m$.

[5 + 5 = 10]

State the duality theorem.
Let S be a nonempty closed convex set in R^n and $y \notin S$. Then there exist a nonzero vector p and a scalar α such that $p^t y > \alpha$ and $p^t x < \alpha$ for each $x \in S$.

[5 + 5 = 10]

Assignment

[30]

Group B

Full Marks: 50

Begin this group on a fresh answer-script.

1. A machine operator has to perform three operations namely turning, threading and knurling on five different jobs, named A, B, C, D, E. The time required to perform these operations (in minutes) for each job is known and is given in the following table:

Job	A	B	C	D	E
Time for Turning	8	10	6	7	11
Time for Threading	5	6	2	3	4
Time for Knurling	4	9	8	6	5

Determine the order in which the jobs should be processed in order to minimize the total time required to finish all the jobs. What is the total elapsed time? What is the idle time at each of the three stages?

[6+5+9=20]

2. Let P_j be the problem: max $f(x)$, $x \in S_j$ $j = 1, 2$; $S_1 \supseteq S_2$. If x^0 is an optimal solution to P_1, then show that $f(x^0) \geq f(x)$ for all $x \in S_2$.

[5]

3. Consider the following LP:

$$\max x_0 = 4x_1 + x_2 + 3x_3 + 5x_4$$

Subject to

$$4x_1 - 6x_2 - 5x_3 + 4x_4 \geq -20$$
$$3x_1 - 2x_2 + 4x_3 + x_4 \leq 11$$
$$8x_1 - 3x_2 + 3x_3 + 2x_4 \leq 23$$
$$x_1, x_2, x_3, x_4 \geq 0$$

One of the simplex iteration tableau of the above LP is given below:

	y_0	$-x_1$	$-x_3$	$-x_6$	$-x_7$
x_0	A	17	-35	-17	11
x_5	B	E	11	0	2
x_4	C	F	-6	-3	2
x_2	D	G	-5	-2	1

Without performing the simplex iterations, find the missing entries $A, \cdots, I$. Show your calculations. [*Hint:* Write the problem in standard form.]

[15]

4. A salesman wants to visit five cities A, B, C, D and E. He does not want to visit any city twice before completing his tour of all the cities and wishes to return to the starting point of his journey. The cost of going from any city to another (in rupees) is shown below:

From↓ To →	A	B	C	D	E
A		2	5	7	1
B	6		3	8	2
C	8	7		4	7
D	12	4	6		5
E	1	3	2	8	

Find the least cost route.

[10]

Indian Statistical Institute

Mid-semester Examination: 2011-2012

M. Tech (QR& OR) II year
Subject: ReliabilityII

Maximum Marks: 100 **Duration: 3 hours** **Date:09.09.2011**

1. (a) Derive the realiability of the system when log(Stress)$\sim$Normal(μ, σ^2) and log(Strength)$\sim$Normal(λ, ρ^2). [20]

(b) The strength **S** and the stress **s** of a device are distributed lognormally with following parameters.
E(**S**) =100000KPa, Standard deviation(**S**) = 10000KPa, E(**s**) = 60000KPa, Standard deviation(**s**) = 20000KPa. Find the reliability of the device. [10]

(c) A component is to be designed for a specified reliability of 0.990. The stress and strength are known to be log-normally distributed with
E(**S**) =1100.00MPa, Standard deviation(**S**) = 100.00MPa, E(**s**) = 850MPa.
Determine the maximum allowable standard deviation of the stress that can be applied to the component which give us the desired reliability. [10]

2. Derive the expression for reliability of the system when
Stress $\sim$Normal(μ, σ^2), Strength $\sim$exp(λ) where λ is the failure rate of the system. [10]

3. (a) Write the effort minimisation algorithm in reliability allocation problem. [10]

(b) A system consists of 4 sub-systems that must function if the system has to function properly. The system reliability goal is 0.950. All the four sub-systems have identical reliability improvement effort functions. The estimated sub-system reliabilities at present time are 0.75, 0.85, 0.90, 0.95. What reliability goal should be apportioned to the sub-systems so as to minimize the total effort spent on the system improvement? [10]

(c) Solve the above problem by ARINC apportionment technique assuming the mission time for all the sub-systems and the system is 20 hours. [5]

(d) Compare the results in (b) and (c) and give your comments. [2]

4. (a) Explain the various facets of software quality and discuss why software reliability is the most important among them. [5]

(b) (i) Describe the assumptions of the Jelinski-Moranda model for software reliability. [4]

(ii) Develop the model from the assumptions and derive the expressions for the estimates of the unknown parameters assuming that the data obtainted are **n** times when the bugs are found. [6]

(iii) What method have you used for getting an optimal estimate? [1]

(iv) Note that the number of bugs initially present in the software is a discrete variable. Can you modify Jelinski-Moranda's approach in line with the above observation? [7]

Indian Statistical Institute

M.Tech (QR & OR), IInd Year, 2011-12

Mid-Semester Examination

Subject : Industrial Experimentation

Date:12.09.11 **Time: 3hours** **Maximum Marks : 100**

[Answer any part of any questions. The maximum marks you can score is 100]

1. (a) A manufacturer of an fuel additive (WOW3) asks you to recommend an experimental design to test the claim that the additive when mixed with petrol will give much better fuel efficiency. Available for this experiment, 50 cars which will be driven in their usual manners during the month of October. These cars are not of same make, model and age. All cars will be tuned at the beginning of the experiment. During the period of experiment, each time a car needs fuel, petrol with or, without additive will be used in accordance with your instructions. Each time the car gets fuel, data will be recorded on total kilometers driven and liters of fuel consumed. What experimental design will you recommend? Provide support for your recommendation? Should you require to make any assumptions, please state them.

1. (b) Construct a Fold-over design starting with the initial $\mathbf{2}^{5-2}$ design. What will be the resultant design and what will be its resolution? Give the defining relation of the resultant design.

[10+15=25]

2. (a) State the model of a Latin Square Design. It is decided to conduct an experiment with a replicated Latin Square design (n=2) in which the number of levels of both the row and column variables remain unchanged but their levels are changed. Give the complete Analysis of Variance table of the replicated Latin Square design.

2. (b) A $\mathbf{2}^{6-2}$ design is required to be constructed. Which of the following two sets of design generators will you choose and why?

(i) 5=1234; 6=123
(ii) 5=123 ; 6=234

Show that it is impossible to have a 2_V^{6-2} design.

[12+13=25]

3. (a) You are required to construct two experimental designs using the following sets of information. What will be the name(s) of the designs? Do the two designs have any special arrangement? How?

(i) Nos. of treatment (a): 4, Nos. of block (b): 6, Nos. of unit per block (k): 2, No. of times each treatments is replicated (r): 3.

(ii) Nos. of treatment (a): 7, Nos. of block (b): 7, Nos. of unit per block (k): 3, Nos. of times each pair of treatments is replicated (r): 3.

3. (b) One of the two shields (either I or, II) is set up on an experimental apparatus. The Sun shines through it and a response (either y_1 or, y_2) is measured on a certain piece of equipment. Each such measurement requires about 15 minutes. The equipment has three adjustable knobs (a,b,c) each at two levels. Given that you have 4 hours of suitable sunshine available on each of four successive days, how would you design an experiment to find the best settings of the three knobs if the object is to find that single setting of the knobs that will produce the largest difference in the measurements made with two shields.

[10+10=20]

4. (a) What is the name of the following design? Analyse the coded results obtained after carrying out an experiment with the design given in the following table.

.Blocks	Positions			
	1	2	3	4
1	A=2	B=9	C=0	D=14
2	B=6	A=5	E=5	C=3
3	C=1	D=9	A=0	E=7
4	D=8	E=8	B=10	A=4
5	E=7	C=6	D=11	B=10

4. (b) Construct a 3^{4-2} with the design generators $D=AB^2$ and $C=AB$.

[25+10=35]

5. (a) What is a Cross-over design? If there are two different methods of antibiotics, 6 ailing human subjects and the effect of the two types of antibiotic are required to be evaluated in two different time periods, what will be the design plan?

5. (b) Give the principal block and two other blocks of the 2^6 design confounded in 2^3 blocks with the following block generators: $B_1=123$, $B_2=2345$, $B_3=1346$.

[7+8=15]

6. There are many different ways to bake brownies. The purpose of this experiment was to determine how the pan material, brand of brownie mix and the stirring method affected the scrumptiousness of brownies. The factor levels were as given in the following table. The response variable was scrumptiousness, a subjective measure derived from the answers obtained on such issues as taste, appearance, consistency, aroma etc. with the help of a questionnaire given to the subjects who sampled each batch of brownies. A three person test panel sampled each batch and filled out the questionnaire.

Brownie batch	A=Pan material	B=Stirring method	C=Brand of Mix	Test Panel Results on scrumptiousness		
				1	2	3
1	Glass	Spoon	Expensive	11	9	10
2	Aluminum	Spoon	Expensive	15	10	16
3	Glass	Mixer	Expensive	09	12	11
4	Aluminum	Mixer	Expensive	16	17	15
5	Glass	Spoon	Cheap	10	11	15
6	Aluminum	Spoon	Cheap	12	13	14
7	Glass	Mixer	Cheap	10	12	13
8	Aluminum	Mixer	Cheap	15	12	15

Analyse the data and offer you comments on results of analysis. Do we really have 3 replicates of the response variable (scrumptiousness) for each treatment combination in the above table? When we take the measure of scrumptiousness as the response variable, do we have to compromise on any assumptions involved in the model of the design?

[20+5=25]

Indian Statistical Institute

M. TECH. (CS) 2 Year : 2011 2012
Mid-semester Examination
Subject: Cryptology

Date: 19/09/2011 Time: 2 hours Maximum Marks: 40

Note: The paper carries 45 marks. Maximum you can score is 40.

1. Prove that every Boolean function has a unique algebraic normal form (ANF) representation. [9]

2. Prove that $|\mathsf{Prob}(l_u = f) - \frac{1}{2}| = \frac{|W_f(u)|}{2^{n+1}}$, where symbols have their usual meaning. [9]

3. Define primitive polynomial. Prove that the sequence generated by LFSR of length d has period $2^d - 1$ if the connection polynomial is primitive. [9]

4. Describe Finney state. Prove that an RC4 state is in Finney state if and only if it's prior state is in Finney state. [9]

5. (a) Explain different kind of attack models on a cryotosystem.
   (b) Define one-time pad. What is its deficiency. How can you overcome the deficiency? [4 + 5 = 9]

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination : (2011-2012)

M.Tech. (Computer Science) II Year.

Database Management Systems

Date : 22.09.2011 **Maximum Marks : 40** **Time : 1.5 Hours.**

1. A flight database in India maintains the detail of flights and their destinations. Each flight detail is recorded with its unique flight-no and the name of the airlines. A flight is either a domestic or an international one. It can also be either a hopping or a direct flight. A hopping flight touches more than one place between the originating point and the final destination. For each domestic destination (final or intermediate), the unique destination name and the state name are stored. In case of an international flight, each international destination name (unique) is associated with the corresponding country name. An international flight may touch more than one domestic airport before flying out of the country. For each flight the arrival and departure time at each destination is stored. The arrival time of the originating point of a flight and the departure time of the final destination of a flight are stored with a value 9999.

   a) From the above description draw an appropriate ER/EER diagram.
   b) To design a relational schema, derive a set of relations from the ER/EER diagram using the standard mapping rules.
   c) Using the relational schema form the following queries:
      i) Find all those hopping international flights with their flight nos. and name of the airlines that touch any domestic airport other than the originating one. [USE RELATIONAL ALGEBRA]
      ii) Find all those flights with their flight nos. and name of the airlines that originate from "Kolkata" and touch "Raipur" either as an intermediate point or final destination. [USE ANY RELATIONAL CALCULUS]

(4x7=28)

2. A relation R(A,B,C,D,E) should satisfy the following set of functional dependencies:

$\{A \rightarrow B, B \rightarrow C, C \rightarrow D, E \rightarrow D\}$

   a) Find the candidate keys of R.
   b) Check whether R is in 2nd Normal Form.
   c) If R is not in 2nd Normal Form, check whether R can be decomposed to create relations free from partial dependencies preserving the given set of functional dependencies.

(5+3+4=12)

※

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination: 2011-2012

M. Tech. (CS) 2nd Year

Artificial Intelligence

Date: 24.09.2011 Maximum Marks: 60 Duration: 2 hours

Answer all questions in brief.

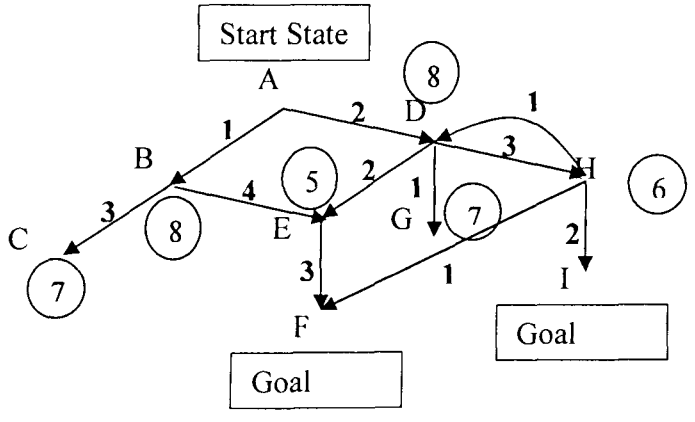
1. Consider a sliding block puzzle with the following initial configuration:

W	W	W	B	B	B	E

There are three white tiles (W), three black tiles (B), and an empty cell (E). The puzzle has the following moves:
(i) A tile may move into an adjacent empty cell with unit cost.
(ii) A tile may hop over at most two other tiles into an empty cell with a cost equal to the number of tiles hopped over.
The goal of the puzzle is to have all of the black tiles to the left of all of the white tiles without regard for the position of the empty cell. Define the problem as a state space graph problem and find a sequence of moves that will transform the initial configuration to a goal configuration. What is the cost of the solution? [4 + 6 = 10]

2. What do you mean by *ridge* and *plateau*? What is the difference between *simulated annealing* and *steepest ascent hill climbing* approach? Describe the crossover and mutation operators of genetic algorithm. [1 + 1 + 2 + 6 = 10]

3. Execute the *uniform cost search* and *best first search* algorithms on the following search graph, and show the solution path, along with its cost and list the expanded nodes for each case (each node of the graph is represented by a letter and the encircled value is the heuristic evaluation of the corresponding node, while the bolded numerical value represents the actual length of the path between two nodes). [5 + 5 = 10]



BTO

4. Define monotone property of a heuristic. Prove that any monotonic heuristic is admissible. Prove that the set of states expanded by algorithm A^* is a subset of those examined by breadth first search. [5 + 5 + 5 = 15]

5. The game of NIM is played as follows: Two players alternate in removing one, two or three coins from a stack initially containing five coins. The player who picks up the last coin loses.
   a) Draw the full game tree.
   b) Show that the player who has the second move can always win the game.
   c) Execute α-β pruning procedure on the game tree. How many terminal nodes are examined? For each cutoff, specify whether it is α-cutoff or β-cutoff. [4 + 3 + 8 = 15]

INDIAN STATISTICAL INSTITUTE
M. Tech (Computer Science) II year, 2011 – 12
Pattern Recognition and Image Processing
Mid-Semestral Examination

Date: 27·09·11 Maximum Marks: 100 Time: 195 minutes

Note: Answer all the questions.

1. State the Bayes decision rule for a 3-class classification problem. Show that it provides the minimum misclassification probability. [3+12=15]

2. Let p_1 and p_2 denote the conditional probability density functions for classes 1 and 2 respectively and, 0.4 and 0.6 be their corresponding prior probabilities. Let

$$p_1(x) = x\ ; 0 < x < 1$$
$$= (2 - x); 1 \le x < 2$$
$$= 0 \ otherwise$$

$$p_2(x) = 1 + \frac{x - 2.5}{1.5}; \quad 1 < x < 2.5$$

and

$$= 2(3 - x); \quad 2.5 \le x < 3$$
$$= 0 \quad otherwise.$$

(a) Find the Bayes decision rule for the above classification problem and find its error probability.
(b) Find the error probability for the following decision rule.
Put x in class1 if $x \le 2$. Otherwise, put it in class 2.
[12+6=18]

3. (i) Describe k-means clustering algorithm.
(ii) Describe the minimum within cluster distance criterion for clustering and state the properties of the clusters obtained by this criterion.
(iii) Provide a dataset and two initial sets of seed points so that the k-means algorithm results in two different clusterings.
(iii) Describe a divisive clustering scheme.
(iv) Describe any two feature selection algorithms.
[6+6+5+6+10=33]

4. Draw the solution tree for branch and bound feature selection algorithm where 3 features are to be chosen from 7 features. [8]

5. Describe perceptron learning algorithm. [10]

(P.T.O)

6. Let $X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix}$ be a random vector with dispersion matrix

$\sum$ where $\sum = \begin{pmatrix} 2 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 \\ 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 5 \end{pmatrix}$. Find the two principal components of X. [10]

7. Write a short note on probabilistic separability measures for feature selection. [6]

Indian Statistical Institute

Semester-I 2011-2012

M.Tech.(CS) - Second Year

Mid-semester Examination (29 September, 2011)

Subject: Compiler Construction

Maximum marks: 40 Duration: 3 hrs.

Please keep your answers brief and to the point.

(a) In the buffering scheme commonly used in lexical analyzers, the input buffer consists of two halves, each of which can hold 1 disk block. Explain why this scheme is used instead of a simple buffer that can hold 1 disk block.

(b) Consider a file that contains a sequence of detailed *search queries*. Each search query consists of a *title*, a *description* and a *narrative* section. These sections are demarcated using opening and closing tags (<title>, </title>, etc.). An example query is given below:

```
<title>corruption </title>
<desc> Anna  Hazare's role in fighting corruption</desc>
<narr>I am looking  for  information  about Anna Hazare's campaign
against corruption.
</narr>
```

Write a Lex program that will take such a file as input and output the queries in a proper format that follows the rules given below.

- The queries are to be consecutively numbered starting from 1. The query number should be enclosed in <num> and </num> tags.
- There should be no white space between a tag and the adjacent text. Thus, the space between corruption and </title> in the above example should be removed in the output.
- Any non-empty sequence of spaces and tabs should be replaced by a single space.
- Each search query should occupy a single line. Thus, the narrative section in the above example should be reformatted to fit in a single line in the output. There should also be a single blank line between two queries.

The expected output corresponding to the above example is as follows.

```
<num>1</num>
<title>corruption</title>
<desc>Anna Hazare's role in fighting corruption</desc>
<narr>I am looking for information about Anna Hazare's campaign against corruption </narr>
```

Hint: In Lex, . is a regular expression that matches any single character other than the newline.

Note: You may make reasonable assumptions where necessary. Clearly state any assumptions you make.

Warning: Lex always tries to find the longest possible match for any given pattern.

3 · 12 · 15

P.T.O.

2. Eliminate left recursion from the following grammar.

$$\begin{array}{lll} A \to Aa & A \to Bb & A \to c \\ B \to Bb & B \to Cc & B \to a \\ C \to Cc & C \to Aa & C \to b \end{array}$$

(A, B, C are non-terminals, while a, b, c are terminal symbols.) [10]

3. Consider the following syntax-directed definition (SDD). Capital letters denote non-terminals, lower case letters denote terminal symbols.

$A \to B_1 C B_2$	$\{\ \boldsymbol{A}.a = C.c\ ;\ C.d = B_1.b + B_2.b\ \}$
$B \to a$	$\{\ \boldsymbol{B}.b = 1\ \}$
$B \to b$	$\{\ \boldsymbol{B}.b = -1\ \}$
$C \to (P)$	$\{\ \boldsymbol{C}.c = C.d * P.p\ \}$
$C \to P_1 \diamond P_2$	$\{\ \boldsymbol{C}.c = C.d * (P_1.p + P_2.p)\ \}$
$P \to xy$	$\{\ \boldsymbol{P}.p = x.val + y.val\ \}$
$P \to yx$	$\{\ \boldsymbol{P}.p = y.val - x.val\ \}$

(a) Compute the *FIRST* set for the right hand side of each syntax rule of the above SDD.

(b) Compute the *FOLLOW* set for each non-terminal. Show your rough work.

(c) Construct the $LL(1)$ parsing table for the above grammar. Is the grammar $LL(1)$? Justify your answer in 1-2 lines.

(d) Construct the canonical collection of $LR(1)$ items for the grammar using the canonical LR method.

(e) Using the above grammar, draw the parse tree for the input string:

$$bxy \diamond yxa$$

(f) For this parse tree, draw the dependency graph between attributes.

(g) Show the steps by which $A.a$ would be calculated for the above parse tree.

[2+4+4+7+2+3+3=25]

INDIAN STATISTICAL INSTITUTE

Midterm Examination: (2011)
M Tech (CS) 2nd year
Internet and Multimedia

Date: 30/09/2011 Maximum Marks 50 Duration: 2 hours

Answer ***all*** questions. The answers should be presented point-wise and **not** in descriptive style. Clearly specify the input and output in case of an algorithm.

1. Write TRUE or FALSE (10):

a) Maximum three vanishing point directions can be estimated in case of orthographic projection.
b) Time domain features of an audio signal can be represented in harmonic spectra.
c) RLE can be used to compress 2D color images.
d) P-frame in MPEG video is always between two B-frames.
e) SNR can be measured in dB units.
f) Gaussian filtering sharpens the signal contrast.
g) CMYK colour representation is used for television broadcasting.
h) Pitch is the perceptual equivalence of the frequency of audio signal.
i) Pixels can have 6-neighbourhood connectivity in a 2D digital image.
j) The 2D DCT is a non-linear transform.

2. Suppose the alphabet is [A, B, C] and the known probability distribution is P=0.5, P=0.4, P=0.1. For simplicity let us also assume that both encoder and decoder know that the length of the message is always 3 alphabets, so there is no need for a terminator. (a) How many bits are needed to encode the message BBB by Huffman encoder? (b) How many bits are needed to encode the message BBB by arithmetic encoder? (5+5=10)

3. Assume 1, 2 and 3 are codes for string A, B and C. State the output of LZW dictionary based compression of an input string ABABBABCABABBA. Also, show that the above input string can be perfectly reconstructed from the compressed output code. (7+3=10)

4. Given coordinates of two points (1,3) and (7,5) find out all the intermediate points lying on the digital line displayed in a computer screen connecting points (1,3) and (7,5). Clearly state the steps. [10]

5. Short questions (5x2=10):
   a. What is *p*64 coding scheme?
   b. What is the advantage of wavelet transformation?
   c. State four time domain features of audio signal?
   d. Define 2D affine and projective transform.
   e. Define Compression Ratio.

12/9/11

Indian Statistical Institute

M.Tech (QR & OR), II year

First Semestral Examination, 2011-12

Subject: Industrial Experimentation

Date: 21.11.11 Time: 3 Hours Maximum Marks:100

Answer any part of any question. Usual notations have been used. The maximum marks you can score is 100

1.

(a) Does rotatability of any response surface design ensure that the parameter $\frac{N \operatorname{Var} \hat{y}(x)}{\sigma^2}$ has stable distribution throughout the explored region of experimentation and is only a function of distance from the design origin? Does the stable distribution of the parameter $\frac{N \operatorname{Var} \hat{y}(x)}{\sigma^2}$ mean that it has uniform value throughout the region of experimentation? If not, what is the proper interpretation of the word "stable distribution"?

(b) Show that a simplex design with k=3 is at the same time rotatable and variance optimal.

[7+8=15]

2.

(a) Is a second order CCD matrix without any centre point singular when $\alpha = \sqrt{k}$?

(b) Choose a value of α in a second order CCD to make the design rotatable with the controllable factors A and B having I=AB for factorial portion of the CCD.

(c) Using the appropriate design moments, prove that the condition of rotatability of a second order response surface CCD is given by $\alpha = \sqrt[4]{n_f}$ where, n_f is the number of point in the factorial portion of the design.

[4+4+7=15]

3.

(a) Construct a Box Behnken second order response surface design with k=4. Can you prove using the design moments that the design is orthogonal and at the same time rotatable?

(b) Consider the following design with k=2.

A	B
1	0
0.5	$\sqrt{(0.75)}$
-0.5	$\sqrt{(0.75)}$
-1	0
-0.5	$-\sqrt{(0.75)}$
0.5	$-\sqrt{(0.75)}$
0	0
0	0
0	0

What is the name of the design? Is the design Spherical? Does the design have the properties of orthogonality and rotatability?

(c) Is rotatability of a second order response surface design always preferable for ensuring quality of prediction? Give reasons.

[10+12+3=25]

4.

(a) What is the canonical form of a second order response surface model when mixture components are the controllable variables?

(b) The following parameter estimates were calculated from multiple observations collected at the points of {3,2} simplex lattice design for a mixture experiment.

b_1=5.5, b_2=7.0, b_3=8.0, $\frac{1}{4}b_{12}$=8.75, $\frac{1}{4}b_{13}$=2.25, $\frac{1}{4}b_{23}$=4.5.

(i) What were the values of the mean responses observed at the lattice points?

(ii) Suppose that, three replicated response values were collected at each binary blends ($x_i = x_j = 0.5$; x_k=0) and four replicated response values were observed at each vertex point of the simplex lattice design. Which of the quadratic coefficients are significantly greater than 0 (at 5% level of significance) if an estimate of the variance of each observation is s^2=3.0?

[3+12=15]

5. Consider a 3-component mixture experiment with the following constraints on the proportions of the mixture components.

$0.34 \leq x_1$, $0.42 \leq x_2$, $0.10 \leq x_3$; $x_1 + x_3 + x_3 = 1$

Find an appropriate design for fitting a second order response surface model for mixture components.

[15]

6.

(a) Given the centre of the design in original mixture components (x_1, x_2, x_3) as $x_0 = (0.39, 0.14, 0.47)'$ and the matrix= diag (0.25,0.14,0.30) associated with an ellipsoidal region defined in terms of the three mixture components, construct a first order rotatable design using four points only and place the points on the boundary of the unit sphere centered at $x_0 = (0.39, 0.14, 0.47)'$. List the component proportions corresponding to the design settings.

(b) If you want to choose the design points symmetrically positioned on (i) the largest sphere and (ii) the largest cube, both centered at x_0 that can be inscribed inside the simplex, what should be the values of the radius multiplier?

(c) If you want to ensure minimum bias of the coefficients of the fitted model, what should be your choice for the radius multiplier?

[30+5+5=40]

7. The fitted model in terms of w_1, w_2 (MRVs) and z_1 using response values observed in an experiment was as follows:

$$\hat{y}(w, z_1) = 19.33 - 1.77w_1 - 0.21w_2 + 4.65z_1 - 0.04w_1z_1 - 1.03w_2z_1$$

Give the full design matrix used for estimating the coefficients of the fitted model and also interpret the various coefficients shown in the above model.

[10]

8.

(a) What are the two step procedures involved in Taguchi's parameter design approach? What are the assumptions involved in this approach? What is the use of Taguchi's Signal--to–Noise ratio in this approach? Is the Signal- to –Noise ratio related to Taguchi's Quadratic loss function? As per Taguchi's approach, do you gainfully exploit control by noise interaction to reduce the variability of response characteristic?

(b) Suppose that the true quality loss is highly skewed. For instance, a smaller windshield has got to be scrapped where-as, a larger windshield can be reworked. The cost of rework is much smaller than the cost of scrap. In this situation, is it still right to bring the location on target as indicated in the two-step procedure? What should be the best approach here? Do you find any conflict in following Taguchi's two step procedure in this case? Explain.

[7+8=15]

INDIAN STATISTICAL INSTITUTE

Semestral Examination: (2011 - 2012)

Course Name: M. Tech. (QROR) **Year: 2nd year**

Subject Name: Database Management Systems

Date: November 23, 2011 **Marks: 100** **Duration: 3 hrs**

Answer to all the questions.

1. A COMPANY database keeps track of a company's employees, departments and projects.

   i) The company is organized into departments. Each department has a unique name and a unique number, and includes a particular employee who manages the department. The database stores the start date when the employee began managing the department. Department may be distributed in several locations.

   ii) Each department controls a number of projects, each of which has a unique name, a unique account number, and a single location.

   iii) Each emploee has a name, PAN (Permanent Account Number), address, basic pay, gender and birthdate. An employee is assigned to one department but may work on several projects that are not necessarily controlled by the same department. The database keeps track of the number of hours per week, which an employee works on each project. It also keeps track of the direct supervisor (manager) of each employee.

   iv) The databse records the details of the dependents of each employee for insurance purpose. Each dependent's name, gender, birthdate and relationship to the employee are also recorded.

   a) Draw an ER diagram for the COMPANY database. 20

   b) Write an expression, in relational algebra, that retrieves the name of each emploee who works on all the projects controlled by the department named "Quality Control". Write also the corresponding SQL statements for the query. 10

P.T.O

2. a) Elaborate in details what a safe tuple relational calculus expression means. 10

b) Consider a binary relation between two schemas: customer (cust_ID, cust_name) and account (cust_ID, account_no, account_balance). The names of the schemas and the attributes are self-explanatory. Write the safe expressions that result in customer names, and their account numbers and corresponding account balance of all the customers, in

i) tuple relational calculus (in reduced form), and

ii) domain relational calculus (in reduced form). 5+5 = 10

3. a) Consider a relational schema EMP_PROJ = (emp_id, proj_id, hours, emp_name, proj_name, proj_location), and a set of functional dependencies as F = {(emp_id, proj_id) → hours, emp_id → emp_name, proj_id → (proj_name, proj_location)}. Check, with proper reason, whether the schema is in second normal form. If not, decompose, by inspection, it into new schemas so that each of them is in second normal form. 2

+ 3 = 5

b) List all the functional dependencies satisfied by the following relation with attributes *A*, *B* and *C*. 5

A	B	C
a_1	b_1	c_1
a_1	b_1	c_2
a_2	b_1	c_1
a_2	b_1	c_3

c) Consider a relational schema $R = (A, B, C, D, E)$ with the set F of functional dependencies, $F = \{A \rightarrow BC, CD \rightarrow E, B \rightarrow D, E \rightarrow A\}$.

(i) List the candidate keys for R. 5

(ii) Compute B^+. 5

(iii) Compute canonical cover F_c. 5

(iv) Check whether R is in BCNF with respect to F. If not, decompose R into two or more relational schemas so that the resulting schemas are in BCNF. 5

4. Consider a relational schema R. Let α, β, γ and δ be subsets of R. The symbols "$\rightarrow$", "$\rightarrow\!\!\rightarrow$" and "$\vdash$", used in the following, stand for "functional dependency", "multivalued dependency" and "logical implication" respectively. Prove or disprove the following:

   a) $\alpha \rightarrow \beta$ and $\gamma \rightarrow \beta \vdash \alpha \rightarrow \gamma$

   b) $\alpha \rightarrow\!\!\rightarrow \beta$ and γ is a superset of $\delta \vdash \alpha\gamma \rightarrow\!\!\rightarrow \beta\delta$

   c) $\alpha \rightarrow\!\!\rightarrow \beta$ and $\beta \rightarrow\!\!\rightarrow \gamma \vdash \alpha \rightarrow\!\!\rightarrow (\gamma - \beta)$

   d) $\alpha \rightarrow \beta \vdash \alpha \rightarrow\!\!\rightarrow \beta$ $5 + 5 + 5 + 5 = 20$

Indian Statistical Institute

Semester I Examination: 2011-2012

M. Tech (QR& OR) II year

Subject: ReliabilityII

Maximum Marks: 100 Duration: 3 hours Date:25.11.2011

Notations used in this paper are usual notations used in the class.
Answer all questions. Marks alloted for each question are given in [ ].

1. (a) Describe the Cox's proportional hazard model. Discuss why it is called a proportional hazard model? What is the significance of it being called a semiparametric model?

(b) Suppose there are survival times of patients who have just undergone brain surgery at one of the 3 hospitals X,Y and Z. The following data are recorded for each patient:

$$z_1 = \begin{cases} 1 \text{ for male,} \\ 0 \text{ for female} \end{cases}$$

$$z_2 = \begin{cases} 1 \text{ if patient attends hospital Y,} \\ 0 \text{ otherwise} \end{cases}$$

$$z_3 = \begin{cases} 1 \text{ if patient attends hospital X,} \\ 0 \text{ otherwise} \end{cases}$$

We have got $\hat{\beta}_1 = 0.071$, $\hat{\beta}_2 = -0.021, \hat{\beta}_3 = 0.016$, using Cox's proportional hazard model. Compare the force of mortality for a female patient of hospital Z with that of
i) male patient of hospital X
ii) female patient of hospital Y.

(c) Explain Breslow's approximation to partial likelihood method of estimation with reference to the above model.

(d) An investigation was carried out into the survival times measured in months of patients in hospital following liver transplants. The covariates are

$$z_{1i} = \begin{cases} 0 \text{ for Placebo} \\ 1 \text{ for treatment X} \end{cases}$$

and z_{2i} =weight of patient (measured in kg), i represents the i^{th} patient.
The observed life times with weights in brackets are as follows:-

Placebo	Treatment X
3(83)	6*(58)
9(68)	11(73)
14(75)	14(68)
16(86)	14*(49)

Observations with * represents censored.

Using Breslow's approximation, find out what contribution to the partial likelihood is made by the deaths at time 14.

[5+6+2+7=20]

2. (a) Discuss the warranty problem faced by a manufacturer.

(b) Point out the differences between guarantee and warranty.

(c) What is meant by Renewing Pro-rata warranty?

(d) Explain what is 2D Combination Free Replacement- Pro Rata Warranty?

(e) Write the differences between preventive and corrective maintenance.

[2+1+2+2+3=10]

3. Prove that a coherent system of independent IFR components has an IFRA life distribution.

[20]

4. (a) Discuss the steps involved in accelerated life testing experiments.

(b) A manufacturer of Bourdon tubes (used as a part of pressure sensors in avionics) wishes to determine its MTTF. The manufacturer defines the failure as a leak in the tube. The tubes are manufactured from 18 Ni (250) maraging steel and operate with dry 99.9% nitrogen or hydraulic fluid as the internal working agent. Tubes fail as a result of hydrogen embrittlement arising from the pitting corrosion attack. Because of the criticality of these tubes, the manufacturer decides to conduct ALT by subjecting them to different levels of pressures and determining the time for a leak to occur. The units are continuously examined using an ultrasound method for detecting leaks, indicating failure of the tube. Units are subjected to 3 stress levels of gas pressures(in psi and the times(in hours) for the tubes to show leaks are given in the following table:

100 psi	120 psi	140 psi
1557	1378	215
5725	2092	431
6207	2656	451
6767	3362	496
7146	3393	565
7346	3477	651
7826	4101	708
8095	4545	743
8871	5030	865
9989	5355	927
11458	5760	966
12102	5968	1124
12512	6783	1158
13536	7329	1293
14997	8440	1385

Determine the mean lives for design pressures of 80 and 90 psi.

[3+12=15]

5. Define likelihood ratio ordering of probability distributions.

Assume a discrete framework for a software release time model. The cumulative maximum gain after k trials in software testing is given by $g_k(\pi(h,t))$ with a prior distribution of number of unidentified bugs as $\pi(h,t)$. Assume that $lim_{k\to\infty} g_k(\pi(h,t)) = g(\pi(h,t))$. Then show that if $\pi'(h_1,t_1) \overset{LR}{>} \pi(h_2,t_2)$ then $g(\pi'(h_1,t_1)) \geq g(\pi(h_2,t_2))$.

[2+23=25]

6. Fill in the following blanks from the given choices in brackets:

(a) The largest class of life distributions among the following is $-------$ *(IFR, IFRA, DMRL, NBU, NBUE).*

(b) If $\pi'(h_1,t_1) \overset{LR}{>} \pi(h_2,t_2)$ then the ratio $\frac{\pi'_n(h_1,t_1)}{\pi_n(h_2,t_2)}$ is $-------$ *(increasing, decreasing)* in n.

(c) If F is *IFR* then $\bar{F}(x|t)$ is $-------$ (increasing, decreasing) in t, $x \geq 0$.

(d) If $\pi'(h_1,t_1) \overset{LR}{>} \pi(h_2,t_2)$ then $\alpha(\pi'(h_1,t_1))$ $-------(\geq,\leq)\alpha(\pi(h_2,t_2))$.

(e) If $\pi'(h_1,t_1) = (1,0,0,\ldots)$ and $\pi(h_2,t_2) = (0,1,0,\ldots)$ then $g(\pi'(h_1,t_1))$ $-------(\geq,\leq)g(\pi(h_2,t_2))$.

[2×5=10]

Indian Statistical Institute

M. TECH. (CS) 2 Year : 2011–2012
Semester Examination
Subject: Cryptology

Date: 28/11/2011 Time: 3 hours Maximum Marks:100

Note: The paper carries 110 marks. Maximum you can score is 100.
Notations used are as explained in the class.

1. Define perfect secrecy. Suppose a cryptosystem achieves perfect secrecy for a particular plaintext probability distribution. Prove that perfect secrecy is maintained for any plaintext probability distribution. [2 + 8 = 10]

2. Suppose a sequence of plaintext blocks, $x_1, \cdots, x_n$ yields the ciphertext sequence $y_1, \cdots, y_n$. Suppose that one ciphertext block, say y_i, is transmitted incorrectly. Show that number of blocks that will be decrypted incorrectly is equal to one if ECB or OFB modes are used for encryption; and equal to two if CBC or CFB modes are used. [10]

3. Describe iterated block cipher. Give a very short, high level description of AES. What is whitening? [6 + 6 + 3 = 15]

4. Give the definition of random oracle model. Prove that under some reasonable assumptions (which you have to state), a collision resistant hash function is preimage resistant. [5 + 10 = 15]

5. Define the RSA cryptosystem. Prove that in RSA $d(e(x)) = x$, for any $x \in \mathbb{Z}_n$. Prove that RSA is not IND-CPA secure. [4 + 8 + 3 = 15]

6. Explain Shamir's (t, w)-threshold scheme. Prove that it is a perfect secret sharing scheme. [3 + 7]

7. Describe the ElGamal signature scheme. Describe the various attack models and the goal of the adversary on a signature scheme. Prove that there is an existential forgery under key only attack on ElGamal signature scheme. [5 + 5 + 5 = 15]

8. Let E be an elliptic curve $y^2 = x^3 + x + 6$ over $\mathbb{Z}_{11}$.

   (a) Determine the points on $E(\mathbb{Z}_{11})$.

   (b) Compute $(2, 7) + (2, 7)$ where $+$ is the addition operation defined on the elliptic curve.

   (c) Compute $(5, 2) + (2, 7)$.

   (d) Compute $-(2, 7)$.

   (e) Compute $(2, 7) + \mathcal{O}$ (where $\mathcal{O}$ is the point at infinity). [8 + 5 + 5 + 1 + 1 = 20]

INDIAN STATISTICAL INSTITUTE
M.Tech (QR & OR) 2nd YEAR
Academic Session: 2011-12
SEMESTER EXAMINATION

Subject: Operations Research-II

Date of Exam: 29-11-2011 **Time: 10:30-13:30** **Maximum marks: 100**

Use separate answer scripts for Group-A and Group-B.

Group – A [50 marks]

Answer any four from (1) to (5).

1. a) Suppose that $f : R^n \to R$ is differentiable at $\bar{x}$. Prove that if there is a vector d such that $\nabla f(\bar{x}) < 0$, then there exists a $\delta > 0$ such that $f(\bar{x} + \lambda d) < f(\bar{x})$ for each $\lambda \in (0, \delta)$.

   b) Suppose that $f : R^n \to R$ is differentiable at $\bar{x}$. Prove that if $\bar{x}$ is a local minimum then $\nabla f(\bar{x}) = 0$ and $H(\bar{x})$ is positive semidefinite.

[4 + 6 = 10]

2. a) Characterize the stationary point of a nonlinear programming problem in connection with the cone of feasible direction and the cone of descent direction.

   b) State the difference critically between Fritz- John necessary condition and KKT necessary condition of optimality.

[4+6=10]

P.T.O

3. a) Formulate a quadratic programming problem as a linear complementarity problem LCP (q, M). State the same formulation in case of a linear programming problem.

b) Solve the following linear complementarity problem, LCP (q, M) by using complementary pivoting algorithm.

$$M = \begin{bmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ 1 & 1 & 2 & 0 \\ 1 & 1 & 0 & 2 \end{bmatrix} \qquad q = \begin{bmatrix} 3 \\ 5 \\ -9 \\ -5 \end{bmatrix}$$

[5+5=10]

4. a) Define positive semidefinite matrix and co-positive matrix.

Consider the matrix $A = \begin{bmatrix} 0 & -1 & 2 & 5 \\ 2 & 1 & 5 & 2 \\ 1 & 0 & 0 & -1 \\ 2 & 4 & 2 & 1 \end{bmatrix}$,

Is the matrix A co-positive-plus?

b) Formulate a linear fractional programming problem as a linear programming problem.

[5+5=10]

5. a) Suggest an additive type iterative model to solve dynamic programming problem.

b) Define complementary matrix and principal pivot transform.

[4+6=10]

6. Assignment

[10]

Group B

Note: Begin this group on a fresh answer-script.
Questions have been set for 55 marks. You may answer as much as you can; but the maximum you can score in this group is 50

1. Solve the following ILP by the cutting plane method:

Maximize $x_0 = x_2$
subject to

$$3x_1 + 2x_2 \leq 6$$
$$-3x_1 + 2x_2 \leq 0$$
$$x_1,\ x_2 \geq 0, \text{integer}$$

[12]

2. Are the following matrices totally unimodular? Give reasons for your answer.

a) $A = \begin{bmatrix} 1 & 0 & -1 \\ 1 & -1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$

b) $B = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 1 & 0 & -1 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 1 \end{bmatrix}$

[3+5=8]

3. A company has an old generator which has to be removed and then replaced by a newer model. This procedure has been performed many times and a careful collection of data has been maintained. The removal time of the generator is as follows:

	Generator Removal					
Time (in hours)	1	2	3	4	5	6
Frequency	3	18	11	9	7	2

The engineers estimate that the times required for fitting the newer model, not used earlier, are as follows:

i) most optimistic time = 2 hours
ii) most likely time = 6 hours
iii) most pessimistic time = 13 hours

Find a time estimate (to the nearest hour) for completion of both the jobs, using 95% confidence level.

[10]

4. a) State clearly the Johnson's optimal sequencing decision rule for n jobs on 2 machines.

   b) The fundamental cut in method of integer forms is given by

$$s = -f_{i0} + \sum_{j \in R} f_{ij}\, x_j\,, \qquad s \geq 0$$

   Show that s must be an integer for an ILP problem.

[5+5=10]

5. Consider a project for which the relevant details are given below:

Activity (i,j)	Normal		Crash	
	Duration	Cost	Duration	Cost
(1,2)	15	600	12	1200
(1,3)	8	700	5	1600
(2,5)	12	750	6	1500
(3,4)	15	650	12	1400
(3,5)	18	700	13	1450
(4,5)	8	500	5	950

   a) How long will it take to complete the project?

   b) What will be its associated cost?

   c) If the management wants the project to be completed in 28 days through crashing, what would be the cost incurred? Show all relevant calculations.

[5+2+8-=15]

INDIAN STATISTICAL INSTITUTE

First-Semester Examination: 2011-2012

M. Tech. (CS) 2nd Year

Artificial Intelligence

Date: 30.11.2011 Maximum Marks: 100 Duration: 3 hours

Answer all questions in brief.

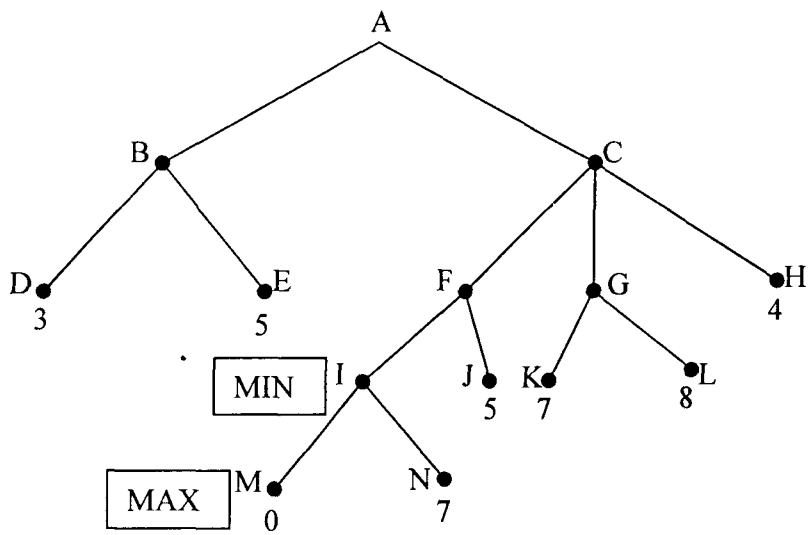
1. a) Prove that if a well formed formula α is a tableau provable, then α is valid.
   b) Consider the following set of sentences: "Mary will get her degree only if she registers as a student and passes her examination. She has registered herself as a student. She has passed her examination." Prove using both semantic tableaux approach and resolution refutation method that "she will get a degree". [5 + (5 + 5) = 15]

2. Answer *any two* from the following: [2 X 10 = 20]
   a) Define the constraint satisfaction problem and solve the following cryptarithmetic problem:

```
    S E N D
  + M O R E
  ------------
  M O N E Y
  ------------
```

[2 + 8 = 10]

   b) Perform the *minimax* search procedure on the game tree shown below in which static scores are all from the first player's point of view and MAX is allowed to move first. Perform the left-to-right and right-to-left α-β pruning procedure on this tree and show how many nodes can be pruned. Discuss why a different pruning occurs. [3 + 3 + 3 + 1 = 10]



   c) In *farmer-fox-goose-grain* puzzle, a farmer wishes to cross a river taking his fox, goose, and grain with him. He can use a boat which will accommodate only the farmer and one of his possessions. If the fox is left alone with the goose, the goose will be eaten. If the goose is left alone with the grain, the grain will be eaten. Draw a state space search tree for this

P.T.O

puzzle using left-bank and right-bank to denote left and right river banks, respectively. [10]

3. a) Explain with example the differences between red cut and green cut in Prolog.
   b) How do you convert a clause of first order predicate logic into clausal representation of logic programming?
   c) Write a program in Prolog for merging two ordered lists. [5 + 4 + 6 = 15]

4. a) Prove that a clause C is a logical consequence of a set of clauses S if and only if the set $S' = S \cup \{\sim C\}$ is unsatisfiable.
   b) Prove using semantic tableaux that the following sentences are mutually inconsistent: "All musicians are singers. A teacher is not a singer. Mary is a teacher. Mary is a musician."
   c) What are the main features of an expert system? Explain with an example the forward chaining approach in a rule based expert system. [6 + 6 + (4 + 4) = 20]

5. Answer *any three* from the following: [3 X 10 = 30]
   a) Let $I = <U, A>$ be a decision table, where $U = \{x1, ..., x7\}$ is a nonempty set of finite objects, the universe, and $A = C \cup D$ is a nonempty finite set of attributes. Here, $C = \{A1, A2\}$ and D are the set of condition and decision attributes, respectively.

U	A1	A2	D
x1	16-30	50	yes
x2	16-30	0	no
x3	31-45	1-25	no
x4	31-45	1-25	yes
x5	46-60	26-49	no
x6	16-30	26-49	yes
x7	46-60	26-49	no

   In the context of rough set theory, explain lower and upper approximations, boundary region, and degree of dependency and significance of an attribute with the above example data. [2 X 5 = 10]
   b) (i) Describe the *semantic network* with a suitable example.
   (ii) Write a program in Prolog for post-order traversal of a binary tree. The traversal method stores the elements of the tree in a list. [5 + 5 = 10]
   c) Discuss the Dempster-Shafer theory of evidence with an example. Suppose an initial observation S_1 confirms some hypothesis h with the belief MB = 0.3. The second observation S_2 confirms the same hypothesis h with the belief MB = 0.5. Find the certainty factor regarding the hypothesis h by the two observations S_1 and S_2. [6 + 4 = 10]
   d) Explain with suitable examples the differences between:
   (i) Simulated annealing and steepest ascent hill climbing approach;
   (ii) Uniform cost search and best first search. [5 + 5 = 10]

INDIAN STATISTICAL INSTITUTE

End Semestral Examination

M. Tech (CS) - II Year, 2011-2012 (Semester - III)

Advanced Algorithms for Graphs and Combinatorial Optimization Problems

Date : 30.11.2011 Maximum Marks : 100 Duration : 3.5 Hours

The paper carries 125 marks. The maximum you can score is 100.

(Q1) You are asked to design a polynomial time deterministic algorithm for finding a clique of size k (k is a fixed integer that you know before designing the algorithm. As an example, you know before designing the algorithm that $k = 5$.). Can you design such a polynomial time deterministic algorithm? If you can, give a sketch of such an algorithm. Now, does that contradict the fact that the *CLIQUE* problem is NP-Complete? Justify. [4+4=8]

(Q2) Prove or disprove the following statement. *If* Π_1 *and* Π_2 *are two* NP*-Complete problems, then* $\Pi_1 \leq_P \Pi_2$. [5]

(Q3) Let a problem $\Pi \in$ co-NP $\cap$ NP. Now, someone proves Π to be NP-Complete. What can be the implication of the above? Deduce it. [8]

(Q4) Given a graph $G = \{V, E\}$, the vertex cover problem asks for a minimum cardinality subset X of V such that every edge $e \in E$ is incident to at least one vertex in X. Now, consider a greedy heuristic for the problem that constructs a set C as a vertex cover. At each step, the greedy heuristic picks up the vertex v with the highest degree in the remaining graph and then deletes v and its adjacent edges from G and adds v to C. Continuing in this fashion, the algorithm ends when G is empty.

Consider a particular graph $\mathcal{G}$ constructed by the following method. Firstly, just consider a set of n vertices v_i, where $1 \leq i \leq n$. Now, $\forall k$, $1 \leq k \leq n$ and for each $1 \leq j \leq n/k$, add a node w_j^k with edges to every node in the set $\{v_{k(j-1)+1}, v_{k(j-1)+2}, \cdots v_{kj}\}$. Analyze correctly the size of the vertex cover of $\mathcal{G}$ that will be returned by the greedy heuristic discussed above and find out the approximation ratio that will be obtained in this case. Note that, the optimal solution can be obtained by picking up the nodes v_i which are n in number. [16]

(Q5) Given a graph $G = \{V, E\}$, a subset of edges E_M of E is said to be a matching if no two edges of E_M share a vertex. A matching of *maximum cardinality* in E is said to be a *maximum matching*. A matching that is *maximal* under inclusion is called a *maximal matching*. Both maximum matching and maximal matching can be solved in polynomial time. Design an approximation algorithm for the problem of finding a minimum cardinality maximal matching in an undirected graph. [12]

(Q6) Let C_n denote a chordless cycle of n vertices. Consider the following statement. *For $n \geq 5$, the graph $\overline{C}_n$ is not triangulated.* If you consider the statement to be true, prove it. Else, give a counterexample. [8]

(Q7) (a) Define the concept of *perfect graph.*

(b) Prove that if the two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ are perfect, then the graphs $G_1 \bigcup G_2$ and $G_1 + G_2$ are also perfect, where $G_1 \bigcup G_2 = (V_1 \bigcup V_2,\ E_1 \bigcup E_2)$, and $G_1 + G_2 = (V_1 \bigcup V_2,\ E_1 \bigcup E_2 \bigcup \{(u, v) | u \in V_1, v \in V_2\})$.

[2+(6+6)=14]

(Q8) (a) Define *simplicial vertex* and *perfect elimination order.*

(b) Show that every triangulated graph $G = (V, E)$ has a simplicial vertex. Show that if G is not complete, then it has two non-adjacent simplicial vertices.

(c) Show that, if G is a triangulated graph, then it has a perfect elimination order.

[Note: If you need any result to prove the above, state it explicitly.]

[(2+2)+8+8=20]

(Q9) Consider a set of horizontal line segments placed in the first quadrant of the coordinate system. Design and analyze an efficient algorithm for computing the minimum number of vertical lines, originating from the x-axis, that are required to stab all the line segments. Note that, if a pair of line segments are overlapping, then they may be stabbed by a single vertical line if its originating position on the x-axis is chosen appropriately. [10]

(Q10) Prove that every interval graph is triangulated. [8]

(Q11) (a) Define a comparability graph clearly explaining the terms used.

(b) Prove or disprove the following statement. *Every bipartite graph is a comparability graph.*

(c) Show that an undirected graph G is an *interval graph* if and only if G is chordal and $\overline{G}$ is comparability.

[4+6+6=16]

INDIAN STATISTICAL INSTITUTE

First-Semester Examination: 2010-11

Course Name: M-TECH (QROR) II Year.

Subject Name: Software Engineering

Date: 02 - 12 - 2011 ***Maximum Marks: 60*** ***Duration: 2½ hours***

Answer any six questions.

1. How *use cases* can be useful at various stages of a software development process. [10]

2. How Structured design methodology is different from Object oriented design methodology? What is common between them? [5+5=10]

3. Explain software reliability. How is it different from Hardware reliability? Describe four metrics that help in measuring the software reliability. [2+2+6=10]

4. Explain in details with an example: (a) Halstead's Complexity measure, and (b) Cyclomatic Complexity. [5+5=10]

5. Why do we need software architecture? How will you evaluate the architecture of a software system (explain one method of evaluation)? [5+5=10]

6. Explain the role of ISO and CMM in the software development process. How these standards influence the software product. [5+5=10]

7. Why testing of software is important? Explain any method of software testing. How will you improve the reliability of software? [3+3+4=10]

8. Explain software quality. Define at least six software quality attributes. [4+6=10]

------**XXX**------

Note: Books, note-books, computers, mobiles etc. are not allowed during examination. Only calculators can be used.

INDIAN STATISTICAL INSTITUTE

Semester Examination : (2011-2012)

M.Tech. (Computer Science) II Year

Database Management Systems

Date: 2.12.2011 **Maximum Marks: 50** **Time: 2 Hours.**

Note: Answer all questions

1. Using the relations R1 (A, B, C) and R2 (C, D, E) the following query is executed:

$$\sigma_{R1.B = "X" \wedge R2.D = "Y"} (R1 \infty R2)$$

Relations R1 and R2 have 100 and 200 tuples respectively. The space requirements for different attributes are: A=10 bytes, B=20 bytes, C=30 bytes, D=20 bytes and E=20 bytes.

a) Considering that in R1 attribute C cannot have any value that is not present in R2, i.e. C is a foreign key in R1, estimate the size of the natural join of R1 and R2. Size is the estimate of total space occupancy in bytes.

b) If R1 and R2 are to be joined using block-oriented loop algorithm, calculate the number of disk accesses necessary in both cases when either R1 or R2 is placed in the outer loop. For each relation, consider 10 tuples occupy a block.

c) What would be the number of disk accesses in both cases above, when relation R1 (the smaller of the two) can be totally accommodated in the main memory during execution of the join.

d) Let V(R,X) represent the variety of values that an attribute X can have in a relation R. Now if in the two given relations, V(R1,B) =20 and V(R2,D)=10 and the distribution of the different values of these two attributes in both the relations is uniform, what would be the size of the resultant relation generated by the above query? Once again, size is the estimate of total space occupancy in bytes. Consider that join is done before selection as shown in the above query.

e) Consider that selection is done early for both R1 and R2 in the above query and the join is done later. So the equivalent query becomes:

$$(\sigma_{R1.B = "X"} (R1)) \infty (\sigma_{R2.D = "Y"} (R2))$$

Also consider that both the intermediate relations can be accommodated in the main memory. Now, what would be the total number of disk accesses to execute the query when both R1 and R2 are sorted on attribute C and Sort-Merge join is used. Ignore the cost of sorting.

(3+(2x2)+(2x2)+4+5=20)

2. For the relation R=(A,B,C,D,E,F), the following dependencies are specified :

A $\rightarrow$BC, CD $\rightarrow$E, B $\rightarrow$D, E $\rightarrow$A.

a) List all the candidate keys of R, involving only two attributes.

b) Is the decomposition (A,B,C) and (C,D,E) lossless? Explain.

c) Justify that the relation R is not in BCNF.

d) Decompose R into a set of normalized relations free from partial and transitive dependencies.

e) If a new multivalued dependency A->->F is introduced, what would be the new set of normalized relations?

(4+4+4+6+4=22)

3. Two transactions T_o and T_1 are executed sequentially as shown below. Both the transactions are manipulating the data item A. If a crash occurs in one of the four places (1 to 4) as indicated in the schedule, explain the recovery action the system would undertake if it follows a deferred update log maintenance strategy with 'redo' and 'undo' routines. Here, 'redo' and 'undo' are two procedures used in the recovery process where 'redo' forces the new values to the involved data items signifying as if no crash had occurred, 'undo' , however, sets the old values signifying as if no execution of transactions was ever done. Consider that the log contains a check point after the commit of T_o and the crash at place 2 occurred before the check point is inserted.

Schedule:

T_o : read (A)
A=A-1
write(A)
----------(1)
commit
----------(2)
T_1 : read (A)
A=A+2
write (A)
----------(3)
commit
----------(4)

(2x4=8)

※

Indian Statistical Institute

Semester-1 2011-2012

M.Tech.(CS) - Second Year

End-semester Examination (5 December, 2011)

Subject: Compiler Construction

Total marks: 110 Maximum marks: 100 Duration 3 hrs.

Please keep your answers brief and to the point.

1. Consider the following syntax-directed definition (SDD) based on a left recursive grammar:

$L \rightarrow$ **id** *Elist*	{ `Elist.array = lookup(id.name);`
	`L.place = Elist.array;`
	`L.offset = Elist.place * elt_size(Elist.array);` }
$L \rightarrow$ **id**	{ `L.place = lookup(id.name); L.offset = 0;` }
Elist $\rightarrow$ *Elist*$_1$, E	{ `Elist1.array = Elist.array;`
	`Elist.ndim = Elist1.ndim + 1;`
	`Elist.place = E.place +`
	`Elist1.place * num_elts(Elist1.array, Elist1.ndim);` }
Elist $\rightarrow E$	{ `Elist.ndim = 1; Elist.place = E.place;` }

   (a) Explain in one line each whether `Elist.array`, `Elist.ndim`, `Elist.place`, `L.place`, `L.offset`, are inherited or synthesized attributes. Is the above SDD (i) S-attributed? (ii) L-attributed? Justify your answer in 1-2 lines.

   (b) Convert the SDD into an equivalent translation scheme (TS) by embedding actions at appropriate places in the RHS of each rule.

   (c) Convert the TS into a form suitable for bottom-up parsing by using markers where necessary.

   (d) Assume that the attributes used in your TS are stored in the usual positions in the value stack during bottom-up parsing. For each semantic action in your TS, write code that manipulates the value stack according to the semantic action.

   (e) Eliminate left-recursion and left factors from the grammar in the above SDD.

   (f) Construct an equivalent translation scheme (TS) that uses the modified grammar (part (e)). Make sure attributes are handled correctly.

   HINT: Draw the parse tree for the string `x[i,j]` using the original grammar, and look at the values of attributes at various nodes. Now construct the parse tree for the same string using the non-left-recursive version of the grammar.

   [7+2+3+8+4+16=40]

2. In Pascal, the **repeat-until** statement has the form: `repeat S until E;`. It is equivalent to

```
do S while (not E);
```

   (a) Manually translate the statement `repeat i = i+1 until (i == N);` into 3-address code.

   (b) Give a syntax directed definition (SDD) for translating a general repeat-until statement into 3-address code using (i) symbolic labels, (ii) backpatching. You do not need to include grammar rules / semantic actions for general statements / Boolean expressions.

   [2+(4+6)=12]

P.T.O.

3. What is the output of the program below, if you assume that the parameter passing mechanism is (a) call-by-value (b) call-by-reference (c) copy-restore (d) call-by-name? Show your rough work.

```
int a = 10;
void f(x, y, z) {  x = x + 1;    a = a * y;    z = z + a;  }
void main(void) {  int b = 5; f(a, a+b, b); printf("%d %d", a, b);  }
```

[3+2+4+2=11]

4. Consider the following function in C. Assume that (i) `data` is an array of `num` integers, and (ii) integers floats and pointers occupy 4 bytes each.

```
float *mean_var(int *data, int num)
{ int i;
  float stats[2];
  /* body starts here */
  stats[0] = stats[1] = 0;
  for (i = 0; i < num; i++) {
      stats[0] += data[i];
      stats[1] += data[i] * data[i]
  }
  stats[0] /= num;
  stats[1] = stats[1]/num - stats[0] * stats[0];
  /* body ends here */
  return stats;
}
```

(a) Convert the body of the procedure into 3-address code. Each time a temporary variable is needed use a new temporary. You may use the name of an array instead of the constant (base address associated with that array. **Do not perform any optimization at this stage.**

(b) Assume that it takes 256 bytes to store the saved machine status. Draw a suitable layout for the Activation Record (AR), including byte offsets, for `mean_var`.

(c) Write the machine code for the calling sequence and return sequence for a call to `mean_var` from `main`. Assume that (i) the stack grows from low addresses to high addresses; (ii) the stack pointer points to the beginning (i.e. lowest address) of an AR; and (iii) the AR for `main` occupies 320 bytes.

(d) Identify the leaders in the 3-address code in (a) and draw the flow-graph.

(e) Optimize your intermediate code in (a) by using whichever of the following techniques are applicable: constant folding, global common sub-expression elimination, copy propagation, dead code elimination, code motion, induction variable elimination.

(f) Briefly explain the single most important flaw in the above code.

[15+6+(6+4)+(2+4)+8+2=**47**]

END

INDIAN STATISTICAL INSTITUTE
M. Tech. (QR OR); II Year (2011-2012)
Semester: I
Subject: Advance Statistical Methods

Date: 07 .12.2011 Maximum Marks: 100 Duration: 3½ hours

Note: This paper carries 118 marks. You can answer any part of any question, but maximum you can score is 100.

1) Write True or False

i) Single linkage method & complete linkage method lead to the same cluster.
ii) Prior Knowledge about the distribution of the variables is a prerequisite for clustering.
iii) Principal components do not always lead to meaningful interpretation
iv) One can go back to the original variable in the case of principal component regression.
v) A discriminant function always minimizes the distance between pairs of observations.
vi) A model developed by multiple linear regression method represents the underlying causal model.

[6]

2) Consider the following problems:

a) In a steel melting shop there are two furnaces. The sources of raw material are the same for both the furnaces and from each heat of steel a single sample is collected and chemical analysis is carried out. Percentages of four elements are recorded. A team of engineers has hypothesized that two furnaces produce steel of different chemical compositions.

b) Two batches of steel components have arrived. Physical properties (three in number) are measured taking samples from each batch. The manager wants to know if the physical properties are the same for both the batches.

For the above problems (a) and (b) answer the following:

i) Write down the hypothesis
ii) Write down the necessary assumptions
iii) Write down the test statistic you propose to use.
iv) What distributions does the test statistic follow?
v) What are the test criteria?
vi) What should be the lower bound for the sample size?
vii) Would you prefer univariate tests instead? Justify your preference.

$(2 + 3 + 2 + 2 + 2 + 2 + 2) = [15]$

3) a) Suppose that Hotelling's T^2 test statistic was used to test the hypothesis $H_o : \mu_1 = \mu_2$ based on two samples of size n_1 and n_2 from two multivariate normal populations; $N_p(\mu_1, \Sigma)$ and $N_p(\mu_2, \Sigma)$ respectively.

i) In the event of rejection of the hypothesis, explain how will you identify the significant variables responsible for rejection of hypothesis?

ii) Write down the sample linear discriminant function, which maximally separate the two populations.

iii) Is there any relationship between the discriminant criteria and Hotelling's T^2 test statistic ?

b) Observations are taken from two bivariate normal populations with equal variance covariance matrix Σ. The data and some relevant computations are given below:

$$Y_1 : \left[\begin{pmatrix}7\\4\end{pmatrix}, \begin{pmatrix}5\\7\end{pmatrix}, \begin{pmatrix}5\\6\end{pmatrix}, \begin{pmatrix}6\\9\end{pmatrix}, \begin{pmatrix}7\\7\end{pmatrix}\right]$$

$$Y_2 : \left[\begin{pmatrix}9\\10\end{pmatrix}, \begin{pmatrix}6\\8\end{pmatrix}, \begin{pmatrix}9\\5\end{pmatrix}, \begin{pmatrix}8\\7\end{pmatrix}, \begin{pmatrix}10\\8\end{pmatrix}\right]$$

$$\bar{y}_1 = \begin{pmatrix}6.0\\6.6\end{pmatrix}, \quad \bar{y}_2 = \begin{pmatrix}8.0\\7.60\end{pmatrix} \quad S_1 = \begin{bmatrix}0.80 & -0.40\\-0.40 & 2.60\end{bmatrix}, \quad S_2 = \begin{bmatrix}1.84 & -0.04\\-0.04 & 2.64\end{bmatrix}$$

i) Construct the discriminant function, which best separates the two populations. Which variable has highest contribution in group separation?

ii) Consider a new observation y' = (4, 8). Does this observation belong to population 1 or population 2 ?

iii) Assuming prior probabilities $p_1 = .7$ and $p_2 = .3$ compute the error rate (APER) of the classification rule.

c) Show that the optimal classification rule to classify k multivariate normal population with equal covariance matrices is:

Allocate $\mathbf{y}$ to G_i if $L_i(\mathbf{y})$ = largest of $\{L_1(\mathbf{y}), L_2(\mathbf{y}), \ldots L_k(\mathbf{y})\}$, $i = 1, \sim\ldots, k$,

where G_i is the ith group, and

$$L_i(\mathbf{y}) = \bar{\mathbf{y}}_i' S_{pl}^{-1} \mathbf{y} - \tfrac{1}{2} \bar{\mathbf{y}}_i' S_{pl}^{-1} \bar{\mathbf{y}}_i + \ln(p_i), \quad i = 1,2,\ldots k$$

$\bar{\mathbf{y}}_i$ is the sample mean corresponding to i^{th} population,

p_i is the prior probability, and

S_{pl} is the pooled sample variance-covariance matrix.

[(4+ 1 + 2) + (6+3+6)) +8 = 30]

4) A soft drink bottler is analyzing the vending machine service routes in his distribution system. He is interested in predicting the amount of time required by the route driver to service the vending machines in an outlet. The service activity includes stocking the machine with beverage products and minor maintenance or housekeeping. The industrial engineer responsible for the study has suggested that the two most important variables affecting the delivery time (Y) are the number of cases of product (X_1) stocked and the distance walked by the route driver (X_2). 25 observations were collected on y, x_1 and x_2.

From the data $(x'x)^{-1}$ was obtained as follows:

$$\begin{bmatrix} 0.11321518 & -0.00444859 & -0.00008367 \\ -0.00444859 & 0.00274378 & -0.00004786 \\ -0.00008367 & -0.00004786 & 0.00000123 \end{bmatrix}$$

The fitted model is

$y = 2.34123+1.615991\ x_1+0.01438x_2$

and following outputs were obtained

SSR = 5550.81092, SSE =223.73168

i) Does X_2 contribute significantly to the model given that X_1 is in the model?

ii) Is there any high leverage point in the data ? What is the effect of high leverage in the fitted model ?

iii) What do you mean by an influential observation? Is there any influential observation in the given data ? (For i) and ii) see the table on the next page)

iv) What do you understand by multicollinearity?

v) How does multicollinearity affect the least square estimates of regression coefficients? Explain considering two regressor variables.

vi) In a multiple linear regression problem with six regressors, for n = 12, the correlation matrix of the regressors is found to be

$$R = \begin{bmatrix} 1.000 & 0.052 & -0.343 & -0.498 & 0.417 & -0.192 \\ & 1.000 & -0.432 & -0.371 & 0.485 & -0.317 \\ & & 1.000 & -0.355 & -0.505 & 0.494 \\ & & & 1.000 & -0.215 & -0.087 \\ & & & & 1.000 & -0.123 \\ & & & & & 1.000 \end{bmatrix}$$

If pair-wise correlation matrix does not give any indication of multicollinearity, does it mean that there is no multicollinearity? Justify your answer. You may utilize the example you are provided

	Observed Value (y)	**Predicted Value y-hat**	Residual (e_i)	Studentized Residual (r_i)	Cook's Distance (D_i)	Leverage (h_{ii})
1	16.68000	21.70808	-5.02808	-1.6277	0.100092	0.1018
2	11.50000	10.35361	1.14639	0.3648	0.003376	0.0707
3	12.03000	12.07979	-0.04979	-0.0161	0.000009	0.0987
4	14.88000	9.95565	4.92435	1.5797	0.077647	0.0854
5	13.75000	14.19440	-0.44440	-0.1418	0.000543	0.0750
6	18.11000	18.39957	-0.28957	-0.0908	0.000123	0.0429
7	8.00000	7.15538	0.84462	~ 0.2704	0.002172	0.0818
8	17.83000	16.67340	1.15661	0.3663	0.003051	0.0637
9	79.24000	71.82030	7.41970	3.2138	3.419313	0.4983
10	21.50000	19.12359	2.37641	0.8133	0.053845	0.1963
11	40.33000	38.09251	2.23750	0.7181	0.016200	0.0861
12	21.00000	21.59304	-0.59304	-0.1933	0.001596	0.1137
13	13.50000	12.47299	1.02701	0.3252	0.002295	0.0611
14	19.75000	18.68246	1.06753	0.3411	0.003293	0.0782
15	24.00000	23.32880	0.67120	0.2103	0.000632	0.041
16	29.00000	29.66293	-0.66293	-0.2227	0.003289	0.1659
17	15.35000	14.91364	0.43636	0.1380	0.000401	0.0594
18	19.00000	15.55138	3.44862	1.1130	0.043978	0.0963
19	9.50000	7.70681	1.79319	0.5788	0.011919	0.0964
20	35.10000	40.88797	-5.78797	-1.8735	0.132445	0.1017
21	17.90000	20.51418	-2.61418	-0.8778	0.050861	0. 1653
22	52.32000	56.00653	-3.68653	-1.4500	0.451045	0.3916
23	18.75000	23.35757	-4.60757	-1.4437	0.029899	0.0413
24	19.83000	24.40285	-4.57285	-1.4961	0.102322	0.1206
25	10.75000	10.96258	-0.21258	-0.0675	0.000108	0.0666

[6+5+5+2 + 6 + 6 = 30]

5) a) What are the similarities and differences between Principal Component Analysis and factor analysis?

b) Write down the factor model and the associated assumptions.

c) Show that the factor model and communalities remain unchanged under orthogonal transformation.

d) Explain how orthogonal transformation help in factor extraction ?

e) If $\Lambda = ((\lambda_{ij}))_{p \times p}$ be the loading matrix then show that the contribution of the j^{th} factor to $V(Y_i)$ is λ_{ij}

[6+ 3 + 6 + 4 + 4 = 23]

6) i) What is the difference between hierarchical and partitioning method of clustering?

ii) Discuss suitable graphical techniques to detect grouping in a set of observations?

iii) Consider the following distance matrix for 4 items. Apply complete-linkage method to obtain the clustering.

	1	2	3	4
1	0			
2	516	0		
3	590	833	0	
4	693	881	464	0

[3·5·6 14]

INDIAN STATISTICAL INSTITUTE
M. Tech.(Computer Science) II Year, 2011-12
Semestral Examination
Pattern Recognition and Image Processing

Date: 7-12-2011 Maximum Marks: 100 Duration: 195 minutes

Note: This paper carries 107 marks. Answer as much as you can.

1. Let $x_1 = (0,0), x_2 = (0,1), x_3 = (1,0)$ and $x_4 = (1,1)$. Let $\theta_1 = \theta_3 = 1$ and $\theta_2 = \theta_4 = 2$. Let θ_i denote the class of x_i for each i. Find the straight line that separates the two classes by applying the Perceptron learning algorithm with learning rate as 0.5, and the initial separating straight line as $x - y = 0$. [15]
2. Let there be 3 classes. Let
   $A_1 = \{(0,0,0),(0,0,1),(1,0,0)\}, A_2 = \{(1,-1,0),(-2,1,-1),(-2,-1,0)\}$ and
   $A_3 = \{(1,1,0),(1,0,-1)\}$. Let the elements in A_i be from the i-th class for all i. Classify the point **x** =(-2,-2,1) to one of the three classes using 3-nearest neighbor rule. [6]

3. Apply histogram equalization method on the following frequency distribution table of gray values, and find the modified histogram. Show all the steps of your method. [10]

Gray value	Frequency of occurrence
0	3
1	15
2	23
3	19
4	7
5	6
6	4
7	3

4. Describe a procedure for introducing salt and pepper noise in a gray scale image. [5]
5. Describe the Canny's edge detection procedure. [15]
6. Describe the Hough transform method for detecting line segments in a binary image. [12]
7. (i) Describe a region based segmentation procedure for gray level images using quad tree. [8]
   (ii) Describe a procedure for object background segmentation using the histogram of gray levels. [6]

(P.T.O)

8.

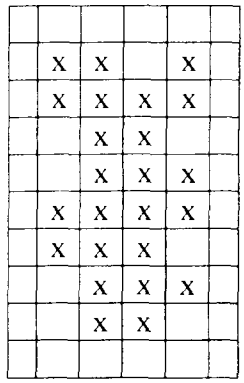


Figure (a)

An object is shown in figure (a). The pixels of the object are represented by the sign 'x'. .

(i) Describe the 8-connected boundary of the object by using chain code.

(ii) Describe a thinning algorithm. Apply it on the object in figure (a), show all the interim modifications on fig. (a), and provide the final output. [3+(8+9)]

9. A dataset is given to you where the number of observations is 80 but the number of features is 1000. You need to get 70 principal components. Your computer can be used to obtain eigen values and eigen vectors of an MxM matrix where M can't be more than 100. State a procedure for obtaining those components. [10]

Indian Statistical Institute

Semester II Examination 2011

M. Tech. (CS) - Second year

Subject: Internet & Multimedia Technologies

Date: 09.12.2011 Full Marks: 100 Duration: 3:00 hrs.

1. Answer both. $2 \times 10 = 20$

a) Modify the Bresenham's midpoint subdivision algorithm to draw the conic $f(x,y) = b^2x^2 + a^2y^2 - a^2b^2$ for the first quadrant.

b) Write algorithmic steps for MPEG-I video compression.

2. Answer any five from the following: $5 \times 4 = 20$

a) What are the primitive datatypes supported by the Java programming language?

b) What do you know about the String class in Java?

c) Is explicit initialization of variables always required in Java? Discuss.

d) Make necessary corrections to the following code.

```
public int increasePositive ( int a) {
int b;
if ( a > 0 ) b = a+1;
return b;
}
```

e) Indicate the wrong statements. (Explanation is not required)

i) float[] test = new float[3] { 1.2, 3.4, 5.6};
ii) float test[3] = {1.2, 3.4, 5.6};
iii) float[] test[] = {{}, new float[] {} };
iv) float test[][] = { {1.2, 3.4}, new float[2] };
v) float test[] = { 1.2, 3.4, 5.6, 7.8, } ;

f) Write down the output of the following code: (Explanation is not required)

```
class Marks {
double program;
double os;
double dbms;
Marks(double p, double o, double d) {
program = p;
os = o;
dbms = d;
```

```
}
public String toString() {
return "Marks are " + program + ", " +  os + " and " + dbms + ".";
}
}
class toStringDemo {
public static void main(String args[]) {
Marks m = new Marks(80, 62, 54);
String s = "Current sem: " + m;
System.out.println(m);
System.out.println(s);
}
}
```

g) What is a 'main() method' in a Java program?
Can a Java main() method be overloaded? If so, give an example.

h) What are (i) JVM and (ii) JRE. Describe them briefly.

3. Answer any five from the following: $5 \times 2 = 10$

a) What will be the results of using shift operators as in the following:
(i) -1 >> 1 and (ii) 1 << 31

b) Explain using one example for each why && and || are called short-circuit logical operators?

c) What will be the output of the following:

```
System.out.println( 1 + 3 + "3" );
System.out.println( "=/" + 1 + 3 + 3 );
```

d) What are called access modifiers in Java programming language?

e) If no access modifier is specified in a Java program, what will be the accessibility of its classes?

f) When does the initialization of static variables take place? Can you give an example of a local variable declared static?

4. Answer any one: 6

a) Write full forms of AWT, SWT and JFC.

b) Compare Java Swing with AWT.

5. Answer any three: $3 \times 3 = 9$

a) Describe various components of the following html element.
This is a link

b) When is an html attribute value is necessarily written using single quotes? How can you display text in an html document right justified?

c) Write html code to display the following ordered list in an html browser:

1. Sachin
2. Rahul
3. MSD

d) Write down three different ways of adding styles to an html page.

e) Write the full form of (i) HTML, (ii) XML, (iii) CSS.

6. Represent the following information in XML format: 6

A group has 3 artists named "Ashish", "Sandeep", "Priti".
They are from the districts "Nadia", "Howrah" and "Kolkata".
They are skilled in "Vocal", "Instrumental" and "Dance" respectively.
Their monthly incomes are respectively Rs. 15000/-, Rs. 5000/- and Rs. 20000/-.
Their levels of efficiency are respectively medium, medium and expert.

7. Answer any three from the following: $3 \times 3 = 9$

a) Name two network applications which need guaranteed delivery of byte streams. Also, name the protocol used by these applications for data transmission purpose.

b) Write down two main responsibilities of IP. What is the use of the IP address 0.0.0.0?

c) What is the full form of ARP? What is the role of ARP?

d) Why is it not possible to determine the network and host addresses from a given IP address? Do the following represent valid subnet masks (give reasons)?
(i) 00000000 00000000 00000000 00000000 and (ii) 11111111 11111111 11111111 11111111

e) What is domain name space? Name four major top-level-domains.

8. Answer any four from the following: $4 \times 5 = 20$

a) Searching for a user's query in a database is easier than searching in documents – explain.

b) Write down three major steps of indexing, a subprocess of a retrieval procedure.

c) Define the two major parameters used to measure the efficiency of a retrieval system.

d) What is a Web Crawler? Name two major web crawlers.

e) What are Internet Robots? Write down their uses.

f) Discuss four technologies used to ensure the security of E-Commerce.

INDIAN STATISTICAL INSTITUTE
M. Tech (Computer Science) II year, 2011-12
Pattern Recognition and Image Processing
Back Paper examination

Date : 11.01.12 **Maximum marks: 100** **Duration: 3 hours**
Note: Answer all the questions

1. State the Bayes decision rule for two-class classification problem and show that it provides the minimum probability of misclassification. [3+8=11]
2. Let there be two classes C_1, C_2 with prior probabilities P and (1-P) respectively, and the class conditional density functions being p_1, p_2 respectively. Let

   $p_1(x) = x; \quad x \in (0,1]$
   $= 2 - x; \quad x \in (1,2)$
   $= 0; \quad otherwise$

   and

   $p_2(x) = x - 1; \quad x \in (1,2]$
   $= 3 - x; \quad x \in (2,3)$
   $= 0 \quad otherwise$

   (a) Find the Bayes decision rule for the above classification problem and find its probability of misclassification.
   (b) Find the probability of misclassification for the following decision rule and verify that it is indeed greater than or equal to that of the Bayes decision rule.

   Put x in C_1 if $x \in (0,0.7) \cup (1,1.5)$. Otherwise, put it in C_2.
   [(5+5)+(5+5)=20]
3. (a) Describe k-means algorithm for clustering.
   (b) Describe single linkage algorithm for clustering and state any three of its properties. [5+(4+3)=12]
4. Describe any two feature selection algorithms. [8].
5. Describe Gaussian smoothing method for gray level images. [6]
6. Describe an enhancement technique by using a nonlinear transformation function. [8]
7. Describe non-maxima suppression in Canny edge detection method. [8]
8. Define skeleton of a path connected set in 2-dimensional Euclidean space. [5]

(P.T.O)

9. Let $\underline{X}' = (X_1, X_2, X_3, X_4)$ be a random vector with dispersion matrix Σ,

where $\Sigma = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 4 & -1 \\ 0 & 0 & -1 & 4 \end{pmatrix}$. Find two principal components of $\underline{X}$.

[10]

10. Describe a procedure for finding line segments in a binary image using Hough transform. [12]

INDIAN STATISTICAL INSTITUTE

Semester Examination (Back Paper) : (2011-2012)

M.Tech. (Computer Science) II Year

Database Management Systems

Date: 11.01.2012 **Maximum Marks: 100** **Time: 3 Hours.**

Note: Answer all questions

1. A set of companies producing cassettes of vocal and instrumental music want to maintain a database about their business. Each company has a unique name. It also maintains its address, annual turnover and location of production units in the database. A company may have production units in more than one location. Each company also maintains a list of the albums released by it. An album is associated with a unique name, price and the number of copies sold. While a particular album is released by only one company, a company can release a number of albums. Each company also maintains a list of musicians for whom it has already released some cassettes. A musician may have a number of albums released from one or more companies. An album may contain the performance of one or more musicians. Every musician has a unique name. The database will also store his/her address and approximate annual income. A musician can either be a Vocalist or an Instrumentalist, but not both. An instrumentalist may play more than one instrument. There are three types of vocalists: folk-singer, pop-singer and classical-singer. A folk-singer may sing more than one type of folk-song, a pop-singer may sing in more than one language and a classical-singer may have a set of favourite ragas. A vocalist may sing more than one type of song; folk-song, pop-song, classical-song. All this information is to be stored in the database.

   i) From the above description, draw an ER/EER diagram to design the database schema.
   ii) Map the ER/EER diagram to a set of relations using the standard mapping rules.

   (10+5=15)

2. Consider the following schema and form the required queries:

   Suppliers (<u>sid</u>, sname, address)
   Parts (<u>pid</u>, pname, color)
   Catalog (<u>sid, pid</u>, price)

   The key fields are underlined. Attributes sid and pid indicate the unique ids for a supplier and a part respectively.
   Form the following queries using relational algebra:

   a) List the names of such parts where no supplier is available when the color is 'red'.
   b) Now provide the names and addresses of the suppliers of the same parts when the color is other than 'red'.

   (10+10=20)

3. A relation R (A,B,C,D,E,F,G) has been decomposed into two relations R_1(A,B,C,D,E) and R_2(C,D,E,F,G). Following set of functional dependencies is given.

   $\{A \rightarrow BC, CD \rightarrow E, B \rightarrow EF, CDE \rightarrow FG\}$

   a) Justify that the above decomposition is neither dependency preserving nor lossless.
   b) With the given functional dependencies, decompose R to a set of normalized relations free from partial and transitive dependencies.
   c) If a multivalued dependency $A \rightarrow\rightarrow G$ is introduced, suggest the new set of normalized relations.

   (10+9+4=23)

4. T1 and T2 are two transactions where T2 is nested within T1. The two transactions operate on three data items a, b and c. If a crash occurs at any of the four places (1-4) as shown below, what actions would be taken by the transaction manager for recovery in each case when the system incorporates

   i) immediate update policy,

   ii) deferred update policy.

   Consider 'redo' and 'undo' are two procedures used in the recovery process where 'redo' forces the new values to the involved data items signifying as if no crash had occurred, while 'undo' sets the old values signifying as if no execution of transactions was ever done. Show the content of the log-file in each case.

```
T1 :   read a;
       modify a;
                <- 1
       write a;
T2 :   read b;
       modify b;
       write b;
                <- 2
       commit;
T1 :   read c;
                <- 3
       modify c;
       write c;
       commit. <- 4
```

(4x8=32)

5. Explain how duplicate elimination can be done in projection operation by hashing technique. Give an estimate of the buffer requirement for this purpose.

(10)

-x-

INDIAN STATISTICAL INSTITUTE

Supplementary (Back Paper) Examination

M. Tech (CS) - II Year, 2011-2012 (Semester - III)

Advanced Algorithms for Graphs and Combinatorial Optimization Problems

Date : 16.01.2012 Maximum Marks : 100 Duration : 3.0 Hours

The paper carries 120 marks. The maximum you can score is 100.

(Q1) Prove or disprove the following statement. If any problem in NP is not polynomially solvable, then no NP-Complete problem is polynomially solvable. [8]

(Q2) Say, one fine morning Captain Archibald Haddock wakes up and tells Professor Cuthbert Calculus that he has found a polynomial time deterministic algorithm $\mathcal{A}$ for a decision problem Π which was earlier proved to be NP-Complete. Professor Calculus checks $\mathcal{A}$ and finds it to be correct. Professor Calculus was jumping with joy!! Deduce out for yourself what is the implication of this invention of Captain Haddock towards the classes of problems belonging to NP. [8]

(Q3) Consider the following statement. Let Π be an NP-Complete problem. If $\Pi \in$ co-NP, then NP = co-NP. Prove or disprove the above statement. [10]

(Q4) Consider the following problem which is known as *Set Cover*. Given a set $\mathcal{U}$ of n elements, a collection $\mathcal{S} = S_1, \ldots, S_m$ of subsets of $\mathcal{U}$, and a number k, does there exist a collection of at most k of these sets whose union is equal to all of $\mathcal{U}$. Show that *Set Cover* is NP-Complete. [12]

(Q5) Let $G = \{V, E\}$ be an undirected graph. We know that a subset $S \subseteq V$ is an INDEPENDENT SET of vertices if and only if $V - S$ is a VERTEX COVER. We also saw that if S is of maximum cardinality then $V - S$ is of minimum cardinality. Now, consider the 2-approximation algorithm discussed in the class for MVC using *maximal matching*. Considering the fact that the problems VERTEX COVER and INDEPENDENT SET are complementary in nature, can we use the 2-approximation algorithm for MVC to find an approximation algorithm for MAXIMUM INDEPENDENT SET? If yes, give the algorithm; if no, explain why? [10]

(Q6) (a) Define (i) a chordal graph and (ii) a perfect elimination order.

(b) Show that if a graph G has a perfect elimination order, then G is chordal.

[(2+2)+4=8]

(Q7) Give an example of a graph G for which $\alpha(G) = k(G)$ and $\omega(G) < \chi(G)$, where $\alpha(G)$ stands for the maximum independent set, $k(G)$ stands for clique cover number, $\omega(G)$ stands for clique number and $\chi(G)$ stands for chromatic number.

Does this contradict the Perfect Graph theorem?

[10+2=12]

(Q8) Show that a triangulated graph on n vertices has at most n maximal cliques, with equality if and only if the graph has no edges. [10]

(Q9) Show that the complementary graph of a chordless cycle of at least 5 vertices is not triangulated. [8]

(Q10) (a) Define *simplicial vertex*.

(b) Show that every triangulated graph $G = (V, E)$ has a simplicial vertex. Show that if G is not complete, then it has two non-adjacent simplicial vertices.

[Note: If you need any result to prove the above, state it explicitly.]

[2+10=12]

(Q11) Consider a set of vertical line segments $\mathcal{L}$ placed in the first quadrant of the coordinate system. Design and analyze an efficient algorithm for computing the minimum number of horizontal lines, originating from the y-axis, that are required to stab all the line segments in $\mathcal{L}$. Note that, if a pair of line segments are overlapping, then they may be stabbed by a single horizontal line if its originating position on the y-axis is chosen appropriately. [12]

(Q12) (a) Define a comparability graph clearly explaining the terms used.

(b) Give an example of a graph of at least 6 vertices that is a chordal graph but not a comparability graph.

[4+6=10]

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination : (2011-2012)
M.Tech.(CS) II Year

Advanced Database Theory & Applications

Date : 21.02.2012 **Maximum Marks : 50** **Duration : 2 Hours**

. Answer all questions.

1 The concurrent schedule shown below involves three transactions T1 to T3 using two data items X and Y for read and write.

Time Slot	T1	T2	T3
1	read(X))		
2			read(Y)
3		write(X)	
4			write(Y)
5			write(X)
6	read(Y))		
7		write(Y)	

a) Drawing a precedence graph show whether the above schedule is conflict serializable?

b) To examine view serializability, draw the labeled precedence graph for two data items separately to show whether they are individually view serializable?
Also draw the composite labeled precedence graph to examine whether the schedule is view serializable considering the two data items together?

(5+10+5=20)

2 Let us consider that timestamp based protocol is used for the execution of the concurrent schedule given in Question 1. According to the schedule, each read or write operation needs one time slot. In case of any time conflict, a transaction is allowed to rollback and restart with a new timestamp higher than all the existing timestamps. No such conflict is ignored. i.e. Thomas' Write Rule is not considered. It is assumed that a rolled back transaction is rescheduled immediately. The unused time slots obtained due to the rollback of a transaction may be utilized by a rescheduled transaction. A transaction may have any number of rollback and restart. Find the total number of time slots required to execute all the three transactions without changing the given schedule. The timestamps of the three transactions are related as,

$$TS(T1) < TS(T2) < TS(T3)$$

For each instruction executed, show the status of the read and write timestamps of the data item involved.

(15)

3 Considering Wound-Wait scheme as deadlock detection mechanism, evaluate the final execution order of the three transactions if they follow the schedule given in Question 1. The three transactions follow the timestamp order given in Question 2. Consider that between two transactions, one having lower value of timestamp has a higher priority.

(15)

Indian Statistical Institute

Periodical Examination of Second Semester (2011-2012)
M. TECH. (CS) 2 Year
Subject: Information and Coding Theory
Date: 26 /02/2012 Time: 2 hours Maximum Marks: 40

Note: The paper carries 45 marks. Maximum you can score is 40.

1. Prove that $I(X;Y) \geq 0$ with equality if and only if X and Y are independent. [5]

2. Define channel capacity, noisy typewriter and the set of jointly typical sequences. [5]

3. Prove that Huffman code is optimal. [10]

4. State the channel capacity theorem and give an outline of the proof. [10]

5. Let $\mathcal{X} = \{1, 2, 3, 4\}$ and $X_1, X_2, X_3 \ldots$ be a time invariant Markov chain where X_is take values from the set $\mathcal{X}$. The transition matrix is

$$P = \begin{bmatrix} 2/3 & 1/6 & 1/6 \\ 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}.$$

   Find the stationary distribution of the Markov chain. [5]

6. Let (X, Y) have the following joint distribution:

Y \ X	1	2	3	4
1	1/8	1/16	1/32	1/32
2	1/16	1/8	1/32	1/32
3	1/16	1/16	1/16	1/16
4	1/4	0	0	0

   Find $H(X), H(Y), H(X|Y), H(Y|X)$ and $I(X;Y)$. [10]

Indian Statistical Institute

Advanced Image Processing

M.Tech.(CS)-II Year: 2011-12

Full marks: 60 Time: 2 Hours

Date: 24.02.2012

Answer **any six** questions. All questions carry equal marks.

1. Assuming pin-hole camera model for perspective projection from 3-D to 2-D, prove that

   (i) a set of parallel straight lines not perpendicular to z-axis is mapped to a set of concurrent straight lines. What is that common point called?

   (ii) distant objects appear smaller.

   [5 + 5]

2. Starting from three basic assumptions of photometric model of image formation, derive the equation (all terms have their usual meaning):

$$g(x,y) = \int_{\infty}^{\infty} \int_{\infty}^{\infty} h(x-\alpha, y-\beta) f(\alpha,\beta) d\alpha d\beta$$

   [10]

3. Describe image compression using Vector Quantization. [10]

4. Consider the following block of graylevels:

6	6	2	3
9	8	4	1
8	2	3	7
4	2	7	8

   Calculate the compressed and reconstructed representation of the block using Block Truncation Coding. Calculate PSNR and bpp. [8 + 2]

5. Prove that the Fourier transform of a two-variable function $f(x,y)$ is rotated by an angle θ if $f(x,y)$ is rotated by the same angle. Obtain the Fourier

transform of the Laplacian of a two-variable function $f(x, y)$. Assume that x and y are continuous variables. [6 + 4]

6. Describe the Fast Fourier Transform algorithm and discuss its computational complexity. [8 + 2]

7. Describe the thresholding method proposed by N. Otsu. How do you extend this method to obtain multiple thresholds. [7 + 3]

8. State and prove the correlation theorem. Show that the Fourier transform of the autocorrelation function of $f(x)$ is its power spectrum. [7 + 3]

INDIAN STATISTICAL INSTITUTE

Periodical Examination

M. Tech (CS) - II Year (Semester - II

Multidimensional Searching and Computational Geometry

Date : 24.2.2012 Maximum Marks : 60 Duration : 2.5 Hours

Note . You may answer any part of any question, but maximum you can score is 60.

1.(a) Show that given a simple polygon P with n vertices, the convex hull of the vertices of P can be computed in $O(n)$ time.

(b) You are given two x-monotone polygonal chains P and Q with a total of n vertices, prove that they can intersect at most $O(n)$ times. [10+10=20]

2.(a) Define the following terms in the context of *Fortune's Line Sweep Algorithm* for constructing the Voronoi diagram of a set of points:
(i) beach line, (ii) site event and (iii) circle event

(b) Show that (i) the only way in which a new arc can appear on the beach line is through a site event, and (ii) the only way in which an existing arc can disappear from the beach line is through a circle event.

(c) State an $O(n\log n)$ time algorithm which takes input a set of points S and reports for each point in S its nearest neighbor in S. If you take help of the algorithm of an well-known problem, then state that problem and the time complexity of that algorithm.

[(3*2)+(6+6)+7=25]

3.(a) Given a set of red $2n$ points and a set of blue $2n$ points such that their convex hulls are linearly separable. Show that there exists a linear time algorithm that can compute a line ℓ which splits both the red points and blue points into two subsets of size n each. Justify the correctness and time complexity of your proposed algorithm.

(b) You are given a set of n vertical line segments in the plane. Present an efficient algorithm to determine whether there exists a line that intersects all of these segments. (Hint: $O(n)$ time algorithm is possible.) Justify your algorithm's correctness and derive its running time. [15 + 10=25]

INDIAN STATISTICAL INSTITUTE

Mid - Semestral Examination: 2011 – 2012

Course Name: M. Tech (CS) - II

Subject Name: Computer Vision

Date: 27/02/2012, Maximum Marks: 50 Duration : 1hr. 30 min

Note, if any: Attempt all the questions

Q1.(i) Write down two main differences between the perspective projection and orthographic projection. 2+2

(ii) What kind of projection will you have for a wide angle lens and a telephoto lens? Justify your answer. 2+2

(iii) Give the physical significance of BRDF? How is it measured? 2+2

Q2. (i) What is meant by the point spread function? How will you check the point spread function of a camera? 2+2

(ii) What is the significance of the unit impulse δ (x, y)? Find the Fourier transform of $\nabla^2 f$.

where $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the Laplacian operator in two dimension. 4+6

Q3. The second moment of any region about an axis inclined at an angle θ can be written in the form

$$E = a\sin^2\theta - b\sin\theta\cos\theta + c\cos^2\theta$$

(i) Compute the major and minor axes of an equivalent ellipse that has the same second moment about any axis through the origin. 5+5

(ii) Use the definition 6

$$\delta_\varepsilon(x,y) = 1/(4\varepsilon^2) \quad for \ |x| \le \varepsilon \ and \ |y| \le \varepsilon$$

$$= 0, \quad otherwise$$

to find the Fourier transform of the unit impulse.

Q4. What do you mean by the reflectance map of a Lambertian object? Sketch the contours of constant brightness in the gradient space. 4+4

INDIAN STATISTICAL INSTITUTE

Mid Semestral Examination

M. Tech (CS) - II Year, 2011-2012 (Semester - IV)

Topics in Algorithms and Complexity

Date : 27.02.2012 Maximum Marks : 60 Duration : 2.5 Hours

Note: Answer as much as you can, but the maximum you can score is 60.

This is a 2-page question paper with 7 questions.

(Q1) Consider the following algorithm where RANDOM(i) is a function that takes an integer i as an input and generates a random integer between 1 and i in constant time.

Algorithm 1: An algorithm to generate a random permutation.

Input: An array $A[1 \ldots n]$

Output: The array $A[1 \ldots n]$ with the same elements, but rearranged into a random permutation.

```
1 for i ← n downto 2 do
2     indexrandom ← RANDOM(i);
3     Exchange A[i] and A[indexrandom];
  end
```

(a) Prove that every possible permutation of A is equally likely to be the output of the above algorithm.

(b) If we change the i in line 2 to n, does the algorithm still produce every possible permutation with equal probability? Explain with proper argument.

[6 + 6 = 12]

(Q2) Consider the randomized algorithm for the two dimensional linear programming studied in class.

(a) Extend the algorithm so that it works for d dimensions, where $d > 2$.

(b) Deduce the time complexity as a function of n and d, where n is the number of linear constraints.

[4 + 6 = 10]

(Q3) Let $\mathcal{A}$ be the set of all vectors of the form $a = (a_1, \ldots, a_r)$, where a_i is an integer in the range $[0, p-1]$ (p is a prime) for each $i = 1, \ldots, r$. For each $a \in \mathcal{A}$, we define the linear function $h_a(x) = (\sum_{i=1}^{r} a_i x_i) \bmod p$. We define a family of hash functions $\mathcal{H} = \{h_a : a \in \mathcal{A}\}$.

Is it necessary for p to be a prime so that $\mathcal{H}$ is a universal class of hash functions? Explain with proper arguments. [8]

(Q4) Consider the following algorithm of selecting a sample of m elements randomly from a set of n elements in an array A, where $m \leq n/2$.

First mark all the n elements in A as *unmarked*. Next, repeat the following steps until exactly m elements have been selected. Generate a random number r between 1 and n. If $A[r]$ is marked *unselected*, then mark it *selected* and add it to the sample.

Analyze the expected running time of this algorithm. [10]

(Q5) Let $G = (V, E)$ ($|V| = n$) be an undirected graph in which each node is incident to exactly d edges. A *dominating set* in G is a subset $S \subset V$, with the property that every node in $V \setminus S$ is adjacent to a node in S.

(a) Deduce a lower bound on the size of any dominating set in G.

(b) Show that for some constant c, a set of $\frac{cn \log n}{d+1}$ nodes chosen uniformly at random from G will be a dominating set with high probability. [Note that the term 'high probability' means 1 upon a polynomial in n.]

[4 + 11 = 15]

(Q6) **(a)** Define a Probabilistic Turing machine. Explain briefly whether or not it is same as a non-deterministic Turing machine.

(b) Distinguish between the complexity classes RP and ZPP. Give an example of a problem in each of these two classes.

[(2 + 2) + (2 + 4) = 10]

(Q7) **(a)** What is the relation between the class BPP and its complement? Justify your answer.

(b) State the Sipser-Gács-Lautemann Theorem. Sketch its proof.

[2 + (2 + 6) = 10]

INDIAN STATISTICAL INSTITUTE

Periodical Examination: (2012)

M.Tech (CS) II Year

Parallel Processing: Architectures and Algorithms

Date: 29/02/2012 Full Marks: 100 Duration: 3 hrs

Answer any Five

1. Show the schematic diagrams of a Parallel Random Access Machine (PRAM), an SIMD machine and an MIMD machine. Compare the features of PRAM model with those of SIMD and MIMD architectures and mention which architecture can best model the PRAM machine.
Mentioning the features of EREW and CRCW models of PRAM, prove that the CRCW model with N processors can always be simulated on the EREW model having the same number of processors with an $O(\log N)$-fold increase in the processing time.

[(6·6)+(2·6) 20]

2. a) A uniprocessor system with 15 MHz clock executes a given program mix with 10 MIPS (million instructions per second) rate. Each memory access needs one cycle delay.
   I) Calculate the effective CPI (cycles per instruction) of the machine.
   II) Suppose the clock is now replaced with a 30 MHz clock. However, the memory system remains unchanged, and consequently requires two cycles for every memory access. In the program mix, if 30% of the instructions require one memory access and another 5% require two memory accesses per instruction, calculate the CPI and MIPS rate of the upgraded machine.
   b) Distinguish between medium-grain and fine-grain multicomputers in their architectures and programming environments.

[(4·(7·3))+6 20]

3. Given a PRAM machine with n^3 processors, write down an efficient algorithm to multiply two $n \times n$ matrices A and B in $O(\log n)$ time (n is a power of 2). Mention which memory model is required. Find out the speedup and utilization. Is it possible to improve the utilization using less number of processors keeping the time complexity unchanged? Justify your answer.

[8·2+4·6 20]

4. a) Prove that a complete binary tree with (N-1) nodes is not contained in an N-node hypercube H_n ($N=2^n$).
   b) In a complete binary tree with $N = 2^n$ leaf nodes, each leaf node is labeled as a unique binary string $x_{n-1}\ x_{n-2}\ \ldots x_1\ x_0$ from left to right. Each leaf node $x_{n-1}\ x_{n-2}\ \ldots x_1\ x_0$ contains exactly one packet to be routed to the leaf node $x_{n-1}'\ x_{n-2}\ \ldots x_1\ x_0$ in the other half (x_{n-1}' represents the complement of x_{n-1}). Find out the number of routing steps required to deliver all the packets to the respective destinations. Also find out the maximum buffer size required at a node to store the incoming packets. Assume all links are bi-directional, and each node can simultaneously transmit/receive over all links.
   c) Find how the use of *Fat Tree* can improve the delay in the above routing.

[6·(6·3)+5 20]

5. a) Show the block diagram of a 16×16 Clos network using $4 \times m$ switches where m takes the minimum value required for non-blocking operation of the network.
   (b) Given a BPC permutation P: $x_2\ x_1\ x_0 \rightarrow x_0\ x_2\ x_1$, show the path matrix of P for 8×8 Omega network. Hence draw the conflict graph and find out the minimum number of passes required to route

P in Omega network with the corresponding partitions of the paths.
Is *P* admissible in the Omega network with one or two additional stages? Justify your answer.

[5+ (6+4+5)=20]

6. Answer in brief:
   a) Prove that H_n, a hypercube of order n, is Hamiltonian, for $n \geq 2$.
   b) Show how in a machine with N processors interconnected by *shuffle-exchange* interconnection network, any processor can broadcast a datum in minimum number of steps ($N = 2^n$).
   c) Find out the reliability of a 4×4 mesh in presence of two node faults.
   d) Show that the bisection width of a pyramid of size $P = 4k$ is $2\sqrt{P}$.

[5 × 4 = 20]

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination : (2011 - 2012)

Course Name : M.Tech (CS) **Year : 2nd year**

Subject Name : Neural Networks & Applications

Date : **Maximum Marks : 50 Duration : 2 hrs**

Answer all the questions.

1. Explain the various issues you need to consider while designing an artificial neural network model. (10)

2. Show how a single-node perceptron model can be used to solve linearly separable two-class pattern classification problem. Consider n-dimesional patterns distributed in 2 classes. (15)

3. Consider a pattern classification problem where n-dimensional input patterns are distributed in l classes. These patterns need to be classified using a multilayer perceptron (MLP) with one hidden layer consisting of m nodes. Derive the expressions for the amount of modification of all the weights of this MLP that is required for its training under the gradient-descent technique for minimization of error. Assume sigmoid functions, with necessary bias, as activation functions of all the hidden and output nodes. (25)

INDIAN STATISTICAL INSTITUTE

Semestral Examination

M. Tech (CS) - II Year (Semester - II)

Multidimensional Searching and Computational Geometry

Date : 23.5.2012 Maximum Marks : 100 Duration : 3.5 Hours

Note : You may answer any part of any question, but maximum you can score is 100.

1.(a) For a given set P of n points in $\mathbb{R}^2$, let ρ be the minimum value of the radius of a k-enclosing circle (i.e., a circle that encloses exactly k points). Present an efficient algorithm for computing a k -enclosing circle whose radius is guaranteed to be less that 2ρ. Derive the time complexity and justify the approximation bound of your proposed algorithm.

(b) Using the same idea that you will use for proposing a 2-approximation algorithm for computing the k-enclosing circle, can you design 2-approximation algorithm for computing the k-enclosing square for the point set P? If so, justify; if not, say the reason. [15+10=25]

2. Consider a bounded height binary tree to be maintained in a secondary storage (disk).

(a) Suggest a suitable organization of the tree and the corresponding record structures in the disk file for this purpose. (b) Write in brief your insertion and search algorithm of a key in this structure, and (c) analyze the number of disk probes for each operation. (* Detailed description of the algorithms are not required. Only the key ideas and their proper justifications will be enough. *) [(5+5)+(5+5)+(4+4)=28]

3. Describe an incremental algorithm for computing the farthest point Voronoi diagram of a set P of n points in $\mathbb{R}^2$, and the input points are given in an array of size n in arbitrary order. State and prove the time and space complexities of your proposed algorithm. Mention all the intermediate steps of your algorithm and state the time complexities of those steps with proper justifications. [15]

4. You are given a set P of n points in $\mathbb{R}^2$. A corridor is a region defined by a pair of lines that are parallel to each other. Write efficient algorithms for computing (i) the widest empty corridor among the points in P and narrowest corridor containing all the points in P. You are not allowed to use more than $O(n)$ space in addition to the space required for storing the input points.

State and prove the time complexity of your proposed algorithms. Mention all the intermediate steps of your algorithms and state the time complexities of those steps with proper justifications. [15+10=25]

5. Let S be a set of n (possibly intersecting) disjoint line segments in $\mathbb{R}^2$. Design an efficient data structure for storing S such that given an arbitrary vertical line L (defined by its x-coordinate), one can report all the members in S intersected by L in $O(k + polylog(n))$ time. Justify the preprocessing time, space complexities, and the query time complexity of your proposed algorithm. [15]

Indian Statistical Institute

M. TECH. (CS) 2 Year : 2011–2012
Semester Examination
Subject: Information and Coding Theory
Date: 24/04/2012 Time: 3 hours Maximum Marks:100

Note: The paper carries 110 marks. Maximum you can score is 100.
Notations used are as explained in the class.

1. Prove that for a binary symmetric channel with crossover probability $p < 1/2$, the maximum likelihood decoding rule is the same as the nearest neighbour decoding rule. [10]

2. Define Hamming distance between two vectors. Prove that the Hamming distance is a metric. [10]

3. Define linear code, generator matrix, parity check matrix, Hamming code, cyclic code, BCH code, RS code and MDS code. [16]

4. Construct a single-error correcting linear code with $n = 10$. Enumerate two codewords. [10]

5. Construct a double-error-correcting BCH code of length 15. [10]

6. Prove that a code with minimum distance d can correct $\lfloor \frac{d-1}{2} \rfloor$ errors. [10]

7. State and prove the Gilbert-Varshamov bound. [10]

8. Describe briefly the different stages of decoding BCH codes. [12]

9. Prove that the dual of a cyclic code is a cyclic code. Also prove that, in general the dual of a BCH code is not a BCH code. [10]

10. Let C be an $[n, k, d]$ code over $GF(q)$. Prove that the following statements are equivalent.

    (a) C is MDS.

    (b) Every k columns of a generator matrix G are linearly independent.

    (c) Every $n - k$ columns of a parity check matrix H are linearly independent. [12]

Indian Statistical Institute

Advanced Image Processing

M.Tech.(CS)-II Year, 2011-2012

Full marks: 100 Time: 3 Hours

Date: 24.04.2012

Answer any **ten** questions. All questions carry equal marks.

1. Define 'morphological filter'. Prove that morphological opening operation is a filter. [2+8=10]

2. How is a gray level image decomposed using multi-scale morphological filters? Describe briefly a contrast enhancement algorithm through image decomposition. [5+5=10]

3. Derive the expression of Wiener filter for image restoration using minimum mean-square estimation approach. [10]

4. (a) Define: (i) path, (ii) connected component, and (iii) genus with respect to a binary image.

   (b) Describe a two-pass algorithm for component labelling. [(3x2)+4=10]

5. State any two necessary properties of watermarking algorithms. Describe Cox's spread-spectrum watermarking algorithm. [2+8=10]

6. Describe briefly Meyer's watershed segmentation algorithm. What is meant by over-segmentation problem and suggest a way to overcome this problem? [8+2=10]

7. What is the difference between tree wavelet expansion and packet wavelet analysis of an image? Describe the fast inverse wavelet transform. [4+6=10]

8. (a) Calculate the sequency of each column of Hadamard matrix of order 8.

   (b) Describe the Hotelling transform and show that it is optimal in the least-square-error sense. [3+(5+2)=10]

9. (a) Define: (i) principal axis of an image and (ii) bi-linear interpolation.

(b) If $\overline{m}_{ij}$ denotes the (i,j)th central moment of an image f, and θ represents the slope of the principal axis, then prove that $\tan 2\theta = \frac{2\overline{m}_{11}}{\overline{m}_{20}-\overline{m}_{02}}$.

[(2+2)+6=10]

10. Consider the following block of gray levels:

0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

Construct the gray level co-occurrence matrices for angle $\theta = 0°, 45°, 90°, 135°$ considering unit pixel distance and compute the angular second moment for each case. [(4x2)+2=10]

11. Compute the time dispersion and spectral bandwidth of the following Gaussian signal: $f(t) = e^{-\frac{t^2}{2\sigma^2}}$. Prove that the signal $f(t)$ achieves the minimum of the uncertainty inequality. You may use the following two results:

$$(i) \int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}; \quad \text{and} \quad (ii) \int_{-\infty}^{\infty} x^2 e^{-x^2} dx = \frac{\sqrt{\pi}}{2}.$$

[(4+4)+2=10]

12. Consider the following digital signal: [1, 2, 4, 1, -1, -2, -1, 1]. Construct the tree wavelet expansion of this signal using the following scaling filter: [0.4830, 0.8365, 0.2241, -0.1294]. [10]

INDIAN STATISTICAL INSTITUTE
Second Semester Examination: 2011-2012
M.Tech (CS) II Year
Advanced Database Theory and Applications

Date: 27.04.2012 **Max. Marks – 50** **Time – 2 Hours.**

1. An organization is working on some construction projects in different parts of the country. It maintains a database of its employees and the projects under execution. The entire database is maintained at the head-quarter of the organization at Kolkata. Appropriate fragments are maintained at different worksites. The present worksites are Delhi, Mumbai and Chennai. Delhi site has only Civil Engineers, Mumbai site has only Electrical Engineers and Chennai site has both Civil and Electrical Engineers. Relevant relations of the database schema are shown below:

   Employee (e_no, e_name, e_type, salary)
   Technical (e_no, specialization, p_no)
   Project (p_no, p_name, budget, st_date, expected_completion_date, e_no, location)

   Usually following two monthly reports are generated at each worksite and sent to the head-quarter.

   1) List of the engineers working at each site along with their salary and specialization, so that their salary can be transferred to the appropriate bank accounts at their respective worksites.
   2) Progress report about the projects getting executed at each worksite.

   Indicate the horizontal/vertical fragments of the above relational schema that need to be maintained at each worksite, in order to generate the above reports. Following assumptions may be made:
   a) Only relevant attributes required to generate the above reports are maintained at each worksite.
   b) Attribute 'e_type' in relation Employee can either be technical or administrative and no information about the administrative employees are maintained at the worksites.
   c) Attribute 'specialization' in relation Technical indicates type of engineering degree a technical employee may have.

   (13)

2. Two sets of relational schemas are given below:

Set-1	Set-2
R_{11} (a, b, c, d, e)	R_{21} (p, q, r, a)
R_{12} (x, y, z)	R_{22} (m, n, x)

   Further information about the attribute relationships in the two schemas are:
   1) The underlined attributes are the primary keys.
   2) Relations R_{11} and R_{12} in Set-1 are similar to the relations R_{21} and R_{22} respectively in Set-2.
   3) Attribute 'a' in R_{11} and R_{21} are synonyms but attribute 'x' in R_{12} and R_{22} are homonyms.
   4) All other attributes are different from each other.

   Using the above information, generate a global conceptual schema.

   (12)

3. There are three sites S_1, S_2 and S_3. S_1 and S_2 maintain the data for relations R_1 and R_2 respectively. Additionally, S_1 knows the schema of R_2 and S_2 knows the schema of R_1 without having the data of these relations. S3 has no data or schema related information. Relations R_1

and R_2 are to be joined and the result is to be sent to S_3. Indicate a method of achieving such join without sending the entire R_1 or R_2 relation to any of the sites who are not maintaining them. Partial transfer of data is however allowed.

(10)

4. Let A be an object belonging to the security class X and B be a subject belonging to the security class Y. Explain in each of the following cases whether B will be permitted to execute the desired operations on A when the concerned DBMS implements both mandatory and discretionary access control mechanisms. The mandatory system follows the Bell-Lapadula model. Discretionary system can be used only if both the subject and object belong to the same security class.

   i) The owner of A has granted explicit read and write privileges to B and B wants to execute these operations when X>Y.

   ii) Under the same set of discretionary access rights, B wants to read and write on A when X=Y.

   iii) B wants to read and write on A when X<Y, but the owner of A has not granted any privilege to B. B, however, has received the required privileges from another subject C who in-turn has received them from the owner of A with grant option. C belongs to security class X.

(15)

-x-

INDIAN STATISTICAL INSTITUTE

Second Semester Examination: 2011 – 2012

Course Name: M. Tech (CS) - II

Subject Name: Computer Vision

Date: 8.5.12 Maximum Marks: 100 Duration: 3hr. +10min.

Note: Attempt questions for 100 marks of your choice.

Q1.(i) Show that when an object undergoes rigid body motion with translational and rotational velocity $t = (U,V,W)^T$ and $\omega = (A,B,C)^T$ respectively, the corresponding optical flow should have its translational and rotational components (u_t, v_t) and (u_r, v_r) as given below.

$$u_t = \frac{-U + xW}{Z}, \qquad u_r = Axy - B(x^2 + 1) + Cy$$

$$v_t = \frac{-V + yW}{Z}, \qquad v_r = A(y^2 + 1) - Bxy - Cx.$$

(x,y) are the coordinates of an image point and (X,Y,Z) are the coordinates of the corresponding object point on the surface of the object in the viewer centered coordinate system. [10]

(ii) For such an object as in (i) can you determine its shape? Justify your answer through the brightness change constraint equation. [2+8]

Q2. (i) What do you mean by Gaussian curvature? Show that the Gaussian curvature of a parametric surface patch in terms of its normal vector $\vec{n}$ and its derivatives $\vec{n}_u, \vec{n}_v$ is

$$K = \frac{[\vec{n}\,\vec{n}_u\,\vec{n}_v]}{|\vec{n}|^4},$$

where [ ] indicates the vector triple product and u, v are the parameters to specify points on the surface. [5+10]

(ii) If $\vec{r} = (x,y,z)^T$ is a vector to a point on the surface then prove that K can also be expressed in the following form

$$K = \frac{[\vec{r}_u\,\vec{r}_v\,\vec{r}_{uu}][\vec{r}_u\,\vec{r}_v\,\vec{r}_{vv}] - [\vec{r}_u\,\vec{r}_v\,\vec{r}_{uv}]^2}{|\vec{r}_u \times \vec{r}_v|^4}.$$ [10]

Q3. (i) From the disparity between the two images in stereo vision, justify that the distance to near objects can be measured with greater accuracy than that for far away objects. What is the epipolar line and what is its importance? [6+2+2]

(ii) Using graylevel matching in stereo pair of images how do you find depth of an object? [10]

Q4. (i) Write true or false for the following statements.

(1) Point spread function provides information about the relative merit of cameras.

(2) Laplacian is a rotationally symmetric operator but this is not true for biharmonic operator.

(3) Gaussian image is defined on a unit sphere.

(4) Extended Gaussian image is unique for a convex polyhedron.

(5) For small surface inclinations, determinant of Hessian is the Gaussian curvature.

(6) In stereographic projection the equator projects to a circle of radius 2.

(7) Conjugate points in stereo vision do not have any role in depth computation.

(8) Image irradiance is proportional to scene radiance.

(9) BRDF is constant for a Lambertian surface.

(10) In gnomonic projection all the points in the southern hemisphere can be mapped onto a plane, tangent to the north pole of the sphere. [10]

(ii) For a Lambertian object, if $\hat{s}_1, \hat{s}_2, \hat{s}_3$ are three unit vectors in the directions of three source positions, show that

$$\rho\hat{n} = \frac{1}{[\hat{s}_1 \hat{s}_2 \hat{s}_3]}(E_1(\hat{s}_2 \times \hat{s}_3) + E_2(\hat{s}_3 \times \hat{s}_1) + E_3(\hat{s}_1 \times \hat{s}_2)),$$

where $\hat{n}$ is the unit surface normal and ρ is the albedo. The brightness vector is given by $E=(E_1, E_2, E_3)^T$. [10]

Q5. (i) For a surface of low inclinations, the Gaussian curvature and the mean curvature are given by $k_1 k_2 \approx z_{xx} z_{yy} - z_{xy} z_{yx}$ and $\frac{1}{2}(k_1 + k_2) \approx \frac{1}{2}(z_{xx} + z_{yy})$.

Express these results in terms of the Hessian matrix H. [10]

(ii) Find z (x, y) so as to minimize the error

$\iint ((z_x - p)^2 + (z_y - q)^2) dxdy$, where p and q are the given estimates of the gradient in the direction of x and y respectively, while z_x, z_y are the partial derivatives of the best fit surface. [5]

Q6. (i) Suppose we use the integral

$e_s = \iint ((\nabla^2 u)^2 + (\nabla^2 v)^2) dxdy$ instead of

$e_s = \iint ((u_x^2 + u_y^2) + (v_x^2 + v_y^2)) dxdy$ as the measure of departure from

smoothness and the integral

$e_c = \iint (E_x u + E_y v + E_t)^2 dxdy$ as the error in optical flow constraint equation. To minimize $e_s + \lambda e_c$ what will be the Euler equations? λ is a parameter that weights the errors in the image motion equation relative to the departure from smoothness. Mention the necessary boundary conditions. [10+5]

(ii) Give reasons why corner points in the graylevel image are chosen as initial points to interpolate optical flow values. [5]

Q7. A surface z(x, y) with continuous second partial derivatives has to satisfy an integrability constraint, i. e., $\frac{\partial^2 z}{\partial x \partial y} = \frac{\partial^2 z}{\partial y \partial x}$ or $p_y = q_x$. Suppose you wish to minimize the brightness error, $\iint (E(x,y) - R(p,q))^2 dxdy$ by suitable choice of the two functions *p(x, y)* and *q(x, y)*, subject to the integrability constraint $p_y = q_x$.

(a) Show that the appropriate Euler equations are $(E(x,y) - R(p,q))R_q = -\lambda_x$ and $(E(x,y) - R(p,q))R_p = +\lambda_y$ where $\lambda(x, y)$ is a Lagrange function and

$$R_p = \frac{\partial R}{\partial p}, R_q = \frac{\partial R}{\partial q}, \lambda_x = \frac{\partial \lambda}{\partial x}, \lambda_y = \frac{\partial \lambda}{\partial y}.$$ [16]

(b) Conclude that the desired functions *p(x, y)* and *q(x, y)* must satisfy the equation

$$((E-R)R_{pp} - R_p^2)p_x + ((E-R)R_{pq} - R_p R_q)(p_y + q_x) + ((E-R)R_{qq} - R_q^2)q_y = (E_x R_p + E_y R_q).$$

[16]

(c) Show that you end up with the same Euler equation if you try to minimize $\iint (E(x,y) - R(z_x, z_y))^2 dxdy$ by suitable choice of *z(x, y)*. [18]

INDIAN STATISTICAL INSTITUTE

Semestral Examination: (2011-2012)

M.Tech. (CS) II Year

Parallel Processing: Architectures and Algorithms

Date: 04/05/2012 Full Marks: 100 Duration: 3 hrs

NOTE: A student may answer all questions but maximum marks attainable is 100.

1. a) Define the topology of a cube-connected cycle (CCC), consisting of $n.2^n$ nodes. Derive the expression for the diameter of the network.
b) In a regular graph with diameter k and node degree d, derive an expression for the *Moore Bound* on the number of nodes N. What is *Moore graph*?
c) Consider an N-node linear array in which each node contains a packet. Assume that the destination of each packet is distinct, and the links are bi-directional. Describe a deadlock-free algorithm for delivering all the packets in at most $(N-1)$ steps. Find out the maximum buffer size required at each node. Prove that your algorithm is deadlock-free.

[(3+3)+(4+2)+(4+2+2)= 20]

2. a) Show that the Cayley graph defined by the set of generators:
{1 2 3 4 6 5, 1 2 4 3 5 6, 2 1 3 4 5 6} is a hypercube of degree 3.
b) Define a 4-pancake graph mentioning the relevant generators. Prove that $(2n-3)$ is an upper bound on the diameter of an n-pancake graph.

[5+(5+8)=18]

3. a) Draw a sorting network for sorting 8 elements by *Batcher's odd-even merge sort* technique.
b) Describe *Batcher's odd-even merge sort* algorithm for sorting a sequence of n elements. Prove its correctness by using *0-1 principle*.
c) Analyze the *speed-up* of the system.

[6+(4+ 6)+6=22]

4. Consider a CRCW SM SIMD computer with n^3 processors. Simultaneous write operations to the same memory location are allowed when all the values to be written are the same. Describe an algorithm to compute the connectivity matrix of an n-node graph on this computer in $O(\log n)$ time. Comment on the optimality of the above algorithm.

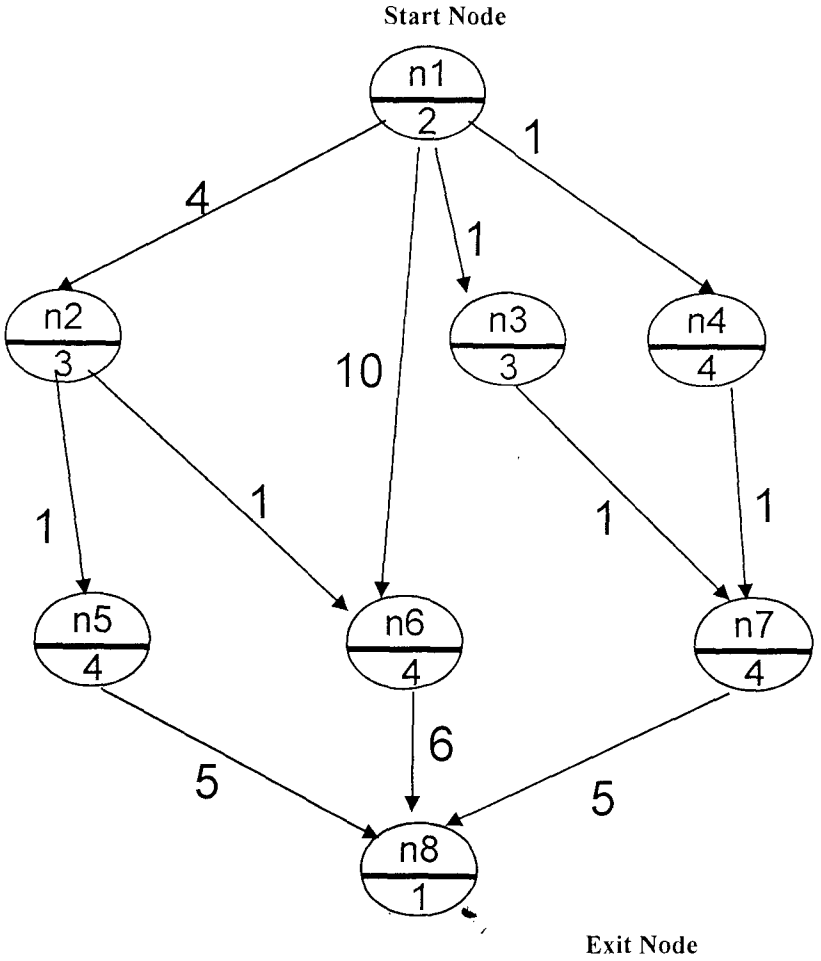
[15+5 = 20]

5. *Gauss-Jordan elimination* is one method for solving the system of n linear equations $\mathbf{AX} = \mathbf{B}$, where $\mathbf{A}$ is the $n \times n$ coefficient matrix, $\mathbf{B}$ is an $n \times 1$ vector and $\mathbf{X} = (x_1, x_2, \ldots, x_n)^T$, $x_1, x_2, \ldots, x_n$ are n unknown variables. Design an $O(n)$ time parallel version for *Gauss-Jordan elimination* algorithm on a SM SIMD Computer. What model of SM is needed and why? Find out the AT cost of the algorithm, where A is proportional to the number of processors and T is the time required.

[12+3+5=20]

P.T.O

6. A parallel program is represented by a directed acyclic graph (DAG) G(V, E) shown in the Fig. below. The nodes represent tasks, the labels show the task-id and the computation time, the edges show the precedence/ communication relations with labels representing the communication time. A task can start only when all its parent nodes have finished and communicated data to it. Initially only the start node is in the ready node list.
Given a set of 3 processors, compute the earliest start time (EST) for each ready node on each processor and select the node-processor pair with minimum EST. In case of tie, select the node with highest static *b-level* (sum of all node weights in the longest path from the exit node). Find out the complete schedule for the tasks on the processors, and calculate the time of completion.



[20]

INDIAN STATISTICAL INSTITUTE

End Semestral Examination

M. Tech (CS) - II Year, 2011-2012 (Semester - IV)

Topics in Algorithms and Complexity

Date : 08.05.2012 | Maximum Marks : 100 | Duration : 3.5 hours

---

Note: Answer as much as you can, but the maximum you can score is 100.

This is a two-page question paper with 10 questions.

(Q1) (i) Let X be a random variable which is the sum of n independent 0-1 valued random variables, i.e., $X = \sum_i X_i$, where each X_i takes the value 1 with probability p_i, $0 < p_i < 1$, and the value 0, otherwise. Let $\mu = E[X]$, where $E[X]$ denotes the expectation of X. Show that for any δ, $0 < \delta < 1$, $\Pr[X < (1-\delta)\mu] < e^{-\frac{\delta^2 \mu}{2}}$. [10]

(ii) Suppose that $p_i = 3/4$ in the above setting. How large must n be so that $\Pr[X < (1-\delta)\mu] < n^{-5}$? [4]

(iii) What can you infer about δ if we want $\Pr[X < (1-\delta)\mu] < e^{-5 \ln n}$? [2]

(Q2) Let S be a set drawn from a totally ordered universe. A leveling with r levels of an ordered set S is a sequence of nested subsets, called levels, $L_0 \supseteq L_1 \supseteq \cdots \supseteq L_{r-1} \supseteq L_r$, such that $L_0 = \emptyset$ and $L_1 = S$.

Show that the number of levels r in a random leveling of a set S of size n has expected value $E[r] = O(\log n)$. [10]

(Q3) Consider throwing n balls uniformly and independently into n bins. Let the random variable X_1 denote the number of balls that fall into the first bin. Determine a quantity ω such that $\Pr[X_1 > \omega] \leq \frac{1}{n^2}$. [5]

(Q4) Let a set of n variables be $x_1, x_2, \ldots, x_n$, each of which can take one of the values in the set $\{0, 1\}$. We are given a set of k equations. The i^{th} equation has the form $x_j + x_{j'} \bmod 2 = b_i$, for some choice of two distinct variables $x_j, x_{j'}$ and for some value b_i that is either 0 or 1. Each equation specifies whether the sum of a pair of variables is even or odd.

Consider the problem of finding an assignment of values to the variables such that the number of equations that are satisfied, is maximum.

(i) Let c^* denote the maximum possible number of equations that can be satisfied by an assignment of values to the variables. Design and analyze a randomized algorithm which satisfies at least $\frac{1}{2}c^*$ number of equations in the expected case. [4]

(ii) Next, modify the k equations mentioned above as follows. The i^{th} equation has the modulo 2 sum of *an arbitrary subset of the variables* equal to a certain value b_i. Design and analyze a randomized algorithm which satisfies at least $\frac{1}{2}c^*$ number of equations in the expected case, where c^* is same as in (i) above. [6]

(Q5) Design and analyze an efficient randomized algorithm for constructing the convex hull of a set of n points. [15]

(Q6) (i) Verify whether $[6, 5, 0, 3, 1, 2]^T$ is entangled. [4]

(ii) Argue whether the tensor product of two matrices $A_{r_1 \times c_1}$ and $B_{r_2 \times c_2}$ is commutative. What is the time complexity to compute this tensor product in the classical way? [4 + 2]

(Q7) (i) Represent each of the following 2-input operations by an appropriate operator matrix: (a) AND, (b) $SWAP$. Mention the representations for the input values. [2 + 3]

(ii) Construct the 2-input $NAND$ operation with one Toffoli gate. How many Toffoli gates are needed for constructing the 2-input OR operation? [3 + 3]

(iii) Present the operation of the Fredkin gate. Is it reversible? [2 + 2]

(Q8) Given a n-vertex graph G, the *triangle problem* is to decide whether G contains a triangle. The graph is specified by a black box that, for any pair of vertices of G, returns a bit indicating whether those vertices are adjacent in G.

(i) What is the query complexity (worst case number of queries made) in classical computation for this problem? [2]

(ii) Using Grover's search algorithm, sketch how to solve the triangle problem. [5]

(iii) What is the time complexity of your algorithm? [3]

(Q9) (i) Present the steps of Shor's algorithm for factoring a given positive integer N, and the necessary quantum operators. [10]

(ii) Find an expression for its speedup over a classical algorithm for this problem. [3]

(iii) How many 2-input Hadamard gates are needed? [2]

(Q10) (i) Prove that the problem of Graph Non-Isomorphism (GNI) for a given pair of n-vertex graphs G_1 and G_2 is in the class IP. Next, give a $AM[2]$ proof protocol for GNI. [5 + 4]

(ii) Discuss the relations of the class BPP with (a) $P_{/poly}$, and (b) BQP. [3 + 3]