"Students governed by the old ordinances will be examined out of 75 marks and their obtained marks shall be proportionately raised."

Note: Answer five questions by selecting at least two questions from each section. The normal distribution table is provided with the question paper.

SECTION – A

1. (a) Use the method of least squares to fit the curve \( y = a \sqrt{x} + b/\sqrt{x} \) for the following data:

\[
\begin{array}{cccccc}
 x & 0.1 & 0.2 & 0.5 & 1 & 2 \\
 y & 16 & 14 & 11 & 6 & 3 \\
\end{array}
\]

(b) Approximate \( x^3 \) by a polynomial of the form \( a + bx + cx^2 \) on [0,1] in the least square sense.

2. (a) Find the values of \( y(0.2) \) and \( y(0.4) \) by using Modified – Euler’s method for the initial value problem:

\[
\frac{d^2y}{dx^2} = x^2 + 3y, \quad y(0) = 1, y'(0) = 1
\]

(chose \( h = 0.2 \))

OR

(a') Find the value of \( y(0.2) \) for the initial value problem given in question 2(a) by using Runge-Kutta method of order 4 by choosing \( h = 0.2 \).

(b) Solve the boundary value problem:

\[
y'' - xy = 0, \quad y(0) + y'(0) 1, \quad y(1) = 1 \quad \text{with} \quad h = \frac{1}{3} \quad \text{using finite – difference method.}
\]

3. (a) Solve the equation

\[
\frac{\partial^2u}{\partial x^2} + \frac{\partial^2u}{\partial y^2} = xy,
\]

\( u(x, y) = x^2 + y^2 \) on the boundary of the domain. The domain is a square with corners at (0,0) and (4,4). Choose \( h = 1 \).

(b) Solve the equation :

\[
\frac{\partial u}{\partial t} = 2 \frac{\partial^2u}{\partial x^2}
\]

with \( u(0,t) = 0 = u(4,t) = 0 \) and \( u(x,0) = x (4-x), \quad 0 \leq x \leq 4 \) by Crank – Nicholson method with \( h = 1 \). (Integrate upto two time levels).
4. (a) Solve numerically the wave equation
\[ \frac{\partial^2 u}{\partial t^2} = 4 \frac{\partial^2 u}{\partial x^2}, \quad 0 \leq x \leq 4, \quad t \geq 0 \]
with the boundary conditions \( u(0, t) = 0, \quad u(4, t) = 0, \quad t \geq 0 \) and the initial condition \( \frac{\partial u}{\partial t}(x, 0) = 0 \) with \( h = 1 \), up to two time levels.

(b) Apply three iterations of iterative method of find an approximate solution of the integral equation
\[ y(x) = 1 + x + \int_0^x (x - s) y(s) \, ds, \quad y(0) = 1. \]
If possible, find the exact solution.

SECTION – B

5. (a) The probability distribution function \( f(x, y) \) of two-dimensional random variable \( (X, Y) \) is given by
\[ f(x, y) = K \left( 2x + 3y \right) \]
for \( x = 0, 1, 2 \), and \( y = 1, 2, 3 \).
Find the value of the constant \( K \), marginal probabilities, \( P(X+Y = 3) \), \( P(X+Y \leq 5|y = 3) \).

(b) The joint probability distribution function of two-dimensional random variable \( (X, Y) \) is given by
\[ f(x, y) = xy^2 + \frac{1}{8} x^2, \]
\[ 0 \leq x \leq 2, \quad 0 \leq y \leq 1. \]
Calculate \( P(X > 1), P(Y < 2), \)
\[ P(X + Y \geq 1), \quad P(X < 1|Y > 2). \]

6. The pdf \( f(x, y) \) of random variable \( (X, Y) \) is given by \( f(x,y) = 8 \, xy, \quad (12) \)
\[ 0 < x < 1, \quad 0 < y < x. \]

(a) Show that the pdf of \( X \) is \( g(x) = 4x^3, \quad 0 \leq x \leq 1. \)
(b) Show that the pdf of \( Y \) is \( h(y) = 4y \, (1-y^2), \quad 0 \leq y \leq 1. \)
(c) Calculate \( E(X), E(Y) \) and \( E(XY) \)
(d) Calculate \( V(X), V(Y) \)
(e) Calculate the correlation coefficient \( \rho \).

7. (a) If \( X \) and \( Y \) are independent random variables with pdf \( g(x) = e^x, \quad x \geq 0 \) and \( h(y) = 3e^{-3y}, \quad y \geq 0 \) respectively, then find that pdf of the random variable \( Z = X/Y \).

(b) \( X \) is exponentially distributed with parameter \( \alpha = 0.001 \). A sample of size 100, \( \{x_1, x_2, \ldots, x_{100}\} \) is chosen at random. Find \( P(900 < \bar{x} < 1050) \) and \( P(M > 6000) \) where \( M \) is the maximum of \( x_1, x_2, \ldots, x_{100}. \)

\[ \text{Contd……3} \]
8. (a) In a telegraph signal problem, the stochastic process \( x(t) \) is defined as
\[
\begin{align*}
x(t) = 1, & \text{ if the total number of points in } (0,t) \text{ is even} \\
= -1, & \text{ otherwise.}
\end{align*}
\]
The probability of getting \( K \) points in \( (0,t) \) is given by
\[
p(K) = e^{-\lambda t} \frac{(\lambda t)^K}{K!},
\]
for \( K = 0, 1, 2, \ldots \). Find \( P(x(t) = 1), \ P(x(t) = -1), \ E\{x(t)\} \) and the autocorrelation coefficient \( R(t_1, t_2) \).

(b) Suppose that \( X \) has distribution \( N(\mu, \sigma^2) \), where \( \mu = 100 \) and \( \sigma = 2 \). To test
\[
H_0 : \mu = 100 \text{ versus} \\
H_1 : \mu < 100,
\]
a sample of size 400 is chosen. If the significance level of the test is to be \( \alpha = 0.01 \), find the value of \( C \).

*Note: The standard Normal Distribution function table is enclosed.*

*****
<table>
<thead>
<tr>
<th>$z$</th>
<th>$(z &lt; -2)d = \frac{np}{z^2-e^{-2}} \frac{x^2}{1} \int (z) = (2)\Phi$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Table I (Continued)</strong></td>
</tr>
</tbody>
</table>

**Table I. Values of the Standard Normal Distribution Function**
2011-2012
M. TECH. (I SEMESTER) EXAMINATION
(COMPUTER SCIENCE AND ENGINEERING)
SOFTWARE ENGINEERING-I
(CO-601)

Maximum Marks: 60
Duration: Three Hours

"Students governed by the old ordinances will be examined out of 75 marks and their obtained marks shall be proportionately raised."

Note: Answer all questions.
Make appropriate assumptions, if required.
Symbols and abbreviations have their usual meanings.

1. (a) (i) What are the factors that have contributed to the present software crisis? Suggest at least two possible solutions to the present software crisis. (5+5)

(ii) Identify at least two activities carried out during each phase of a spiral model.

OR

(a') (i) Explain the problems that might be faced by an organization if it does not follow any software life cycle model. (5+5)

(ii) Identify the definite stages through which a software product undergoes during its life time.

(b) Which software development process model could you be selecting for high risk projects and why? (5)

2. (a) You have been appointed a software project manager for a company that services the genetic engineering world. Your job is to manage the development of a new software product that will accelerate the pace of gene typing. The work is R & D oriented, but the goal is to produce a product within the next year. What team structure would you choose and why? (5)

(b) Write short note on any TWO of following - (5+5)
   (i) Process vs Product metrics
   (ii) Defect removal efficiency
   (iii) Empirical estimation structures

3. (a) Discuss the baselines concept of software configuration management. (5)

(b) Explain with example the control hierarchy framework of software design concepts.

OR

(b') What are the design heuristics for effective modeling? (5)
4. (a) (i) Differentiate between verification and validation in the context of software testing.

(ii) Discuss the criteria for completion of testing.

OR

(a') (i) Explain why boundary value analysis is so important for the design of black-box test suit for a problem.

(ii) Explain with examples methods of condition and loop testing.

(b) Draw the control flow graph for the following function named bin_ser.

From the control flow graph, determines its Cyclomatic complexity.

```c
/* num is the number the function searches in presorted integer array arr */

int bin_ser (int num)
{
    int min, max;
    min=0; max = 100;
    while (min != max)
    {
        if ( arr [(min + max)/2] > num )
            max = (min + max)/2;
        else if ( arr [(min + max)/2] < num )
            max = (min + max)/2;
        else return ( (min + max)/2 );
    }

    return (-1);
}
```
2011-2012
M.TECH. (I Semester) Examination
WEB MINING AND SEARCHING
(CO-616)

Maximum Marks: 60
Duration: Three hours

"Students governed by the old ordinances will be examined out of 75 marks and their obtained marks shall be proportionately raised."

Answer all questions.

1.(a) Differentiate between classification and clustering. Discuss different parameters used for evaluation of a classification technique.

1.(b) What are the benefits of decision tree induction in data mining? Write down the basic decision tree induction algorithm. When we stop growing a decision tree? Also, mention some of the enhancements in basic decision tree induction.

OR

1.(b') Differentiate between splitting attribute and splitting criterion. Name and discuss the two indices that can be used to select splitting attribute. Compute the values of these indices for the attribute Age for the data given in Table 1 for a 2-class system.

2.(a) What do you mean by overfitting? What are the disadvantages of overfitting? Name and discuss the two approaches to avoid overfitting in decision tree based classification algorithm.

2.(b) Describe the major clustering approaches and for each clustering approach, name at least two clustering algorithms, which are based on that approach.

2.(c) Discuss in detail the k-nearest neighbor algorithm for classification.

OR

2'(a) Differentiate between lazy learning and eager learning. What is instance based learning? Name few instance based learning algorithms.

2'(b) What is outlier discovery? Discuss different applications of outlier discovery.

2'(c) Discuss different approaches used for outlier discovery.
3.(a) What do you mean by sequential pattern mining? Name the two approaches used for sequential pattern mining. Describe different applications of sequential pattern mining.

3.(b) Write down the SPADE algorithm and PrefixSpan algorithm for sequential pattern mining. Also, discuss the advantages and limitations of both the algorithms.

OR

3'.(a) Write technical notes on episode pattern mining and periodicity analysis.

3'.(b) Why we need to parallelize data mining algorithms? Discuss different types of parallel association rule mining algorithms and illustrate their operation with the help of a suitable example.

4.(a) What do you mean by Web spiders? Describe different functions of Web Spiders. What is an intelligent Web spider?

4.(b) Define precision and recall. What is the significance of precision at k in Web based Information retrieval? Can actual recall be computed for Web based information retrieval? If not, how we can realize relative recall?

4.(c) What do you mean by Web usage mining? List different Web log analyzer programs and discuss different parts of a Web server log.

OR

4.(c') Draw the standard architecture of a Web search engine and discuss with the help of a proper example, how inverted files are created.

5.(a) Discuss the notion of hub and authority. What are the problems associated with the Web's hyperlink structure and what are the two difficulties arising due to of ignorance of textual contexts in Web structure mining?

5.(b) Write technical notes on Cosine similarity measure and Boolean similarity measures.

5.(c) Discuss the advantage and disadvantages of description based image retrieval. Can we use standard information retrieval techniques for description based image retrieval? If yes, describe the way to use the same.
### Table 1

<table>
<thead>
<tr>
<th>Age</th>
<th>$p_i = \text{Examples from } P \text{ class}$</th>
<th>$n_i = \text{Examples from } N \text{ class}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;=30</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>30-40</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>&gt;40</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>
"Students governed by the old ordinances will be examined out of 75 marks and their obtained marks shall be proportionately raised."

- Attempt All questions.
- Symbols and notation used have their standard meanings.
- Assume suitable data if required.

1(a) Differentiate between Mealy and Moore machine

(b) Generate the state diagram and state table of a 4-bit even counter

(c) Why it is necessary to satisfy the Setup and Hold timing of the signal with the clock in a sequential machine?

OR

1'(a) A state diagram of a sequential system is given in figure (1). Use state reduction technique to eliminate the redundant state and assign suitable codes to the states. What type of state machine is this state diagram represents?

```
Figure (1)
```

(b) What is a “State” in a state machine? Why the state reduction is done?

2(a) Draw the Mealy state diagram and state table for a serial even parity checker. The circuit receives a word of 4-bit serially on its single input ‘X’ and produce an even parity bit after the fourth bit is received. The single output ‘Z’ remain ‘0’ except when the final (fourth) bit is received and the total number of 1’s in the word is odd. The machine is restarting (resetting) machine.

(b) What are the different modes of a port in VHDL? Describe very briefly. Write the VHDL statement, for accessing a component declared in a package.

OR

2'(a) Write the VHDL codes for 4:1 MUX using Data Flow Model.

(b) Consider the sequential system shown in figure 2.
i. Explain whether this sequential system is Mealy or Moore system.

ii. Generate the state table and state diagram of the system and identify its output and function.

3(a) Use selected concurrent statements to write the VHDL codes for circuit that implement the sum of the product (SOP) function \( F = \sum(1, 3, 5, 6, 7, 8, 9, 13, 14, 15) \)

OR

(a') Give the VHDL codes that describe the sequential behavior of JK flip flop using "BLOCK and GUARDED statements"

(b) Where, the generate statements are used in the system modeling? What are the different kinds of generate statements? Give the syntax of these statements.

4(a) Describe different types of wait statements used in the process statement of algorithmic model.

(b) Write down the VHDL codes for the circuit shown in figure 2.

5(a) Write a function in VHDL that compute sum and carry of two number

(b) Write down the VHDL codes for the machine which has two inputs \( X_1, X_2 \) and one output \( Z \). the output of the machine is determine by the following.

i. \( Z \) does not change when \( "X_1X_2"="00" \)

ii. \( Z='1' \) if \( "X_1X_2"="01" \)

iii. \( Z='0' \) if \( "X_1X_2"="10" \)

iv. \( Z \) changes its value if \( "X_1X_2"="11" \)