Paper / Subject Code: ME134 / Elective-II Data Mining

Total No. of Printed Pages:02

M.E (Computer Science & Engineering) (Sem-I) (Revised Course)
EXAMINATION Nov/Dec 2019
Elective-II
Data Mining

[Duration : Three Hours] **[Total Marks:100]**

Instructions:
1) Answer **five** full questions.
2) Make **suitable** assumptions **wherever** necessary.

Q.1 a) Explain the characteristics that have to be applied to data sets in order to have a significant impact on the data mining techniques. Explain in detail with examples the various data sets. **(10)**

b) Consider the following data set for a binary class problem. **(10)**

A	B	C	Class Label
T	F	1.0	+
T	T	6.0	+
T	T	5.0	+
T	F	4.0	-
T	T	7.0	+
F	F	3.0	-
F	F	8.0	-
F	F	7.0	-
T	T	5.0	-
T	F	2.0	-

i) What is the entropy of the above data set with respect to the positive class?
ii) For C, which is a continuous attribute, compute the information gain for split 1 and split 2.
iii) Calculate the information gain when splitting on A and B. Which attribute would the decision tree induction algorithm choose?
iv) Calculate the gain in the Gini index when splitting on A and B. Which attribute would the decision tree induction algorithm choose?
v) What is the best split (between A and B) according to the classification error rate?

Q.2 a) Explain the methods that are commonly used to evaluate the performance of a classifier. **(10)**

b) Explain the characteristics of a rule – based classifier. **(03)**

c) Explain the following w.r.t. Indirect Method for Rule Extraction **(04)**
i) Rule Generation
ii) Rule ordering

d) Explain the k- Nearest Neighbor Classification Algorithm. **(03)**

Q.3

a) Explain the alternative methods for generating frequent itemsets. **(06)**

b) Explain how the association patterns are evaluated. **(04)**

c) Consider a training set that contains 29 positive examples and 21 negative examples. For each of the following candidate rules, **(05)**

R1: A → +(covers 12 positives and 3 negative examples),

R2: B → +(covers 7 positives and 3 negative examples),

R3: C → +(covers 8 positives and 4 negative examples).

Determine which is the best and worst candidate rule according to

i. The Laplace measure
ii. The Likelihood ratio statistics
iii. m- estimate (with k=2 and p_+=0.58)
iv. FOIL's information gain

d) Explain the various alternative counting schemes for a sequential pattern mining. **(05)**

Q.4

a) Explain the issues with the basic K- means clustering algorithm. **(06)**

b) The original association rule mining formulation uses the support and confidence measures to prune uninteresting rules. Draw a contingency table for each of the following rule using the transactions shown in the table below. Compute and rank the rules in decreasing order according to the following measure. **(08)**

i) Interest ii) IS

Rules:
{b} → {c}, {c} → {a}, {e} → {c}, {a} → {d}

TID	Items Bought
1	{a, b, d, e}
2	{b, c, d}
3	{a, b, d, e}
4	{a, c, d, e}
5	{b, c, d, e}
6	{b, d, e}
7	{c, d}
8	{a, b, c}
9	{a, d, e}
10	{b, d}

c) Explain briefly the supervised measure for cluster validation. **(06)**

Q.5

a) Explain the collaborative recommender system. State the advantages of a recommender system. **(06)**

b) Explain the HITS (Hyperlink – Induced Topic Search) algorithm to find authoritative pages in mining the World Wide Web. **(06)**

c) Explain the following trends in data mining **(04)**

i) Biological data mining
ii) Web mining
iii) Data mining and software engineering
iv) Distributed data mining

d) Explain the different approaches proposed for similarity- based search in multimedia data. **(04)**

c) Given that a mail is classified as not spam, the probability of the mail actually being not spam

d) Find the probability that the mail is misclassified.

Q.3 a) Write and explain Brute-force MAP learning algorithm

b) Consider the following data

Origin	Manufacture	Color	Decade	Type	Ex. Type
Japan	Honda	Blue	1980	Economy	+ve
Japan	Toyota	Green	1970	Sports	-ve
Japan	Toyota	Blue	1990	Economy	+ve
USA	Chrysler	Red	1980	Economy	-ve
Japan	Honda	White	1980	Economy	+ve

Answer the following questions.

i) Compute a maximally specific hypothesis using FIND-S algorithm.

ii) Compute a version space containing all hypotheses consistent with the above examples using the candidate elimination algorithm.

iii) Does the order of presentation of the training examples (according to Table) to the learner affect the finally learned hypothesis?

Q.4 a) What is the Vapnik-Chervonenkis (VC) Dimension? When do we say that a set of instances (S) is shattered by hypothesis space(H)? Consider the set X of instances corresponding to points on the x, y plane and let H be the set of all linear decision surface in the plane. What will be the VC(H). 10

b) Explain the perceptron training rule? Show how an AND Boolean function can be modelled using a perceptron. Also show the corresponding weights. There are some Boolean functions that cannot be represented by a single perceptron? Which are they? And Why they cannot be represented by a single perceptron? 10

Q.5 a) Discuss the Probably Approximately Correct learning model. Define the PAC learnability. 10

b) What is the difference between the Maximum Likelihood Estimation and the Maximum-A-Posteriori Hypothesis? Using Maximum Likelihood Estimation for the parameter estimation, compute the MLE estimate for the univariate Gaussian distribution. 10

6 a) What is the naive assumption in Naive Bayes algorithm? Consider the following dataset: 12

Age	Income	Student	Credit_rating	Buys_computer
youth	High	No	Fair	No
Youth	High	No	Excellent	No
middle aged	High	No	Fair	Yes
senior	Medium	No	Fair	Yes
senior	Low	Yes	Fair	Yes
senior	Low	Yes	Excellent	No
middle aged	Low	Yes	Excellent	Yes
youth	Medium	No	Fair	No
youth	Low	Yes	Fair	Yes
senior	Medium	Yes	Fair	Yes
Youth	Medium	Yes	Excellent	Yes
middle aged	Medium	No	Excellent	Yes
middle aged	High	Yes	Fair	Yes
senior	Medium	No	Excellent	No

Buys_computer is the class.
Classify the test instance X given below using a Naive Bayes Classifier.
X=(Age=youth, income=medium, student=yes, credit_rating=fair)
What are the advantages and disadvantages of Naive Bayes classifier? In what ways can these disadvantages be handled in the Bayesian belief networks?

b) What is machine learning? Explain the steps in designing a learning system? 8

Paper / Subject Code: ME136 / Advances in Computer Architecuture

ME136

Total No. of Printed Pages:02

M.E (Computer Science & Engineering) (Sem-I) (Revised Course)
EXAMINATION Nov/Dec 2019
Advances in Computer Architecture

[Duration : Three Hours] **[Total Marks:100]**

Instructions to candidates:

1) Assume suitable data if necessary.
2) Answer any five questions
3) Draw neat diagrams if required.
4) Write question numbers legibly while answering.
5) Solve problems with appropriate assumptions if required.

Question No (1)

a) *'If two instructions are dependent, they are not parallel'*. Explain the statement. What are these dependencies? Elaborate with suitable examples. [8]

b) Suppose we have two implementations of the same instruction set architecture. Computer C1 has a clock cycle time of 250 ns and a CPI of 2.0 for some program, and computer C2 has a clock cycle time of 500 ns and a CPI of 1.2 for the same program. Which computer is faster for this program and by how much? Assume that number of instructions in the program is 1,000,000,000. [6]

c) Two code sequences have been provided for a particular machine. Based on the hardware implementation, there are three different classes of instructions: [6]
Class A, Class B, and Class C, and they require one, two, and three cycles (respectively).
The first code sequence has 30 instructions: 10 of A, 10 of B, and 10 of C.
The second sequence has 28 instructions: 15 of A, 8 of B, and 5 of C.
As a compiler designer you are needed to address the following:
i. Which sequence will be faster? How much?
ii. What is the CPI for each sequence?
iii. Does lesser number of instructions imply lessor execution time? Explain you answer.

Question No (2)

a) Explain the basic schemes for enforcing cache coherence in multiprocessor systems. Discuss snooping protocols in detail. [10]

b) Suppose we have made the following measurements: [10]

- Frequency of FP(Floating Point) operations=25%
- Average CPI of FP operations=4.0

- Average CPI of other instructions = 1.33
- Frequency of FPSQR= 2%
- CPI of FPSQR=20

Assume that two design alternatives are to decrease the CPI of FPSQR to 2 or to decrease the average CPI of all FP operations to 2.5.
Compute these two design alternatives using processor performance equations. Also state Amdahl's law.

Question No (3)

a) List the advantages of message – passing communication in a large – scale multiprocessor. [6]

b) Define a trace. How is it used in a VLIW processor? List out the advantages and disadvantages of VLIW processor. [10]

c) What does the ratio of computation to communication in parallel programs signify? [4]

Question No (4)

a) Discuss the design Challenges in Simultaneous multithreading (SMT). [08]

b) Briefly discuss the architectures of Intel i860 and SUN SPARC processors. [12]

Question No (5)

a) Explain the architecture support required for Tomasulo's approach. Describe how it helps to avoid data hazards. [10]

b) Explain how to reduce branch costs by static branch prediction techniques, with an illustrative example. [10]

Question No (6)

a) Explain guidelines and principles useful in the design and analysis of computers. [10]

b) What are the difference between scalar instructions and vector instructions? Give at least four differences. [10]

PPP-A – 32

M.E. (Computer Science and Engineering) (Semester – I)
Examination, November/December 2017
DATA MINING (Elective – II)
(Revised Course)

Duration : 3 Hours | Total Marks : 100

Instructions *: 1) Answer* ***five full*** *questions.*
2) Make ***suitable*** *assumptions* ***wherever*** *necessary.*

1. a) Draw a neat labeled diagram to show the complete data mining process and describe the same by listing at least one technique at each step. 7

b) Consider the following Mushroom data set where some of the mushrooms have been determined as poisonous and others as not. (2+4+4)

Example	Is heavy	Is smelly	Is spotted	Is smooth	Is poisonous
A	0	0	0	0	0
B	0	0	1	0	0
C	1	1	0	1	0
D	1	0	0	1	1
E	0	1	1	0	1
F	0	0	1	1	1
G	0	0	0	1	1
H	1	1	0	0	1
U	1	1	1	1	?
V	0	1	0	1	?

Answer the following questions w.r.t. Mushroom dataset :

i) What is the entropy of 'Is Poisonous' ?

ii) Write ID3 algorithm.

iii) Demonstrate ID3 algorithm to develop a decision tree and predict for the cases U and V.

c) Discuss the significance and working principle of Principal Component Analysis (PCA). 3

P.T.O.

2. a) For the following vectors, x and y, calculate the jaccard, cosine and euclidean distance measures. 6
   $x = (1, 1, 1, 1)$; $y = (2, 2, 2, 2)$.

   b) Write the algorithm and explain the K-nearest neighbour algorithm for classification. 7

   c) What is overfitting in decision trees ? Demonstrate with an example an effective solutions to overcome overfitting. 7

3. a) Given a simple transactional database X : 10

TID	Items
T01	A, B, C, D
T02	A, C, D, F
T03	C, D, E, G, A
T04	A, D, F, B
T05	B, C, G
T06	D, F, G
T07	A, B, G
T08	C, D, F, G

   Using the threshold values support = 25% and confidence = 60%, find :
   i) All frequent itemsets in database X.
   ii) Strong association rules for database X.

   b) Both k-means and k-medoids algorithms can perform effective clustering. 5
   i) Illustrate the strength and weakness of k-means in comparison with k-medoids.
   ii) Illustrate the strength and weakness of these schemes in comparison with the hierarchical clustering scheme (e.g., AGNES).

   c) Prove that in DBSCAN, the density-connectedness is an equivalence relation. 5

4. a) What is a multidimensional cube ? Give its significance. 3

   b) The age values of the data tuples are : (2+5)
   20, 20, 21, 22, 22, 25, 25, 25, 25, 13, 15, 16, 16, 19, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.
   i) What is an outlier ? Point out an outlier from the given 'age' data ?
   ii) Demonstrate a method for automatically detecting an outlier and its effectiveness.

c) Consider the following data set : 10

No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

i) Construct the Naive Bayes Classifier.

ii) Classify the test case 'Red, Domestic, SUV'.

5. a) Discuss the steps involved in construction and mining of the fp-tree. 6

b) Suggest four methods which can be used for improving the efficiency of Apriori method. 4

c) Write short notes on **(any two)** : (5+5)

i) Spatial data mining

ii) OPTICS algorithm

iii) Rule based classifier.

c) Consider the following set of frequent 3-itemset. Assume that there are only 5 items in the data set. {1, 2, 3}, {1, 2, 4}, {1, 2, 5}, {1, 3, 4}, {1, 3, 5}, {2, 3, 4}, {2, 3, 5}, {3, 4, 5}. 6

i) List all candidate 4-itemsets obtained by a candidate generation procedure using $F_{k-1} \times F_1$ merging strategy.

ii) List all candidate 4-itemsets obtained by a candidate generation procedure in Apriori algorithm.

iii) List all candidate 4-itemsets that survive the candidate pruning step of Apriori algorithm.

5. a) Explain the various text mining tasks 6

i) Document classification analysis

ii) Document clustering analysis

b) Can we construct a data cube for multimedia data analysis ? Justify your answer. 4

c) Explain the visual and audio data mining with examples. 6

d) Explain data mining, privacy and data security. 4

PPP – 27

M.E. (Computer Science and Engineering) (Semester – I)
Examination, May 2016
DATA MINING (Elective – II)
(Revised Course)

uration : 3 Hours

Total Marks : 100

Instructions *: 1) Answer* ***five full*** *questions.*
2) Make suitable assumptions ***wherever*** *necessary.*

1. a) Explain the various data preprocessing steps that are applied to make the data more suitable for data mining. 10
   b) Explain the direct methods for rule extraction, to build a rule based classifier. 10
2. a) Explain the various factors that affect the computational complexity of the Apriori algorithm. 10
   b) Explain the frequent item set generation, in frequent pattern growth algorithm. 10
3. a) Explain the Bisecting K-means clustering algorithm. State the K-means strengths and weakness. 10
   b) Explain the unsupervised cluster evaluation using cohesion and separation. 10
4. a) Explain the different approaches proposed for similarity-based search in multimedia data. 5
   b) Explain the spatial mining association and co-location patterns in spatial databases. 5
   c) Explain the following dimensionality reduction techniques used in text mining. 10
      i) Latent semantic indexing.
      ii) Probabilistic latent semantic analysis.
      iii) Locality preserving indexing.
5. a) Explain the HITS (Hyperlink-Induced Topic Search) algorithm to find authoritative pages in mining the world wide web. 7
   b) Explain the techniques used for statistical data mining. 3
   c) Discuss the various examples of "ubiquitous and invisible" social impacts of data mining that affect everyday things from the product stocked at supermarket, to the ads seen while surfing the internet. 10

5. a) Explain the visual and audio data mining with examples. 10

   b) Explain the following trends in data mining 10

      i) Biological data mining

      ii) Web mining

      iii) Graph mining

      iv) Constraint based mining

      v) Mining software bugs.

PPP-A – 58

M.E. CSE (Semester – I) Examination, Nov./Dec. 2012
DATA MINING (Elective – II)

uration : 3 Hours

Max. Marks : 100

Instructions: *1) Answer **any five** questions.*

*2) Make **suitable** assumption is required.*

. a) Define knowledge discovery in databases. 2

b) Given two objects represented by the tuples (20, 2, 41, 11) and (19, 1, 35, 9) : 8

i) Compute the Euclidean distance between two objects.

ii) Compute Manhattan distance between two objects.

iii) Compute the Minkowski distance between two objects, using q = 3.

c) With the help of neat block diagram, explain the data mining steps in the process of knowledge discovery. 10

. a) Suppose that data mining task is to cluster the following eight points With (x, y) representing location) into three clusters. 10

D1(2, 10), D2(2, 5), D3(8, 4), E1(5, 8), E2 (7,5), E3(6, 4), F1(1, 2), F2(4, 9)

The distance function is Euclidean distance. Suppose initially we assign A1, B1 and C1 as the center of each cluster, respectively. Use K-mean algorithm to find out :

i) The three clusters center after first round execution

ii) The final three cluster. (Show all steps.)

P.T.O.

5. a) Write an apriori algorithm to mine sequence pattern. Using above algorithm identify sequence pattern from given data. Min_support = 50%. Interpret the patterns. 10

Object	TimeStamp	Event
A	1	1,2,4
A	2	2,3
A	3	5
B	1	1,2
B	2	2,3,4
C	1	1,2
C	2	2,3,4
C	3	2,4,5
D	1	2
D	2	3,4
D	3	4,5
E	1	1,3
E	2	2,4,5

b) What is Bayesian classification ? Explain how to classify data using Bayes theorem. Comment on the final output generated and compared this method with decision induction tree. 10

6. a) Explain the decision tree induction. Write a recursive algorithm to implement generate decision tree and trace it. 10

b) What is transactional database. List and explain various advance database system. 10

6. a) Explain how to calculate cost of an integrated circuit and also describe how cost becomes price by taking an practical example. 10

   b) Explain different bench marking tools available for evaluating the performance of computer systems. 10

7. Write short notes on **any four**. (4×5=20)

   a) Non-linear pipelines

   b) CPU architectures

   c) Super computing environment

   d) Multi-core system architectures

   e) Floating point arithmetic performance enhancement measure.

PPP-A – 24

M.E. Computer Science and Engineering (Semester – I) Examination, November/December 2015
ADVANCES IN COMPUTER ARCHITECTURE
(Revised Course)

Duration : 3 Hours | Total Marks : 100

Instructions : 1) ***Assume*** *suitable data if necessary.*
2) *Answer* ***any five*** *questions.*
3) *Draw* ***neat*** *diagrams if* ***required***.
4) *Write question numbers legibly while answering.*
5) *Solve problems with* ***appropriate*** *assumptions if required.*

1. a) Our favorite program runs in 10 seconds on computer A, which has a 2 GHz clock. We are trying to help a computer designer build computer B, which will run this program in 6 seconds. The designer has determined that a substantial increase in the clock rate is possible, but this increase will affect the rest of the CPU design, causing computer B to require 1.2 times as many clock cycles as computer A for this program. What clock rate should we tell the designer to target ? 6

b) Suppose we have two implementations of the same Instruction Set Architecture (ISA). For some program,

Machine A has a clock cycle time of 10 ns. and a CPI of 2.0

Machine B has a clock cycle time of 20 ns. and a CPI of 1.2

Which machine is faster for this program, and by how much ?

Assume that number of instructions in the program is 1,000,000,000. 4

c) A compiler designer is trying to decide between two code sequences for a particular computer. The hardware designers have supplied the following facts :

	CPI for each instruction class		
	A	B	C
CPI	1	2	3

P.T.O.

For a particular high-level language statement, the compiler writer is considering two code sequences that require the following instruction counts :

Code sequence	Instruction counts for each instruction class		
	A	B	C
1	20	10	20
2	40	10	10

Which code sequence executes the most instructions ? Which will be faster ? What is the CPI for each sequence ? 6

d) The design team for a simple, single-issue processor is choosing between a pipelined or non-pipelined implementation.

Here are some design parameters for the two possibilities :

Parameter	Pipelined Version	Non-Pipelined Version
Clock Rate	500 MHz	350 MHz
CPI for ALU instructions	1	1
CPI for Control Instructions	2	1
CPI for Memory Instructions	2.7	1

i) For a program with 20% ALU instructions, 10% control instructions and 70% memory instructions, which design will be faster ? Give a quantitative CPI average for each case.

ii) For a program with 80% ALU instructions, 10% control instructions and 10% memory instructions, which design will be faster ? Give a quantitative CPI average for each case.

2. a) Suppose that the cost of processing a single 200mm-diameter wafer in a particular process is Rs. 2,000 and that 90% of the dice fabricated are good. Determine the processing cost per good die, for a die size of 0.5 cm.

b) i) A cooling door for a rack costs $4000 and dissipates 14 KW (into the room; additional cost is required to get it out of the room). How many servers with a Sun Niagara 8-core processor, 1 GB 240-pin DRAM, and a single 5400 rpm hard drive can you cool with one cooling door ?

ii) You are considering providing fault tolerance for your hard drive. RAID 1 doubles the number of disks. Now how many systems can you place on a single rack with a single cooler ?

iii) In a single rack, the MTTF of each processor is 4500 hours, of the hard drive is 9 million hours, and of the power supply is 30K hours. For a rack with 8 processors, what is the MTTF for the rack ?

Use the below figure for your power calculations :

Component

Type	Product	Performance	Power
Processor	Sun Niagara 8-core	1.2 GHz	72-79 W peak
	Intel Pentium 4	2 GHz	48.9-66 W
DRAM	Kingston X 64 C3 AD 21 GB	184-pin	3.7 W
	Kingston D2 N3 1 GB	240-pin	2.3 W
Hard drive	Diamond Max 16	5400 rpm	7.0 W read/seek, 2.9 W idle
	Diamond Max Plus 9	7200 rpm	7.9 W read/seek, 4.0 W idle

Power consumption of several computer components. 6

c) Suppose that we are considering an enhancement to the processor of a server system used for web serving. The new CPU is 10 times faster on computation in the web serving application than the original processor. Assuming that the CPU is idle for 20% of time, does computation for 30% of time and is waiting for I/O for 50% time, what is the overall speed up in this case ? 4

d) A program has 10% divide instructions. All non-divide instructions take one cycle. All divide instructions take 50 cycles.

i) What is the CPI of this program on this processor ?

ii) What would the speedup be if we speed up divide by 2 ?

iii) What would the speedup be if we speed up divide by 50 ? 6

3. a) Define a trace. How is it used in a VLIW processor ? List out the advantages and disadvantages of VLIW processor. 10

b) Describe three performance metrics that are critical in any hardware communication mechanism. 10

PPP-A – 56

M.E. Computer Science and Engg. (Semester – I) (Revised Course) Examination, Nov./Dec. 2014
ADVANCES IN COMPUTER ARCHITECTURE

Duration : 3 Hours

Max. Marks : 100

Instruction : *Answer **any** 5 questions.*

1. 1) Suppose you have a load/store computer with the following instruction mix :

Operation	Frequency	No. of clock cycles
ALU ops	25%	1
Loads	25%	2
Stores	25%	2
Branches	25%	3

a) Compute the CPI.

b) We observe that 25% of the ALU ops are paired with a load, and we propose to replace these ALU ops and their loads with a new instruction. The new instruction takes 1 clock cycle. With the new instruction added, branches take 5 clock cycles ; compute the CPI for the new version.

c) If the clock of the old version is 15% faster than the new version, which version has faster CPU execution time and by how much percent ? 10

2) You have a system that contains a special processor for doing floating point operations. You have determined that 60% of your computations can use the floating point processor. When a program uses the floating-point processor, the speed up of the floating-point processor is 50% faster than when it doesn't use it.

a) Overall speed up by using the floating-point processor.

b) In order to improve the speed up you are considering two options :

- **Option 1** : Modifying the compiler so that 60% of the computations can use the floating-point processor. Cost of this option is Rs 500 k.
- **Option 2** : Modifying the floating point processor. The speed up of the floating-point processor is 100% faster than when it doesn't use it. Assume in this case that 60% of the computations can use the floating-point processor. Cost of this option is Rs 600 k. Which option would you recommend ? Justify your answer quantitatively. 10

P.T.O.

Consider the following values of voltage in each generation.

Processor	Voltage
80286 (1982)	5
80386 (1985)	5
80486 (1989)	5
Pentium (1993)	5
Pentium Pro (1997)	3.3
Pentium 4 Willimette (2001)	1.75
Pentium 4 Prescotte	1.25
Core 2 ketsfield (2007)	1.1

d) Find the average capacitive loads, assuming a negligible static power consumption.

e) Find the largest relative change in voltage between generations.

f) Find the geometric mean of the voltage ratios in the generation since pentium. 12

3. A) What is dynamic scheduling ? Explain how it is used to reduce data hazards. 10

B) Explain to Masulo's approach. How it is minimizers data Hazards ? 10

4. A)

			Reference
	Computer A (Seconds)	Computer B (Seconds)	Computer (Seconds)
Program 1	1	2	6
Program 2	2	2	1
Program 3	4	2	1

Table 1

The execution times of three programs on three computers are listed in table 1 above. Calculate the following :

i) The arithmetic means of the execution times on each computer for the of programs. Which computer has the highest performance ? Mention potential problem with comparing performance using arithmetic means.

ii) For each of the programs, calculate the ratio of the execution time computer A and the reference computer for each of the programs and the do the same for computer B.

iii) Now use geometric mean to calculate the ratio of the performance o computer A and B. Which of the computers has the highest performance using this approach ?

B) We want to reduce the execution time of a program using a multi processor system. The execution time of the original sequential program is 12 seconds. Now assume that 75% of the sequential program execution can be distributed evenly across processors in a multiprocessor system. What is the speed up achieved ? What is the maximum speed-up that can be ever achieved ?

C) Assume that a memory instruction has an average CPI = 4 and that the average CPI for the rest of the instructions is CPI = 2. Given that 20% of the instructions are memory instructions, that a program executes 1,000,000,000 instructions and the clock frequency is 1 GHz, what is the execution time of the program ? Also derive an expression for computing the performance of computers in general case.

5. A) Describe basic compiler techniques for exposing instruction level parallelism.

B) What is virtual memory ? Write techniques for fast address translation.

C) What is multithreading ? How to exploit thread level parallelism ?

6. A) Consider a symmetric multiprocessing architecture machine which is using a snooping cache coherency protocol. The protocol is "write invalidate", that is invalidate messages are sent to the bus whenever a cache block is over written.

Let us now consider a write broadcast based cache coherency protocol.

i) Draw the diagram of the cache states for this protocol.

ii) Discuss the differences between the write – Broadcast and the write invalidate protocol.

iii) Which one is easier to implement ?

iv) Which one is more efficient ? Under what conditions ? Discuss. 10

B) Explain various techniques for reducing cache miss penalty. 10

7. Write short notes on **any four** : (4×5=20)

a) Vector processors.

b) Non linear pipelines.

c) Cache Coherence.

d) Super Computers.

e) Floating-point arithmetic performance enhancement.

GPP-A – 31

M.E. Computer Science and Engineering (Semester – I) Examination, Nov./Dec. 2013
ADVANCED COMPUTER ARCHITECTURE (New Course)

Duration : 3 Hours | Total Marks : 100

Instructions : *1) **Assume** suitable data **if necessary**.*
*2) Answer **any five** questions.*
3) Draw neat diagrams if required.
4) Write question numbers legibly while answering.
*5) Solve problems with appropriate assumptions **if required**.*

1. A) Consider the MIPS program in Figure 1 which operates on an array with 64-bit elements. The register R1 points to the beginning of the array from the beginning and R2 the end. The array always contains 1000 elements.

You are using this program as a benchmark when comparing two design options. Your basic design is a single-issue pipeline where each instruction executes in 1 cycle except LD, SD,DMUL, and BNE. The LD and SD takes 14 cycles extra penalty if a data cache miss occurs. The DMUL and the branch, BNE, instructions both takes 8 cycles to execute. The miss rate for the data cache has been measured to be 15%. No other stalls occur.

You are considering the option of increasing the clock frequency by 40%. This would however hurt some of the latencies. The BNE would be unchanged, but the DMUL would need yet an another extra cycle which would make it execute in 9 cycles instead. The data cache miss time would stay the same measured in absolute time (but the actual stall cycles for LD and SD would increase).

i) What is the expected speedup when using the high frequency design compared to the basic design ? **6**

ii) Explain Ahmdal's law and use it to calculate the additional speedup obtained if we could halve the data cache miss stall time in the high frequency design. **6**

iii) Another option would be to reduce the miss rate by using a larger data cache. This would however influence the yield. Explain the term yield when manufacturing integrated circuits. What influences the yield and what is its relation to cost ? **4**

P.T.O.

iv) Instead of increasing the clock frequency to get higher performance, one option would be to parallelize the program and run it on two processor cores with lower clock frequency instead. What is the general trade-offs between these options when looking specifically at power use ?

```
        ANDI R3, R3, 0
LOOP :  LD R4, 0(R1)
        DMUL R5, R4, R4
        DADD R5, R3, R5
        SD R5, 0(R1)
        DADDI R3, R4, 0
        DADDI R1, R1, 8
        BNE R1, R2, LOOP
```

FIGURE 1 MIPS program

2. Consider again the MIPS program in Figure 1 (see previous question).

   A) Generally, you can find three different types of dependencies in programs. Explain these different dependencies and give at least one example for each one in the program in Figure 1. 6

   B) For data hazards, you often find the terms RAW, WAR, and WAW. What does RAW, WAW, and WAR stand for ? Try to give examples of each type of data hazard in the program in Figure 1. If you think there are no examples, state the explicitly. 6

   C) Two techniques often used by compilers are instruction scheduling and loop unrolling. Show how the loop in the program in Figure 1 can be unrolled and how the instructions can be rescheduled. In your new program, point to the specific changes you have made that you think will improve the performance and explain why the performance will be improved. 8

3. A) What is dynamic scheduling ? Explain how it is used to reduce data hazards ? 10

   B) Explain Tomasulo's approach, how it is minimizes data hazards ? 10

4. A)

	Computer A (seconds)	Computer B (seconds)	Reference Computer (seconds)
Program 1	1	2	6
Program 2	2	2	1
Program 3	4	2	1

Table 1

The execution times of three programs on three computers are listed in the table 1 above. Calculate the following :

i) the arithmetic means of the execution times on each computer for the set of programs. Which computer has the highest performance ? Mention a potential problem with comparing performance using arithmetic means. 3

ii) For each of the programs, calculate the ratio of the execution time on Computer A and the reference computer for each of the programs and then do the same for Computer B. 2

iii) Now use geometric mean to calculate the ratio of the performance on Computer A and B. Which of the computers has the highest performance using this approach ? 3

B) We want to reduce the execution time of a program using a multiprocessor system. The execution time of the original sequential program is 12 seconds. Now assume that 75% of the sequential program execution can be distributed evenly across processors in a multiprocessor system. What is the speed-up achieved ? What is the maximum speed-up that can be ever achieved ? 5

C) Assume that a memory instruction has an average CPI =4 and that the average CPI for the rest of the instructions is CPI = 2. Given that 20% of the instructions are memory instructions, that a program executes 1,000,000,000 instructions and the clock frequency is 1 GHz, what is the execution time of the program ? Also derive an expression for computing the performance of computers in general case. 7

5. A) Consider the following loop. First, list all the dependencies. Can this loop, as written, be unrolled ? Explain. Can this loop be rewritten so loop unrolling can be applied ? Explain.

```
For (i=1;i<100;i=i+1){
A[i] = A[i] + B[i]; /*S1 */
B[i]=2*B[i];/*S2 */
C[i] = A[i–1] + B[i–1]; /*S3 */
D[i] = D[i] * A[i–1]; /* S4 */
}
```

10

B) Consider the following piece of code given figure 2, and derive the branch prediction accuracy for the two branch predictors defined in what follows :

```
style="mso-ansi-language:PL";
 lang "PL";
 i = 12 ;
while (i<104) do { /* Branch b1 */
      style="mso-ansi-language:PL";
      lang="PL";
      a[i] = a[i]*b[i];
       i = i+1;
}
while(i<1500) do { /*branch b2 */
      c[i] = c[i]*10;
      i = i+1;
}
```

T* – taken

NT*– not taken

Figure 2

First assume you have a 1 bit branch predictor set to T*. Derive branch prediction accuracy for the above code using the 1 bit predictor. Next, assume you have a 2 bit branch predictor as given in Figure 2. Compute the branch prediction accuracy for the above code assuming both branches history bits are set to NT*. Write state transition diagram for 2 bit branch predictor. 6

C) Give an precise formula for memory access time in terms of cache hit rate, cache hit time (Assume all cache misses are memory hits and that memory hit time exceeds cache hit time). Also derive expression for memory access time using two level caches (L-1 Cache, L-2 Cache). 4

6. A) Consider a symmetric multiprocessing architecture machine which is using a snooping cache coherency protocol. The protocol is "write invalidate", that is invalidate messages are sent to the bus whenever a cache block is overwritten. Let us now consider a write-broadcast based cache coherency protocol.

i) Draw the diagram of the cache states for this protocol.

ii) Discuss the differences between the write-broadcast and the write-invalidate protocol.

iii) Which one is easier to implement ?

iv) Which one is more efficient ? Under what conditions ? Discuss. 10

B) Explain various techniques for reducing cache miss penalty. 10

7. A) Describe basic compiler techniques for exposing instruction level parallelism. 8

B) What is virtual memory ? Write techniques for fast address translation. 8

C) What is multithreading ? How to exploit Thread Level Parallelism ? 4

PPP – A27

M.E. Computer Science and Engineering (Semester – I) Examination, May/June 2013
DISTRIBUTED OPERATING SYSTEMS

Duration : 3 Hours

Total Marks : 100

Instructions : *1)* ***Assume*** *necessary data.*
2) Answer ***any 5*** *questions.*

1. a) Differentiate between FX sequent computers and Intel's Hypercube. 4
   b) How multiprocessor system reduces memory access time ? 3
   c) Which are the factors based on that programmer designs multiprocessor systems ? 3
   d) Consider the case where creating, maintaining and switching between processes occurs frequently. Which processes will give solution for this situation. Explain it in detail. 6
   e) What are the various problems which arise in designing memory management in Mach OS ? 4

2. a) Suppose there are 3 sites S_1, S_2, S_3 making request to CS using (2, 1), (1, 2), (1, 2). Apply lamports algorithm to implement mutual exclusion on the sites to enter CS. Optimize solution using Ricart Agrawal algorithm. 12
   b) Why information structure is needed in the mutual exclusion algorithms. 8

3. a) How naming and location works in sun network file system ? 4
   b) In what aspects log file structured file system is different from other file system. 8
   c) What are the benefits of grouping files into volumes in coda ? 8

4. a) How fault detection mechanism works in sequio OS. 5
   b) Explain fault recovery in sequio OS. 4
   c) How efficiency has increased in Mach virtual memory systems ? 4
   d) Explain memory protection and machine independence in MACH OS. 7

5. a) Discuss the features of the following : 12
      i) Linux
      ii) Solaris
      iii) Real Time OS.
   b) How Caching can be done in distributed systems ? 8

P.T.O.

6. a) Discuss various methods to implement distributed shared memory. 8
   b) Why and how coherence semantics improve performance of shared memory? 10
   c) How Cache Coherence is implemented in PLUS system. 2

7. a) Is graph theoretic model better than two types of preliminaries in deadlock detection ? Discuss advantages and disadvantages of various methods. 8
   b) Explain various mechanisms for building distributed file system. 10
   c) Why deadlock detection and resolution is difficult to implement ? 2

8. a) Real Time OS uses control area network. Is it true ? How it is different from other networks ? 6
   b) How process management plays important role in the multimedia OS ? 6
   c) Round Robin scheduling algorithm is better than all scheduling policies. Illustrate your answer. 8

PPP-A – 57

M.E. Computer Science and Engineering (Semester – I) Examination, Nov./Dec. 2012
ADVANCED IN COMPUTER ARCHITECTURE

Duration: 3 Hours | Total Marks: 100

Instructions : *1) Assume suitable data **if necessary**.*
*2) Answer **any five** questions.*
*3) Draw **neat** diagrams **if required**.*
*4) Write question numbers legibly **while** answering.*
*5) Solve problems with appropriate assumptions **if required**.*

1. 1) Suppose you have a load/store computer with the following instruction mix :

Operation	Frequency	No. of clock cycles
ALU ops	25%	1
Loads	25%	2
Stores	25%	2
Branches	25%	3

a) Compute the CPI.

b) We observe that 25% of the ALU ops are paired with a load, and we propose to replace these ALU ops and their loads with a new instruction. The new instruction takes 1 clock cycle. With the new instruction added, branches take 5 clock cycles; Compute the CPI for the new version.

c) If the clock of the old version is 15% faster than the new version, which version has faster CPU execution time and by how much percent ? **10**

2) You have a system that contains a special processor for doing floating-point operations. You have determined that 60% of your computations can use the floating-point processor. When a program uses the floating-point processor, the speedup of the floating-point processor is 50% faster than when it doesn't use it.

a) Overall speedup by using the floating-point processor.

b) In order to improve the speedup you are considering two options.

- Option 1 : Modifying the compiler so that 60% of the computations can use the floating-point processor. Cost of this option is Rs. 500 K.

P.T.O.

Consider the following values of voltage in each generation.

Processor	Voltage
80286 (1982)	5
80386 (1985)	5
80486 (1989)	5
Pentium (1993)	5
Pentium Pro (1997)	3.3
Pentium 4 willimette (2001)	1.75
Pentium 4 prescott	1.25
Core 2 Ketsfield (2007)	1.1

d) Find the average capacitive loads, assuming a negligible static power consumption.

e) Find the largest relative change in voltage between generations.

f) Find the geometric mean of the voltage ratios in the generations since pentium. 12

3. 1) Consider the following instruction sequences :

	Instruction sequences

A	Lw \$1,40(\$6)
	Add \$6, \$2, \$2
	Sw \$6, 50(\$1)

B	Lw \$5, -16(\$5)
	Sw \$5,-16(\$5)
	Add \$5, \$5, \$5

a) Indicate dependences and their type.

b) Assume there is no forwarding in this pipelined processor. Indicate hazards and add nop instructions to eliminate them.

c) Assume there is full forwarding. Indicate hazards and add nop instructions to eliminate them. The remaining problems in this exercise assume the following clock cycle times.

	Without forwarding	With full forwarding	With ALU-ALU forwarding only
A	300 ps	400 ps	360 ps
B	200 ps	250 ps	220 ps

d) What is the total execution time of this instruction sequence without forwarding and with full forwarding ? What is the speedup achieved by adding full forwarding to a pipeline that has no forwarding ? 12

2) Briefly give two ways in which loop unrolling can increase performance and one in which it can decrease performance. 8

4. 1) Distinguish between flow dependency (true), anti-dependency and output dependency of instructions. Give one example of each of these dependencies.

a) Why should one be concerned about dependencies in pipelined execution of the programs ?

b) If in pipelined execution of programs out of order completion of instructions is avoided, will these dependencies matter. 10

2) Define a trace. How is it used in a VLIW processor ? List out the advantages and disadvantages of VLIW processor. 10

5. 1) Explain how to reduce branching cost by static branch prediction technique. 8

2) Caches are important to providing a high performance memory hierarchy to processors. Below is a list of 32 bit memory address references given as word addresses.

A	1, 134, 212, 1, 135, 213, 162, 2, 44, 41, 221
B	6, 214, 175, 214, 6, 84, 65, 174, 64, 105, 85, 215

a) For each of these references, identify binary address, the tag and the index given a direct a mapped cache with 16 one words blocks. Also list if each reference is a hit or miss, assuming the cache is initially empty.

b) For each of these references, identify the binary address, the tag and the index given a direct mapped cache with two word blocks and a total size of eight blocks. Also list if each reference is hit or miss, assuming the cache is initially empty.

c) You are asked to optimize a cache design for the given references. There are three direct mapped cache design possible, all with a total of eight words of

Data : C1 has one word blocks, C2 has two word blocks and C3 has four word blocks. In terms of miss rate, which cache design is best ? If the miss stall time is 25 cycles, and C1 has access time of 2 cycles, C2 takes 3 cycles and C3 takes 5 cycles, which is the best cache design ? 12

OR

2) Consider a cache (M1) and memory (M2) hierarchy with the following characteristics :

M1 : 16 K words, 50 ns access time

M2 : 1 M words, 400 ns access time.

Assume eight-word cache blocks and a set size of 256 words with set-associative mapping.

a) Show the mapping between M2 and M1.

b) Calculate the effective memory access time with a cache hit ratio of h = 0.95. 12

6. 1) What is multithreading ? How to exploit Thread Level Parallelism (TLP) ? 8

2) Compare the following multiprocessor hardware organization.

Time share bus

Cross bar switch

Multi port memories. 6

3) Explain Flynn's classification of computer architectures : give the definition of the alternative architectures and draw for each one a block diagram. 6

7. Write short notes on **any four** : (4×5=20)

a) Vector processors

b) Non linear pipelines

c) Cache coherence

d) Super computers

e) Floating point arithmetic performance enhancement.

PPP-A – 30

M.E. (CS&E) (Semester – I) (Revised Course) Examination, November/December 2017
DISTRIBUTED OPERATING SYSTEM

Duration : 3 Hours | Total Marks : 100

Instructions: *1) Answer **any 5** questions.*
*2) Assume **necessary** details if **needed** and **state**.*
*3) Draw **neat** labelled diagrams using pencil.*
*4) Answer questions in the **same** sequence.*
*5) Answer every **new** question on a **fresh** page.*

1. a) Define and discuss Cache coherence problem. Give your suggestive solutions. **10**
   b) Explain hypercube architecture support your answer with a neat diagram. **10**

2. a) List 3 basic classes of multiprocessor systems and discuss each one. List merits and demerits of each one. **10**
   b) Define threads and discuss features of following :
      i) User level threads
      ii) Kernel level threads. **10**

3. a) Discuss following design issues with reference to 'MACH' operating system : **10**
      i) Portability
      ii) Protection
      iii) Data sharing
      iv) Efficiency.
   b) Explain structure of Kernel based operating system design approach. Support your answer with a neat diagram. **10**

4. a) List and discuss any 2 models of deadlocks. **8**
   b) List and explain 4 motivations of distributed systems. **8**
   c) How resources of distributed systems are made available to users ? Explain both schemes. **4**

5. a) Discuss features, merits and demerits of ISO-OSI reference model. Provide an over view of ISO-OSI layers. **10**
   b) Provide and explain HO – Ramamoorthy's two phase algorithm for centralised deadlock detection. Discuss its merits and demerits. **10**

P.T.O.

6. a) Explain architecture of distributed file system. Support your answer with a neat diagram. 8
   b) List and explain advantages of distributed shared memory. 7
   c) Explain why false sharing does not occur in object based systems. 5

7. Write detailed note on **any 4** topics : (4×5=20)
   a) Components of load distribution algorithm.
   b) Semaphores and its uses.
   c) Dining philosophers problem and its solutions.
   d) Deadlock handling strategies.
   e) Network topologies.
   f) Write invalidate protocol.
   g) Process synchronization in multiprocessor operating system.

PPP-A – 55

M.E. (Computer Science and Engg.) (Semester – I)
(Revised Course) Examination, Nov./Dec. 2014
DISTRIBUTED OPERATING SYSTEMS

Duration : 3 Hours Max. Marks : 100

Note : *1) Answer **any 5** questions.*
*2) Assume **necessary** details.*
*3) **Wherever** needed draw labelled diagram(s) using PENCIL.*
*4) Answer questions in the **same** sequence.*
*5) Answer **every new** question on a **fresh** page.*

1. a) A task consists of several sub tasks, if these sub tasks communicate synchronously with each other frequently, which scheduling policy is would you recommend and why ? Discuss. 10

   b) List and discuss design issues of Memory Management in 'MACH' Operating System. 6

   c) Write a detailed note on :
      i) Threads
      ii) Reliability and Fault tolerance
      iii) Process synchronization
      iv) Process or Scheduling. 4

2. a) With the help of a neat diagram explain Architecture of a Distributed System. 5

   b) Discuss any 5 issues in Distributed Operating System. 5

      Issues :
      i) Global Knowledge
      ii) Naming
      iii) Scalability
      iv) Compatibility
      v) Resource Management
      vi) Security.

   c) With help of a neat diagram explain functioning of Remote Procedure Calls. Discuss design issues. 10

P.T.O.

3. a) Provide and prove that Lamport's algorithm can achieve mutual exclusion. 6

   b) Provide and explain Ricart-Agarwala Algorithm. Prove that it can achieve mutual exclusion. Compare this with Maekawa's algorithm. 10

   c) Discuss below Listed Deadlock Handling Strategies in Distributed Systems. (any 2). 4

      i) Deadlock Prevention

      ii) Deadlock Avoidance

      iii) Deadlock Detection.

4. a) Provide and discuss any one Hierarchical deadlock Detection Algorithm. List its merits and demerits. 8

   b) With the help of a neat diagram explain Architecture of a distributed file system. How data is accessed by a client in this system ? Discuss. 12

5. a) List any 5 advantages of Distributed Shared Memory and discuss. 5

   b) Provide and discuss central server Algorithm for implementation of DSM. Compare this algorithm, with migration Algorithm. Give your comments on each algorithm. 12

   c) Explain features of 'MIRAGE' with reference to DSM. (DSM : Distributed Shared Memory). 3

6. a) With the help of neat flow chart discuss receiver initiated load Sharing Algorithm. 8

   b) Define and explain features of tightly coupled and loosely coupled systems. Give one example each. List its merits and demerits. 6

   c) What is Caching ? Explain, discuss Cache Coherence problems. Suggest solutions. 6

7. Write detailed note on **any 4**. **(4×5=20)**

   1) Hyper cube Architectures.

   2) 8×8 Omega Multistage Interconnection Network.

   3) Kernel level threads.

   4) Issues in Process Synchronizations.

   5) Open SOLARIS.

   6) Open Source Softwares.

5. a) Explain the hierarchical deadlock detection algorithm. 8
   b) Explain the multimedia process scheduling in a multimedia operating system. 8
   c) How are the disadvantages of user-level threads overcome by using a scheme based scheduler activation ? 4

6. a) Explain the location policy component for sender-initiated load distributing algorithm. 8
   b) Explain the features of Linux operating system. 4
   c) How is the log structured file system different from other file system ? Also explain the disk space management. 8

PPP-A – 56

M.E. (Computer Science and Engineering) (Semester – I) Examination, Nov./Dec. 2012
DISTRIBUTED OPERATING SYSTEMS

Duration : 3 Hours

Total Marks : 100

Instructions : *i) Attempt **any five** questions.*
*ii) Make suitable assumptions **if required**.*

1. a) Why is heterogeneity unavoidable in many distributed systems ? What are some of the common types of incompatibilities encountered in heterogeneous distributed systems ? What are the common issues with which the designer of heterogeneous distributed systems must deal ? 6

   b) Suggest three different routing strategies for use in computer networks. List the relative advantages and disadvantages of the strategies suggested by you. 6

   c) Is it always necessary for the sender of a message to know that the message arrived safely at its destination ? Give two examples in support of your answer. 4

   d) What do you mean by a microkernel ? Why is it considered to be a promising technology for distributed operating system ? 4

2. a) Explain the ways of achieving synchronization using Ada Rendezvous. Give a solution to producers-consumers problem using Ada tasks. 10

   b) Discuss the concept of path expression to achieve process synchronization. Explain what the following path expression do.

   path {openread; read}; {openwrite; write}; close end. 6

   c) Comment on the types of resources and resource access in a distributed system. 4

3. a) Explain the general resource system. 2

   b) Consider the following preemption method to prevent deadlocks. All processes are assigned unique priorities that can be totally ordered. A sequencing process is allowed to preempt another process that holds the needed resource only if the requesting process has higher priority, otherwise, it is blocked. Show that this method prevents deadlocks. 8

P.T.O.

c) When distributed systems are designed and engineered, certain fundamental properties have to be taken into account, including : concurrent execution of components, independent failure modes, communication delay and no global time. Give three examples of the implications of these properties on the engineering of large-scale, widely distributed systems. **10**

4. a) Give and explain the Maekawa's algorithm for distributed mutual exclusion. **8**

b) Compare the features of non-token based algorithms with token based algorithms to achieve distributed mutual exclusion. **4**

c) Discuss the need of centralized deadlock detection algorithms. Compare the one-phase and two-phase algorithms proposed by Ho-Ramamoorthy. **8**

5. a) What are the applications of Agreement Algorithms ? Explain. **5**

b) Show how a solution to the consensus problem can be used to solve the interactive consistency problem. **6**

c) Discuss the mechanisms for distributed file systems. **9**

6. a) What do you understand by memory coherence ? Give different forms of memory coherence. **5**

b) Comment on the impacts of Granularity and page replacement policy on the design of distributed shared memory. **5**

c) Explain the working principle of Receiver initiated load distribution algorithms. **5**

d) What are the different aspects to be considered while selecting a suitable load sharing algorithm ? **5**

7. Write short notes on **any four** : **(4×5=20)**

a) Distributed features of Open SOLARIS

b) RTOS Classifications

c) RTOS Interrupt Handling

d) Cloud Computing

e) Features of IVY.

PPP – 52

M.E. (Computer Science and Engineering) (Semester – I) Examination, November/December 2016 (Revised Course) DISTRIBUTED OPERATING SYSTEM

Duration : 3 Hours | Max. Marks : 100

***Note** : 1) Answer **any 5** questions.*
*2) Assume **necessary** details, **if** needed.*
*3) **Wherever** needed draw **neat** labelled diagram (s) using **Pencil**.*
*4) Answer questions in the **same** sequence.*
*5) Answer every **new** question on a **fresh** page.*

1. a) Explain the Cache Coherence Problems. Give two remedies. **8**

   b) Discuss following design issues of operating system. **(4×3=12)**

      i) Threads

      ii) Process synchronization.

      iii) Process scheduling

      iv) Reliability.

2. a) With the helps of a neat diagram explain features of key components of 'Mach' operating system. **10**

   b) Do you think the page replacement algorithm for multiprocessor system should be different from that of a uniprocessor system ? Discuss. **5**

   c) If the subtask of a task have large critical sections. Which scheduling policy its most desirable ? Explain. **5**

3. a) With the help a neat diagram explain structure and functioning of 8 × 8 omega multistage interconnection network. List its advantages and disadvantagves. **10**

   b) Explain design issues of RPC. **5**

   c) List and explain any 2 fundamental causes of Deadlocks. **5**

P.T.O.

PPP – 52

4. a) Compare Resource deadlocks vs Communication deadlock. Give one example each. 6

   b) Provide and discuss Ho.Ramamorthy's two phase centralised deadlock detection algorithm. 10

   c) Write short notes on (**any 1**) : 4

      i) Deadlock avoidance

      ii) Deadlock detection.

5. a) Explain Lamport's algorithm for distributed Mutual Exclusion. 10

   b) Explain Suzuki – KASAMI's Broadcast Algorithm for distributed Mutual Exclusion. 10

6. a) With the help a diagram explain typical data access action in distributed file systems. 10

   b) Explain why false sharing does not occur in object based systems. 5

   c) Explain LRU policy with classes in page replacement. 5

7. Write short note on **any 5** : (5x4=20)

   a) Semaphores

   b) Producer – Consumer Problem.

   c) Deadlock Handling strategies.

   d) Resource Management in Distributed operating systems.

   e) Raymond's tree based Algorithm for Distributed Mutual Exclusion.

   f) Stateless Server.

   g) Fault tolerance.

   h) Features of open SOLARIS.

PPP-A – 23

M.E. (Computer Science and Engg.) (Semester – I) Examination, November/December 2015
DISTRIBUTED OPERATING SYSTEMS
(Revised Course)

Duration : 3 Hours

Total Marks : 100

Note : *1) Answer **any 5** questions.*
*2) Assume **necessary** details.*
*3) **Wherever** needed draw labelled diagram(s) using pencil.*
*4) Answer questions in the **same** sequence.*
*5) Answer every **new** question on a fresh page.*

1. a) Compare tightly coupled systems with loosely coupled systems. Give one example each. 8

   b) With the help of a neat diagram explain working of 8 × 8 omega multistage interconnection network. 7

   c) List and discuss 05 design issues of operating system. 5

2. a) List and discuss 04 motivations behind distributed systems. 8

   b) Discuss following issues with reference to distributed operating systems.
      i) Global knowledge
      ii) Naming system
      iii) Scalability
      iv) Compatibility
      v) Resource management. 5

   c) List and discuss features of any 2 types of structures of operating system. 4

   d) Compare packet switching with circuit switching. List its merits and demerits. 3

3. a) Provide and explain 'LAMPORT's algorithm for distributed mutual exclusion. List its merits. 10

   b) Provide and discuss RAYMOND's tree based algorithm for distributed mutual exclusion. List its demerits. 10

P.T.O.

4. a) List and compare 02 types of distributed deadlocks. 5
   b) Discuss deadlock prevention strategies in distributed systems (02). 5
   c) Provide and explain any one hierarchical deadlock detection algorithm. 6
   d) What do you understand by false deadlock ? Explain. 4

5. a) List and discuss issues in load distribution with reference to distributed scheduling. 8
   b) Discuss details of following policies with reference to load distribution algorithm. 8
      i) Transfer Policy.
      ii) Selection policy
      iii) Location policy
      iv) Information policy.
   c) With the help of a neat diagram discuss features of architecture of distributed file system. 4

6. a) Explain process of selecting a suitable load sharing algorithm. 4
   b) Explain migration algorithm for implementation of DSM. List its merits and demerits. 8
   c) Discuss features of following types of consistencies with reference to memory coherence under DSM.
      i) Weak consistency
      ii) General consistency
      iii) Sequential consistency
      iv) Release consistency. 8

7. Write short notes on **any 4** : **(4×5=20)**
   a) Non-preemptive transfer with reference to load distribution under distributed scheduling.
   b) CONDOR.
   c) NORMA Architecture details.
   d) Distributed operating system performance issues.
   e) Resource reservations and its effects.
   f) Kermel level threads and its uses.

GPP-A – 39

M.E. (Computer Science and Engineering) (Semester – I) Examination, November/December 2013
DISTRIBUTED OPERATING SYSTEMS (New Course)

Duration : 3 Hours | Total Marks : 100

Instructions : *1) Assume **necessary** data.*
*2) Answer **any 5** questions.*

1. a) "Solaris is open source software". What are the various features of solaris makes it open source software ? 10

   b) Explain various activities for implementing security functions in network operating system. 10

2. a) There are 2 resources (R_1, R_2) stored at sites S_1 and S_2. Suppose two transactions T_1 and T_2 has started simultaneously at S_3 and S_4. Apply centralized deadlock detection algorithms to detect deadlock. Optimize solution using one phase algorithm. 12

   b) There are 7 machines in distributed system. Files of system is stored on 6^{th} machine. How 2^{nd} machine can access files, illustrate your answer with respect to architectures ? 8

3. a) Two communication primitives models are used to develop distributed operating systems. Do you agree on it ? 6

   b) What are various requirements of mutual exclusion algorithms ? 4

   c) Suppose distributed system is made up of 7 sites (S_1, S_2, ... S_7}. How mutual exclusion is implemented using Lamport, Ricart Agrawala algorithm if S_1 wants to enter CS ? 10

4. a) "Resource Management is done in the real time OS by resource reservation". Do you agree ? Illustrate your answer. 8

   b) To manage interrupts in real time OS which mechanisms are applied in RTOS ? 4

   c) Differentiate between multimax of encore corporation, CM* of CMU and Intel's hyper cube. 8

P.T.O.

5. a) Explain in detail various strategies of handling deadlock strategies in distributed systems. 6

   b) Which problems arise in deadlock detection and resolution ? 4

   c) Explain various mechanisms underlying for building distributed file systems ? 10

6. a) Programmers achieves some goals using MACH OS. Illustrate your answer. 6

   b) Explain various load sharing policies with respect to the scheduling in distributed system. 8

   c) How efficiency is considered in Mach OS ? 2

   d) Explain fault tolerance with respect to Sequio system. 4

7. a) What are the purposes of the multiprocessors systems ? 4

   b) How the classification of multiple instruction multiple data architecture is done ? 10

   c) "In Hydra on C-MMP all processors are treated as autonomous". How is it different from Cyber 170 ? 6

8. a) In what manner resources of distributed system is shared among the various processors ? 6

   b) Explain various organization of the structure of OS in distributed systems. 6

   c) There are 7 computers, 5 terminals and 8 printers. Explain various methods of communication between various devices. 8

PPP-A – 33

M.E. Computer Science and Engineering (Semester – I) (Revised Course) Examination, November/December 2017
ADVANCED DATA STRUCTURES AND ALGORITHMS

Duration : 3 Hours | Total Marks : 100

Instructions: *1) Answer **five full** questions.*
*2) Make suitable assumptions **wherever** necessary.*

1. a) Explain the data structure which has features of both a Binary Search Tree and a Min Heap order property. Explain how insertion works for the following data = {I : 73, E : 23, A : 10, H : 7, B : 5, C : 25, D : 9, F : 2, G : 4} into the above data structure.
Also delete the following data = {D : 9, F : 2}. **10**

   b) Using backtracking algorithm, explain the turnpike reconstruction problem. **10**

2. a) Create a Fibonacci Heap for the following list L = <20, 10, 15, 5, 35, 22, 17, 47, 30, 33, 39>. After creation of the Fibonacci Heap extract the minimum key node and change the key value of 30 to 13. Write the algorithms to extract the minimum key node. Analyze the algorithm with respect to running time. **14**

   b) Explain the internal representation of the Binomial Heap and Fibonacci Heap. Comment on the amortized time complexity used to analyze the performance of Mergeable Heap operations as implemented for Fibonacci Heap. **6**

3. a) Write the Johnson's algorithm to find the all pair shortest path. Use the Johnson's algorithm to find the shortest path between all pairs of vertices for the given weighted directed graph G = (V, E, W) where V = {a, b, c, d, e}, E = {<a, b>, <a, c>, <a, e>, <b, d>, <b, e>, <c, b>, <d, a>, <d, c>, <e, d>} and W = {3, 8, –4, 1, 7, 4, 2, –5, 6}. **10**

   b) Explain the various Top-Down Splay Tree rotations. Draw the Top-Down Splay Tree for the given input J, B, H, I, E, K, N, C, F, A, M, G, L. Using Top-Down Splay access D. **10**

P.T.O.

4. a) Write the algorithms to compute the best ordering of matrix multiplication. Include the algorithm to print out the actual ordering. What is the optimal way to compute $A_1, A_2, A_3, A_4, A_5, A_6$ where the dimensions of the matrices are $A_1 : 5 \times 30$, $A_2 : 30 \times 10$, $A_3 : 10 \times 40$, $A_4 : 40 \times 5$, $A_5 : 5 \times 10$, $A_6 : 10 \times 20$. 10

   b) A freight train has n railroad cars. Each is to be left at different station. They're numbered through n and freight train visits these stations in order n through 1. The railroad cars are labeled by their destination. To facilitate removal of the cars from the train, we must rearrange them in ascending order of their number (i.e. 1 through n). When cars are in this order, they can be detached at each station. The cars are rearranged at a shunting yard that has input track, output track and k holding tracks between input and output tracks (i.e. holding track). Consider a shunting yard with k = 3, holding tracks H1, H2 and H3, also n = 9. The cars initially are in the order 5, 8, 1, 7, 4, 2, 9, 6, 3 from back to front. Using appropriate ADT solve the above problem. 10

5. a) Create the Red-Black Tree that results by successively inserting the keys 30, 10, 20, 50, 40, 70, 80, 15, 90, 100 into an empty tree. Delete the keys 50, 40, 15, 30. Explain the various rotations and write the algorithm for insertion and delete operation. 14

   b) Find the articulation points and the biconnected components for the given undirected graph G = (V, E) where V = {A, B, C, D, E, F, G, H, I, J}, E = {<G, B>, <G, I>, <B, I>, <B, C>, <B, E>, <E, C>, <E, F>, <F, A>, <F, D>, <A, D>, <F, H>, <F, J>, <H, A>, <J, D>}. 6

PPP – 26

M.E. (Computer Science and Engineering) (Semester – I) (Revised Course) Examination, May 2016
ADVANCED DATA STRUCTURES AND ALGORITHMS

Duration : 3 Hours | Total Marks : 100

Instructions : *i) Attempt **any five** questions.*
*ii) Make **suitable** assumptions if **required**.*

1. a) Order the following by asymptotic growth rates : 3
n^3, $\log(\log^* n)$, $(\sqrt{2})^{\log n}$, $(n+1)!$, 367, 982n, $n^{1/\log n}$, $2^{\log n}$, $n \log n$.

b) Give an example of non-negative functions f(n) and g(n) such that $f(n) \neq O(g(n))$ and $f(n) \neq \Theta(g(n))$. 3

c) An algorithm takes 0.5 ms for input size 200. How long a problem can be solved in 10 minutes if the running time is the following (assume low order terms are negligible) ? 4
i) linear ii) O(n log n) iii) quadratic iv) cubic

d) Give an efficient algorithm if there exists an integer j such that $A_j = j$ in an array of integers $A_1 < A_2 < A_3 < \ldots < A_n$. What is the running time of the algorithm ? 4

e) Define an ADT called Purse that contains other Purses and purse items. Identify atleast six data elements and six operations that can be used with ADT Purse. Suppose it is required to organize N instances of ADT Purse. Suggest two different implementation schemes, other than arrays, along with worst case running time estimation for each operation. 6

2. a) Suppose a binary tree has levels $l_1, l_2, \ldots, l_M$ at depths $d_1, d_2, \ldots, d_M$ respectively. Prove that $\sum_{i=1}^{M} 2^{-d_i} \leq 1$ and determine when the equality is true. 6

b) Derive an expression for the minimum number of nodes in an AVL tree of height h. Using the expression, compute the minimum number of nodes in an AVL tree of height 15. 6

c) Define a Splay tree. Insert the elements 1, 2, . . . , 7 into an initially empty tree. Explain the use of zig, zig-zag and zig-zig operations during the search of 1, 2, 3 and 4. Comment on the amortized running time of search operation. 8

P.T.O.

3. a) Given two sorted lists, L_1 and L_2. Write an algorithm to compute $L_1 \cup L_2$ using only the basic list operations. 4

   b) Write algorithms to implement three stacks in one array. Your stack functions should not declare an overflow unless every slot in the array is used. 5

   c) What do you understand by a 1-2-3 deterministic Skip List ? Give a 1-2-3 deterministic Skip List. 6

      i) Show that at most 2N nodes are used.

      ii) Write insertion and deletion procedure.

   d) Give a non-recursive top-down implementation of AA Trees. 5

4. a) Suppose X is a node in a Binomial Tree and is not a root node. If the degree of X is D, what is the degree of the parent of X ? Justify your answer with an example. 4

   b) Write algorithms to perform Extract_Minimum and Decrease Key Operations with a Fibonacci Heap. Illustrate the working of algorithms with an example. 8

   c) What are the advantages of using doubly linked list to implement root list and sibling list of a Fibonacci Heap ? 3

   d) Explain zig, zig-zag and zig-zig rotations in a splay tree. 5

5. a) Discuss the importance of the topological sort on a graph. If a stack is used instead of a queue for the topological sort algorithm, does a different ordering result ? Why might one data structure give a "better" answer ? 5

   b) Develop an efficient algorithm to solve single source shortest path problem in a graph G with no negative cost edge. 5

   c) Define the maximum flow problem. Explain the Ford-Fulkerson algorithm to solve the maximum flow problem. 5

   d) What is the significance of a Minimum Spanning Tree in a Graph G ? Explain the Kruskal's algorithm with the help of an example. 5

6. a) Explain the notion of the classes P and NP. 2

   b) What is a polynomial time reduction algorithm ? 2

   c) Write short notes on : 16

      i) Dynamic Hashing ii) NP Completeness

      iii) Greedy Algorithms iv) Divide and Conquer.

PPP-A – 25

M.E. (Computer Science and Engineering) (Semester – I) Examination, Nov./Dec. 2015
(Revised Course)
ADVANCED DATA STRUCTURES AND ALGORITHMS

Duration : 3 Hours Max. Marks : 100

Instructions : *i) Attempt **any five** questions.*
*ii) Make suitable assumptions if **required**.*

1. a) Let f(n) and g(n) be asymptotically negative functions. Prove that max (f(n), g(n)) = Θ(f(n) + g(n)). 3

b) Find two functions f(n) and g(n) such that neither f(n) = O(g(n)) nor g(n) = O(f(n)). 3

c) An algorithm takes 0.5 ms for input size 200. How long will it take for input size 5000 if the running time is as following (assume low order terms are negligible) ?
   i) Linear ii) O(n log n)
   iii) Quadratic iv) Cubic 4

d) Give an efficient algorithm along with running time analysis to compute the minimum positive subsequence sum in a sequence of both positive and negative integers. 4

e) Define an ADT called Bag that contains other Bags and items. Identify atleast six data elements and six operations that can be used with ADT Bag. Suppose it is required to organise N instances of ADT Bag. Suggest two different implementation schemes, other than arrays, along with worst case running time estimation for each operation. 6

2. a) Let the height of a tree be the number of nodes in the longest path from the root to the leaves. Provide a recursive algorithm that when given a binary tree determines the height of the tree. 6
   i) What is the time and space complexities of your algorithm ?
   ii) What is the time and space complexities of your algorithm, if only AVL trees are allowed ?
   iii) What is the time and space complexities of your algorithm, if only Red-Black trees are involved ?

P.T.O.

5. a) What is the purpose of a topological sort on a graph G ? Write a linear time algorithm to perform a topological sort on a graph. 5

b) Given as input a weighted graph G = (V, E) and a distinguished vertex s, develop an algorithm to find shortest weighted path from s to every other vertex in G. 6

c) Compare the running time of Prim's algorithm and Kruskal algorithm to compute the Minimum Spanning Tree of a given graph G. 4

d) Give a method to determine articulation points in a graph G. How is it useful to decide biconnectivity of a Graph ? 5

6. a) Explain the terms : Polynomial time solvable and polynomial time verifiable. 4

b) Compare the decision problems with optimizing problems. 4

c) Write short notes on (**any three**) : 12

i) NP Hard Problems

ii) Pairing Heaps

iii) Randomized Algorithms

iv) Greedy Algorithms.

PPP-A – 54

M.E. Computer Science and Engineering (Semester – I)
Examination, Nov./Dec. 2014
ADVANCED DATA STRUCTURES AND ALGORITHMS
(Revised Course)

ation : 3 Hours Max. Marks : 100

Instructions : *1) Answer **five full** questions.*

*2) Make **suitable** assumptions **wherever** necessary.*

1. a) Write the algorithms to perform insertion, deletion and searching in skip list. **12**

   b) Into an empty two-pass min pairing heap, insert elements with priorities 200, 10, 5, 18, 6, 12, 14, 9, 8 and 22 (in this order). Show the min pairing heap following each insert. Delete the minimum element from the final min pairing heap. Show the resulting heap. **8**

2. a) Convert infix expression to postfix using an ADT, specify all the operations. Using the algorithm, convert the expression : $a + (b - c) * c + d / (e - a * b) + c$ to postfix. **10**

   b) Write the Ford-Fulkerson algorithm to get the maximal flow through network. Show the step by step implementation of Ford-Fulkerson algorithm for the given directed weighted graph G = (V, E, W) where V = {s, 1, 2, 3, 4, t},

   $E = \{ \langle s,1\rangle, \langle s, 2\rangle, \langle 2,1\rangle, \langle 1, 2\rangle \langle 1, 3\rangle, \langle 2, 4\rangle, \langle 3, 2\rangle, \langle 4, 3\rangle, \langle 3, t\rangle, \langle 4, t\rangle \}$,

   W = {16, 13, 4, 10, 12, 14, 9, 7, 20, 4}. **10**

3. a) Create the red-black tree that results by successively inserting the keys 41, 38, 8, 2, 31, 12, 19, 7 into an empty tree. Delete the keys 8, 12, 19, 2, 31, 38, 41. Explain the various rotations and write the algorithm for delete operation. **14**

   b) Find the articulation points and the biconnected components for the given undirected graph G = (V, E) where V = {A, B, C, D, E, F, G, H, I, J},

   $E = \{\langle G, B\rangle, \langle G, I\rangle, \langle B, I\rangle, \langle B, C\rangle, \langle B, E\rangle, \langle E, C\rangle, \langle E, F\rangle, \langle F, A\rangle, \langle F, D\rangle \langle A, D\rangle, \langle F, H\rangle, \langle F, J\rangle, \langle H, A\rangle, \langle J, D\rangle\}$ **6**

P.T.O.

4. a) Let A = ⟨7, 2, 4, 17, 1, 11, 6, 8, 15, 10, 20⟩. Construct a binomial heap whose keys are elements of A. Extract the node with minimum key. Write the algorithms to extract the element with minimum key. State its time complexity. 10

   b) Explain the on-line and off-line bin packing algorithms using greedy design techniques. 10

5. a) Create a B-tree of order 5 for the given data : 3, 7, 9, 23, 45, 1, 5, 14, 25, 24, 13, 11, 8, 19, 4, 31, 35, 56. Initially the B-tree is empty. Delete these keys : 4, 5, 7, 3, 14. Write the algorithms to perform insert and delete operations on the B-tree of order m. 14

   b) The rank of an element in a sequence is the no of smaller element in the sequence plus the number of equal elements that appear to its left. Write an algorithm for computing rank for a given sequence and state its time complexity. 6

PPP-A – 62

M.E. (Computer Science and Engineering) (Semester – I)

Examination, May/June 2014

ADVANCED DATA STRUCTURE AND ALGORITHM

(Revised Course)

Duration : 3 Hours | Total Marks : 100

Instructions : *i) Answer **any five** questions.*
*ii) **Make** suitable assumptions **if necessary**.*

1. a) Prove that for any two functions f(n) and g(n) we have f(n) = Θ (g(n)) iff f(n) = O(g(n)) and f(n) = Ω (g(n)). 6

   b) The rank of an element in a sequence is the no. of smaller element in the sequence plus the number of equal elements that appear to its left write an algorithm for computing rank for a given sequence and estimate the complexity of this algorithm. 5

   c) Provide algorithms for performing selection sort and Bubble sort on an array of elements. Compare these algorithms in terms of time complexity and space complexity. 6

   d) Define little on notation (o) and discuss its usage. 3

2. a) A deque is a data structure consisting of a list of items on which the following operations are possible.

   Push (x) : Insert item x on the front end of the deque.

   Pop (x) : Removes the front item from the deque and returns it.

   inject (x) : Inserts item x to rear end of the deque.

   eject (x) : Removes element from rear of the deque and returns it.

   Write algorithms to support the deque that takes 0(1) time per operation. 8

   b) Write an algorithm for performing level-order traversal in a binary search tree. 5

   c) Evaluate the bounds on the number of nodes in an AVL tree of height h stated in terms of n. 5

   d) Explain RR and RL rotations in a splay tree. 2

P.T.O.

3. a) Write an algorithm for performing insertion into a max heap.

   Trace the algorithm with following elements :

   25, 5, 3, 22, 17, 6, 55, 33, 13, 0.

   b) Explain cascading cut operation in a Fibonacci heap. Illustrate with an example.

   c) Write an algorithm to carry out deletion in a binomial heap.

4. Consider a machine shop that has m identical machines and n jobs require to be processed. The processing time required by job i is t_i. This time includes the time required to set up and remove jobs on/from a machine. A schedule is an assignment of jobs to time intervals on machine such that :

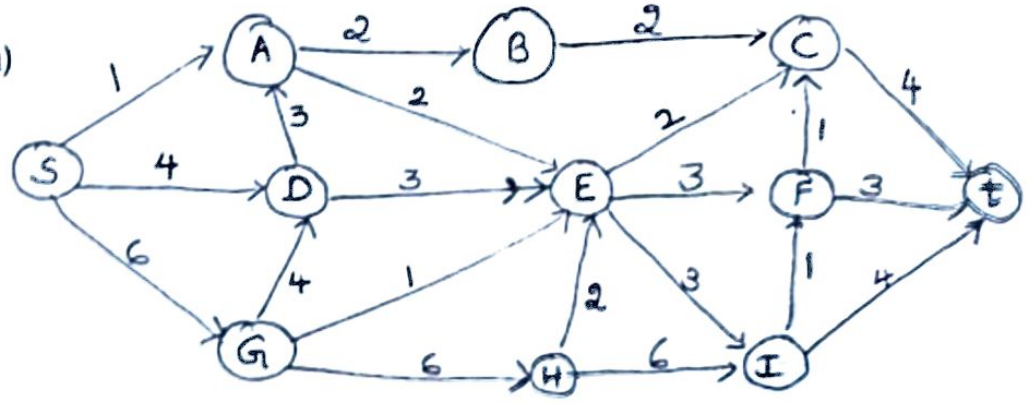
   i) No machine processes more than one job at any time.

   ii) No job is processed by more than one machine at any time.

   iii) Each job i is assigned for a total of t_i units of processing.

   Finish time or length of schedule is the time at which all jobs have completed. If non-preemptive shortest processing time first policy is employed for scheduling what would be the length of the schedule and schedule sequence. Design a solution to the above problem using appropriate data structure. Clearly state the data structure used, the solution strategy employed and give algorithms for its implementation. 2

5. a)



   i) Write an algorithm to calculate maximum flow in a given network.

   ii) Find the maximum flow in the above network.

b) A student needs to take a certain no. of courses to graduate and these courses have pre-requisites that must be followed. Assume that all courses are offered every semester and that the student can take an unlimited number of courses. Given a list of courses and their pre-requisites, use an appropriate structure to model the problem and write an algorithm to compute a schedule that requires the minimum number of semesters. 10

6. a) Given n items of sizes S_1, S_2, ... S_n were all sizes satisfy $0 < Si \leq 1$, the problem is to pack these items in the fewest number of bins given that each bin has unit capacity. Write algorithms to implement the solution using foll. approaches :

i) Best fit

ii) Next fit (in same bin as last item)

iii) First fit. 14

b) What are randomized algorithms ? Discuss their applications. How would you design a randomized version of quick sort ? 6

7. a) Write an algorithm to perform insertion operation in a Red-Black tree using the top-down procedure. 10

b) Explain skew and split operations in A-A trees. 6

c) Are 2-d trees equivalent to binary search trees. Justify your answer. 4

GPP-A – 30

M.E. Computer Science and Engineering (Semester – I) Examination, Nov./Dec. 2013

ADVANCED DATA STRUCTURE AND ALGORITHM (New Course)

ration : 3 Hours

Total Marks : 100

Instructions : *i) Answer **any five** questions.*
*ii) Make suitable assumptions if **necessary**.*

1. a) Evaluate the asymptotic time complexity of following code : 10

i) Private static int rSearch (int lastpos)

```
{  if (lastpos < 0)
     return – 1
   if (x equals (a [lastpos])
     return lastpos
  return rSearch (lastpos – 1);
}
```

ii) Public static void add (int [ ] [ ] a,

```
     int [ ] [ ] b, int [ ] [ ] c, int rows, int cols)
{
     for (int i = 0 ; i < rows ; i++)
     for (int j = 0 ; j < cols ; j++)
     c [ i ] [ j ] = a [ i ] [ j ] + b [ i ] [ j ]
}
```

b) Write algorithms using two different approaches to initialize a max heap. Compare these algorithm and discuss their complexities. 10

2. a) For a binary tree T

i) Prove that the maximum number of nodes on level i is 2^{i-1} $i \geq 1$. 2

ii) Write a function to erase the tree. Assuming t is created using a linked list data structure. The nodes on erasing are returned to free space list. 8

P.T.O.

b) There are several servers in a system each having its own priority queue of processes. If a server shutsdown, its priority queue is combined with the priority queue of its nearest neighbour. Use an appropriate data structure and write an algorithm to perform this combine operation. 10

3. a) Differentiate between static and dynamic hashing. Write an algorithm to implement search in a hash table using the linear open addressing method for handling overflows. 7

b) For a Binomial heap data structure :

i) Give its node structure

ii) Write algorithms for operations carried out on this data structure. 10

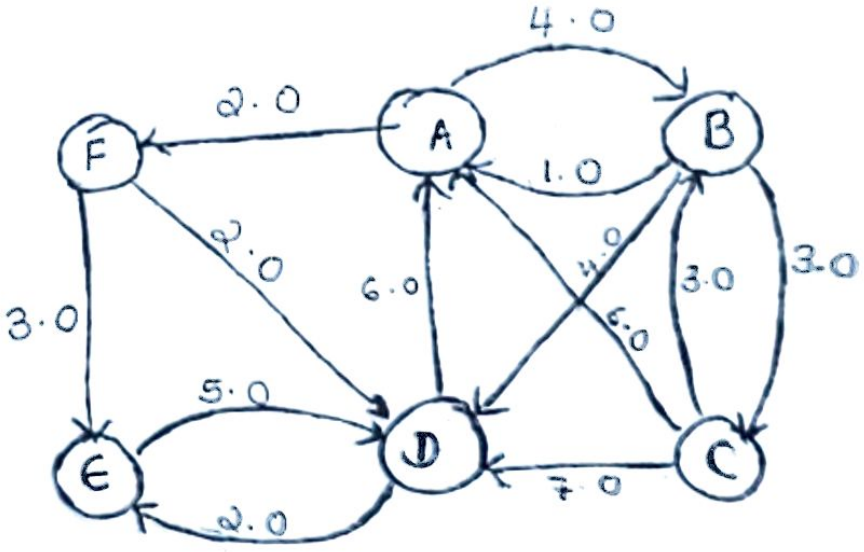
c) Write recursive algorithms for in-order, pre-order and post-order traversal in a binary tree. 3

4. a) Give algorithms for the following :

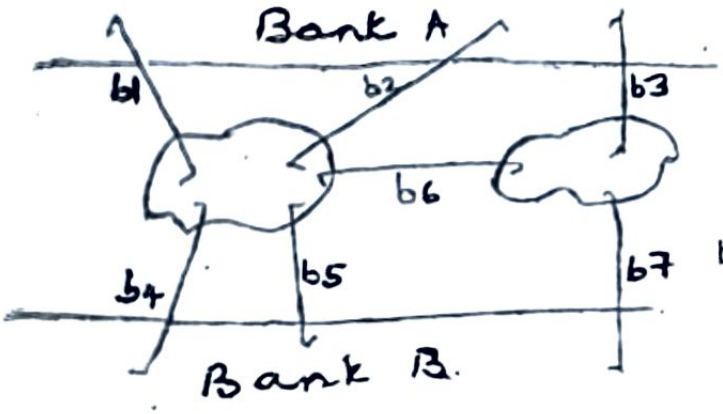
i) to find maximum spanning tree.

ii) to find argumenting path that permits maximum flow. 10

b) Write pseudocode for Dijkstra's shortest path algorithm. Trace the algorithm for the full graph. 10



a) A river with two islands in it are connected to each other by seven bridges. It is to be determined if there is a way to take; starting on any one bank of the river or on one of the islands and crossing each bridge exactly once, to return to the starting location. All bridges must be crossed, but only once.



$b_1 - b_7$: bridges

Model the problem and write algorithm for solution. **15**

b) Explain how insertion operation is carried out on a skiplist. **5**

a) Write an algorithm to carryout top-down deletion in Red-Black trees. **8**

b) Prove that the height of a Red-Black tree is at most 2 log N. **4**

c) Every Red-Black tree can be represented as a B-tree of order 4. True or false ? Justify your answer suitably with an example. **8**

PPP – A26

M.E. Computer Science and Engineering (Semester – I) Examination, May/June 2013
ADVANCED DATA STRUCTURE AND ALGORITHM

ration : 3 Hours | Total Marks : 100

Instructions : *i) Answer* ***any five*** *questions.*
ii) ***Make*** *suitable assumptions* ***if necessary.***

1. a) Prove that for any two functions f(n) and g(n) we have $f(n) = \Theta(g(n))$ iff $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$. 6

   b) The rank of an element in a sequence is the no of smaller element in the sequence plus the number of equal elements that appear to its left write an algorithm for computing rank for a given sequence and estimate the complexity of this algorithm. 5

   c) Provide algorithms for performing selection sort and Bubble sort on an array of elements. Compare these algorithms in terms of time complexity and space complexity. 6

   d) Define little on notation (o) and discuss its usage. 3

2. a) A deque is a data structure consisting of a list of items on which the following operations are possible.

   Push (x) : Insert item x on the front and of the deque.

   Pop (x) : Removes the front item from the deque and returns it.

   inject (x) : Inserts item x to rear end of the deque

   eject (x) : Removes element from rear of the deque d returns it.

   Write algorithms to support the deque that takes 0(1) time per operation. 8

   b) Write an algorithm for performing level-order traversal in a binary search tree. 5

   c) Evaluate the bounds on the number of nodes in an AVL tree of height h stated in terms of n. 5

   d) Explain RR and RL rotations in a splay tree. 2

P.T.O.

3. a) Write an algorithm for performing insertion into a max heap.

   Trace the algorithm with following elements :

   25, 5, 3, 22, 17, 6, 55, 33, 13, 0. 8

   b) Explain cascading cut operation in a Fibonacci heap. Illustrate with an example. 5

   c) Write an algorithm to carry out deletion in a binomial heap. 7

4. Consider a machine shop that has m identical machines and n jobs require to be processed. The processing time required by job i is t_i. This time includes the time required to set up and remove jobs on/from a machine. A schedule is an assignment of jobs to time intervals on machine such that :

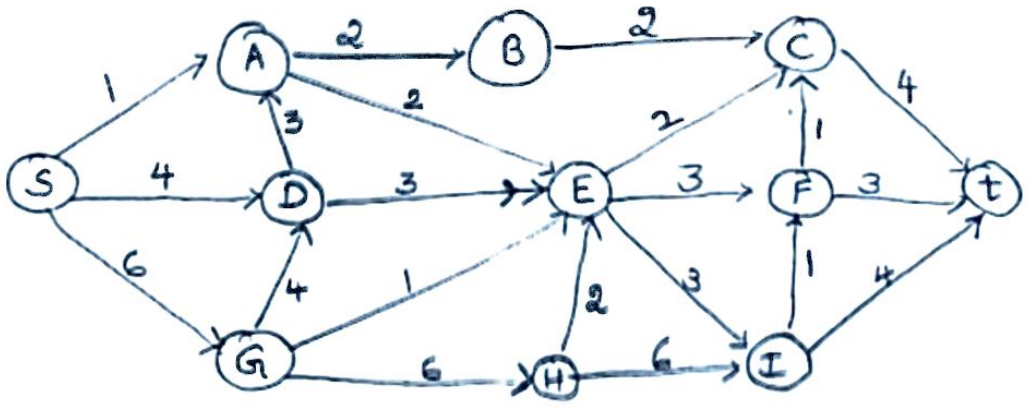
   i) No machine processes more than one job at any time.

   ii) No job is processed by more than one machine at any time.

   iii) Each job i is assigned for a total of t_i units of processing.

   Finish time or length of schedule is the time at which all jobs have completed. If non-preemptive shortest processing time first policy is employed for scheduling what would be the length of the schedule and schedule sequence. Design a solution to the above problem using appropriate data structure. Clearly state the data structure used, the solution strategy employed and give algorithms for its implementation. 20

5. a)



   i) Write an algorithm to calculate maximum flow in a given network. 5

   ii) Find the maximum flow in the above network. 5

PPP-A – 77

M.E. Computer Science and Engineering (Semester – I) Examination, Nov./Dec. 2012
ADVANCED DATA STRUCTURE AND ALGORITHM

Duration : 3 Hours Total Marks : 100

Instructions : *i) Answer **any five** questions.*
*ii) **Make** assumptions **if necessary**.*

1. a) Show that if f(n) and g(n) are monotonically increasing functions then so are functions f(n) + g(n) and f(g(n)) and if f(n) and g(n) are in addition non-negative, then f(n) · g(n) is monotonically increasing. 5

b) Evaluate time complexity using step count for foll. pieces of code :

i) Public static void insert (...)

```
{
    if (a. length < a + 1)
    throw new IllegalArgument Exception
    ("Insert.insert : array not large enough");
    // find proper place for X
    int i
    for (i = n – 1; i > = 0 & & x compareto (a[i]) < 0, i ....)
    a [i + 1] = a[i]
    a [i + 1] = a
}
```

ii) Public static void transpose (....)

```
{
    for (int i = 0; .....)
    for (int j .....)
    {
      // code here for swapping
          a[i] [j] and a[j] [i]
    }
}
```

9

c) Write an algorithm to perform bubblesort and evaluate its space and time complexity. 6

P.T.O.

2. a) Write an algorithm to list the nodes of an AVL tree in descending order of data key data members evaluate its time complexity.

   b) Write a pseudiocode algorithm for performing insertion into a β-tree.

   c) Evaluate the maximum height for a splay tree after n insertions are made into an initially empty splay tree.

3. a) Given input {4281, 3215, 2417, 6135, 4186, 7321, 989} and a hash function h(x) = x mod 10 show the result for the foll. :

      i) Separate chaining hash table

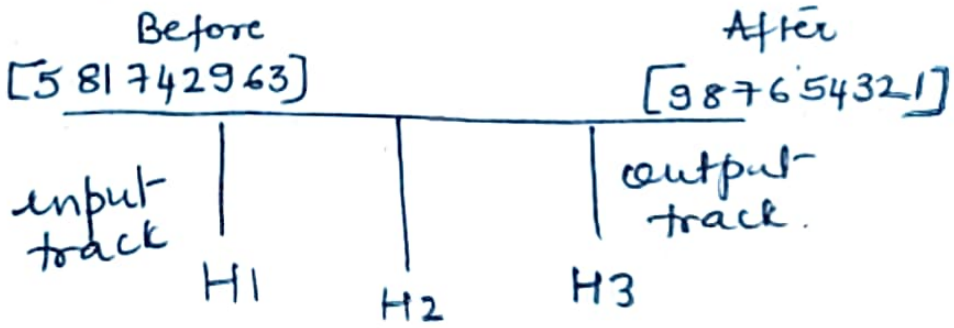
      ii) Hash table using linear probing

      iii) Hash table using quadratic probing

      iv) Hash table with second hash function h2(x) = 7 – (x mod 7).

   b) Write a program to implement extensible hashing. If the table is small enough to fit in the main memory, how does its performance compare with separate chaining and open addressing hashing.

4. A freight train has n railroad cars. Each is to be left at a different station. Assume that n stations are numbered 1 through n and that the freight train visits these stations in the order n through 1. The railroad cars are labelled by their destination. To facilitate removal of the railroad cars from the train, we must reorder the cars so that they are in the order 1 through n from front to back. When the cars are in this order, the last car is detached at each station. The cars are rearranged at a shunting yard that has an input track, an output track and a holding track.



Design a solution strategy using an appropriate data structure. Write algorithms to implement solution and discuss its complexity.

5. a) Write algorithms and analyze complexity of fall sort algorithms
   i) Heap sort
   ii) Merge sort. 16

   b) Explain the median of three partitioning strategy for quick sort. 4

6. a) Provide two greedy solutions with pseudo code to find the minimum spanning tree in an undirected graph. Compare the methods. 16

   b) Depth-first search is a generalization of pre-order traversal true or false ? Justify your answer. 4

7. a) With a neat algorithm/pseudo code explain how insertion is carried out in Treaps. 8

   b) Explain how range search is carried out in 2D-trees. 8

   c) What is pairing heap ? Discuss its application. 4

GPP – 55

M.E. (Computer Science and Engineering) (Semester – I) Examination, May/June 2018
DATA MINING (Elective – II)
(Revised Course)

Duration : 3 Hours Total Marks : 100

Instructions : *1) Answer **five full** questions.*
*2) Make **suitable** assumptions **wherever** necessary.*

1.

a) Draw a neat labeled diagram to explain the KDD process and emphasize on relation of data mining with KDD. (6)

b) Consider the following two binary vectors: (4)
x = 010101001
y = 010011000
Compute the Simple Matching Coefficient and Jaccard distance between x and y vectors.

c) What is overfitting in decision trees? Demonstrate pruning and types of pruning technique used to overcome the effect of overfitting in decision trees. (6)

d) What is numerosity reduction in data mining? Suggest one method to incorporate the same. (4)

2.

a) Why is the k-nearest neighbor called as a lazy classifier? Explain with the help of an example by comparing it with the decision tree classifier. (5)

b) Consider the following data for the 'price' attribute: (4, 8, 9, 15, 21, 21, 24, 25, 26, 28, 29, 34). Partition the same into 3 bins using: i) Equi-depth binning ii)Smoothing by bin means iii) Smoothing by bin boundaries (6)

c) Suppose S is a set of 14 examples in which one of the attributes is wind speed. The values of Wind can be Weak or Strong. The classification of these 14 examples are 9 YES and 5 NO. For attribute Wind, suppose there are 8 occurrences of Wind = Weak and 6 occurrences of Wind = Strong. For Wind = Weak, 6 of the examples are YES and 2 are NO. For Wind = Strong, 3 are YES and 3 are NO. Calculate Gain(S,Wind). (7)

d) With the help of an example explain hierarchical clustering and list its types. (2)

3.

a) What is a data cube? List the operations which can be performed on a data cube. (7)

b) Consider the transactions in the table given below: (4+3+2)

Transaction_ID	Items_Bought
1	{a,b,d,e}
2	{b,c,d}
3	{a,b,d,e}
4	{a,c,d,e}
5	{b,c,d,e}
6	{b,d,e}
7	{c,d}
8	{a,b,c}
9	{a,d,e}
10	{b,d}

Assume min_sup=30%.

i) Write an algorithm to generate frequent itemsets using an iterative level-wise approach based on candidate generation.

ii) Compute the percentage of frequent itemsets with the help of the above algorithm involving generation of candidate itemsets on the table 2 given above.

iii) Write an algorithm to prune frequent itemsets containing infrequent subsets.

c) Consider the following items to cluster: (9)

{2, 4, 10, 12, 3, 20, 30, 11, 25}

i) Using the k-means clustering, cluster the given items considering k = 2

ii) Write the k-means algorithm

4.

a) Discuss the steps involved in construction and mining of the fp-tree. (5)

b) Explain the significance of 'Naive' used by the Naive Bayesian Classifiers. Briefly describe the difference between Naive Bayes Classification and Bayesian Belief Networks. (6)

c) Describe how data mining (e.g. classification) could be used in an E-commerce setting to decide which customers could be contacted by email with an offer to buy a new book. Justify. (6)

d) Identify the kind of attribute? (2)

i. Height of a person

ii. State assuming values 'GA', 'BM', 'MH'

PPP-A – 29

M.E. (Computer Science and Engineering) (Semester – I) Examination, Nov./Dec. 2017
(Revised Course)
EL-1 : MACHINE LEARNING

Duration : 3 Hours | Total Marks : 100

***Instructions** : 1) **All five** questions are **compulsory**.*
*2) Make suitable assumption **if required**.*

1. A) Define Learning. Model the following learning problems in terms of the Task, Performance measure and Training Experience : **8**
   a) A chess learning problem
   b) Image recognition problem
   c) Signature Recognition problem
   d) Classification of various fruits.

   B) Derive the size of the hypothesis space for the following learning task and training data. How would the number of possible instances and possible hypothesis increase with addition of the attribute Water Current which can take on the values : Light, Moderate or Strong ? More generally, how does the number of possible instance grow with the addition of a new attribute A that takes on k possible values ? **8**

Sky	*Temp*	*Humidity*	*Wind*	*Water*	*Forecast*	*EnjoySport*
Sunny	Warm	Normal	Strong	Warm	Same	Yes
Sunny	Warm	High	Strong	Warm	Same	Yes
Rainy	Cold	High	Strong	Warm	Change	No
Sunny	Warm	High	Strong	Cool	Change	Yes

P.T.O.

PPP – 54

M.E. (Computer Science and Engineering) (Semester – I) Examination, Nov./Dec. 2016
EL – I : MACHINE LEARNING (Revised Course)

Duration : 3 Hours

Total Marks : 100

Instructions: *1) Attempt **all five** questions.*
*2) Make suitable assumptions **if required**.*

1. a) Describe what factors have to be considered for choosing the training experience while designing the learning system. 6

   b) "A less optimal method to Bayes Classifier is Gibbs algorithm". Justify. 5

   c) Explain the PAC learnable concept class. 5

   d) List the problems under which decision tree learning is best suited. 4

2. a) Consider the following sequence of positive and negative training examples describing the concept : "Japanese Economy Car". Apply candidate elimination algorithm and derive a hypothesis that describes the concept. 8

Origin	Manufacturer	Color	Decade	Type	Example Type
Japan	Honda	Blue	1980	Economy	Positive
Japan	Toyota	Green	1970	Sports	Negative
Japan	Toyota	Blue	1990	Economy	Positive
USA	Chrysler	Red	1980	Economy	Negative
Japan	Honda	White	1980	Economy	Positive
Japan	Toyota	Green	1980	Economy	Positive
Japan	Honda	Red	1990	Economy	Negative

P.T.O.

5. a) Explain the significance of 'Naive' used by the Naive Bayesian classifiers. Briefly describe the difference between Naive Bayes classification and Bayesian belief networks. 6

b) For the following learning task and training data, derive the size of the hypothesis space. How would the number of possible instances and possible hypothesis increase with addition of the attribute water current which can take on the values: Light, Moderate or Strong ? More generally, how does the number of possible instance grow with the addition of a new attribute A that takes on k possible values ? 8

Sky	Temp.	Humidity	Wind	Water	Forecast	Enjoy Sport
Sunny	Warm	Normal	Strong	Warm	Same	Yes
Sunny	Warm	High	Strong	Warm	Same	Yes
Rainy	Cold	High	Strong	Warm	Change	No
Sunny	Warm	High	Strong	Cool	Change	Yes

c) Write short notes on **(any one)** : 6

i) Pessimistic error pruning

ii) Version space representation theorem

iii) Weighted-Majority Algorithm.

PPP-A – 57

M.E. (CSE) (Semester – I) (Revised Course) Examination, Nov./Dec. 2014
MACHINE LEARNING (Elective – I)

Duration : 3 Hours

Max. Marks : 100

Instructions : *1) Answer **any five** questions.*

*2) Assume **suitable** data **wherever** necessary.*

1. a) Design a learning system for credit card approval. Discuss the various issues encountered while designing the system and represent the final design, using various components of learning. 8

b) In Candidate Elimination Algorithm, given the assumption, $c \in H$, each inductive inference performed by the algorithm can be justified deductively. Explain. 4

c) Given the version space, with both general and specific boundary, do we require enumerating each and every hypothesis from version space for classification of a given instance ? If not, what is the efficient test for classification of an instance ? 4

d) Define the following : 4

i) Version Space ii) General boundary.

2. a) State and explain version space representation theorem. 3

b) Consider the following data which represents the target concept, "Healthy Meal".

Example	Breakfast	Lunch	Dinner	Healthy Meal
D1	Upma	Thali	Pulav	Yes
D2	Vadapav	Thali	Noodles	Yes
D3	Vadapav	Burger	Pulav	No
D4	Vadapav	Dosa	Pulav	Yes
D5	Upma	Burger	Noodles	Yes
D6	Vadapav	Dosa	Noodles	No
D7	Upma	Dosa	Pulav	Yes
D8	Vadapav	Burger	Noodles	No

P.T.O.

a) Calculate the following : 5.

i) Number of instances

ii) Number of syntactically distinct hypothesis.

iii) Number of semantically distinct hypothesis

iv) Number of concepts.

b) Compute a maximally specific hypothesis using FIND-S algorithm.

c) Compute S and G boundary using Candidate Elimination Algorithm.

d) Construct a decision tree to infer a target concept "Healthy Meal". Represent the learned tree with the help of IF-THEN rules. 6

3. a) Construct a decision tree for the following :

i) 4-AND (4 input AND gate)

ii) $(A \wedge B) \vee (C \wedge D)$

b) List out the steps involved in Rule Post Pruning of decision tree learning.

c) Differentiate between Candidate Elimination algorithm and ID3 algorithm, with respect to search space and search strategy.

d) Consider the following attribute, Outlook, with one missing attribute value.

Outlook	Sunny	Rainy	Cloudy	Sunny	Sunny	Rainy	?
EnjoySport	No	No	Yes	Yes	Yes	No	No

Discuss the various methods involved in handling training examples with missing attribute values. Apply any such two methods to calculate the missing value of outlook attribute.

4. a) Gradient Descent technique used in BACKPROPAGATION algorithm is not always the most efficient method for minimizing error function. What are the alternative error minimization procedures used in neural networks ?

b) List out the common heuristics which help to alleviate local minima in BACKPROPAGATION algorithm.

c) What are the characteristics of the problems that can be addressed using neural network learning ?

5. a) Show that every consistent hypothesis in a version space is a MAP hypothesis. 6

b) Explain the Bayes optimal classifier. Suggest some alternative approaches to improve the performance of the classifier. 6

c) Draw a belief network with three causal nodes A, B and C, and two symptom nodes D and E. What are the independence relationships in the network ? Give the Conditional Probability Table for each of the node. Write a joint probability distribution. What will happen if we reverse some arrows in the network ? 8

6. a) Explain the Weighted-Majority algorithm with respect to mistake bound model of learning. 6

b) Define "PAC learning of concept class, C by a learner, L using hypothesis space, H". 4

c) What do you mean by ε-exhausting the version space ? 4

d) Compute a general bound on the number of training examples sufficient for any consistent learner to learn a target concept. 6

7. Write short notes on **any four** : 20

a) Perceptron

b) Grammatical Inference

c) Gibbs algorithm

d) Effects of reduced error pruning in decision tree learning.

e) VC dimension for neural networks.

PPP-A – 59

M.E. (Computer Science and Engineering) (Semester – I) Examination, May/June 2014 (Revised Course) MACHINE LEARNING (Elective – I)

Duration : 3 Hours | Max. Marks : 100

Instructions : *i) Attempt **any five** questions.*
*ii) Make suitable assumptions **if required**.*

1. a) Consider the problem of Internet shopping of books. Explain the steps in designing a program to learn to purchase books using Internet. List the available design choices at each step. 10

   b) Give the definition of a concept learning task in a general form. 4

   c) Explain the FIND_S algorithm. What is the inductive bias of FIND_S algorithm ? 6

2. a) Define a "more general than" relationship between two hypotheses. Show the use of this relationship for general to specific ordering of hypotheses. 5

   b) Suppose we have a version space computed with the available training examples. How can this version space be used to classify new instances with the same degree of confidence as if the target concept had been identified ? 3

   c) Consider the following data about the term exams conducted by faculty. 12

Exam	Use of Calculator	Duration	Lecturer	Term	Difficulty
1	Yes	3 hr	Ravi	Odd	Easy
2	Yes	3 hr	Ravi	Even	Difficulty
3	No	3 hr	Saini	Even	Difficulty
4	No	2 hr	Ravi	Odd	Easy
5	No	2 hr	Sanghi	Odd	Easy
6	Yes	3 hr	Ram	Odd	Easy
7	No	2 hr	Ravi	Even	Difficulty
8	Yes	3 hr	Saini	Even	Difficulty
9	Yes	2 hr	Sanghi	Even	Easy
10	Yes	3 hr	Saini	Odd	Easy
11	No	2 hr	Sanghi	Even	Easy
12	No	3 hr	Ram	Odd	Difficulty
13	Yes	2 hr	Ravi	Even	Easy
14	No	3 hr	Saini	Odd	Easy

   i) Compute a version space containing all hypotheses consistent with the above examples using Candidate Elimination Algorithm.

   ii) Construct a decision tree to learn a function that predicts the difficulty level of the exams.

P.T.O.

3. a) Construct a decision tree when one of the three coins a, b or c is a counterfeit coin. It is given that two of the coins are of equal weight and the third one is either heavier or lighter. 7

   b) Explain the ID3 algorithm with reduced error pruning. 7

   c) How to handle the training examples with missing attribute values and differing costs in a decision tree learning ? 6

4. a) Show that the maximum likelihood hypothesis is the one that minimizes sum of squared errors. 5

   b) Discuss the applicability of Bayes theorem in concept learning. 5

   c) Explain the advantages of Naïve Bayes Classifier. 4

   d) Draw a belief network for "smoking", "lung cancer", "coughing", "positive x-ray", and "black lung". What are the independence relationships in the network ? Give the Conditional Probability Table for each of the node. Write a joint probability distribution. What will happen if we reverse some arrows in the network ? 6

5. a) What do you understand by the PAC learnability ? 5

   b) Explain the Vapnic-Chervonenkis dimension with an example. 6

   c) Discuss the importance of Grammatical Inference. Suggest methods to infer a class of regular grammars from a set of strings. 9

6. Write short notes on **any four** : (4×5=20)

   a) Sample Complexity for finite hypothesis spaces

   b) The mistake bound model of learning

   c) EM algorithm

   d) Learning Bayesian Belief Networks

   e) Application of machine learning to data mining.

PPP-A – 27

M.E. (Computer Science and Engineering) (Semester – I) Examination, November/December 2015
MACHINE LEARNING (Elective – I)
(Revised Course)

Duration : 3 Hours

Total Marks : 100

Instructions : 1) *Attempt* ***all five*** *questions.*
2) ***Make*** *suitable assumptions if* ***required***.

1. a) Consider a problem of a system doing stock market trend prediction. Explain how this process fits into general learning model, identifying each of the components of the model as appropriate. 5

   b) With suitable illustration explain the following solutions to over fitting in decision trees : 6

   i) Stop growing the tree

   ii) Pruning using cross-validation.

   c) State and explain version space representation theorem. 3

   d) Draw diagram and explain the modeling of an inductive system by equivalent deductive system. 6

2. a) Consider the following sequence of positive and negative training examples describing the concept "pairs of people who live in the same house". Each training example described an ordered pair of people, with each person described by their sex, hair color (black, brown or blonde), height (tall, medium or short) and nationality (US, French, German, Irish, Indian, Japanese or Portuguese). (3+3+3)

   + ((male brown tall US) (female black short US))

   + ((male brown short French) (female black short US))

   – ((female brown tall German) (female black short Indian))

   + ((male brown tall Irish) (female brown short Irish))

   Consider a hypothesis space defined over these instances, over these instances, in which each hypothesis is represented by a pair of 4-tuples and where each attribute constraint may be a specific value, "?" or "ϕ".

   i) Write the representation made by the following hypothesis :

   ((male ? tall ?) (female ? ? Japanese))

P.T.O.

ii) Write the Candidate-Elimination algorithm and trace the same for the above given training examples and hypothesis language.

iii) How many distinct hypothesis from the given hypothesis space are consistent with the following single positive training example ? Write the definition for the same

+ ((male black short Portuguese) (female blonde tall Indian))

b) What is the use of VC Dimension to determine sample complexity of infinite hypothesis space ?

c) Explain the steps involved in Rule Post pruning technique.

3. a) Consider the following training data set : (4+3+5

Attribute 1	Attribute 2	Attribute 3	Class
A	70	True	Class 1
A	90	True	Class 2
A	85	False	Class 2
A	95	False	Class 2
A	70	False	Class 1
B	90	True	Class 1
B	78	False	Class 1
B	65	True	Class 1
B	75	False	Class 1
C	80	True	Class 2
C	70	True	Class 2
C	80	False	Class 1
C	80	False	Class 1
C	96	False	Class 1

i) Calculate the gain on Attribute 1 as Gain(x_1).

ii) What is the entropy of this collection of training examples with respect to the target function classification ?

iii) Explain the main steps in construction of the ID3 decision tree using the information in i).

b) Show that any learning algorithm that minimizes the squared error between the output hypothesis predictions and the training data will output a maximum likelihood hypothesis. 8

4. a) When can you say that the version space is ϵ-exhausted with respect to concept c and instance distribution D ? Explain. 5

b) A doctor knows that pneumonia causes a fever 95% of the time. She knows that if a person is selected randomly from the population, there is a chance of the person having pneumonia. 1 in 100 people suffer from fever. You go to the doctor complaining about the symptom of having a fever (evidence). What is the probability that pneumonia is the cause of this symptom (hypothesis) ? 5

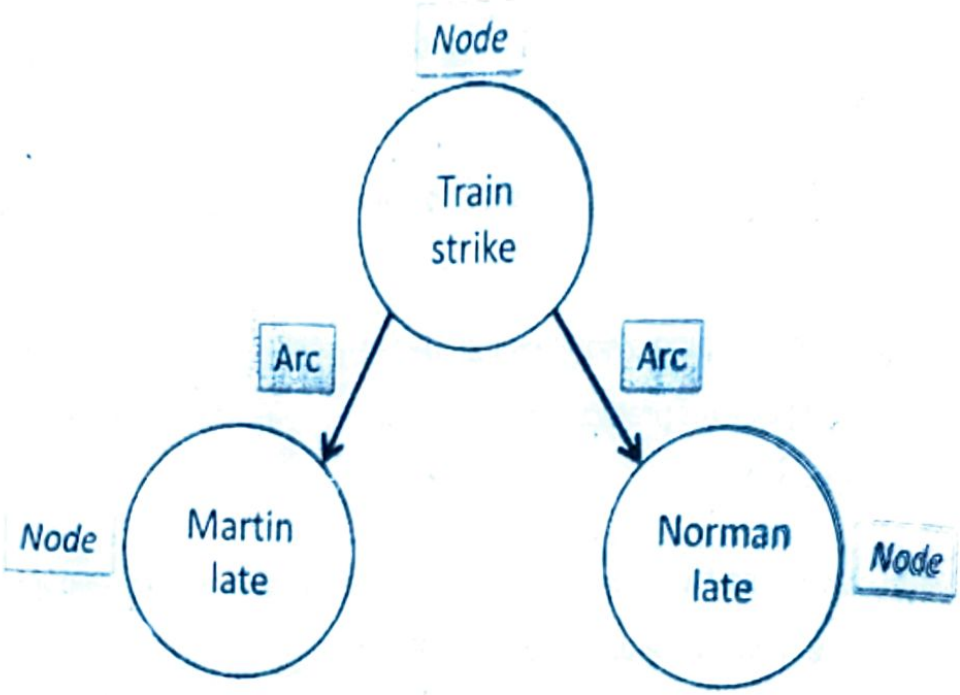
c) With the help of an appropriate illustration explain : 6

i) MAP hypothesis

ii) ML hypothesis.

d) "ID3 searches for just one consistent hypothesis, whereas the Candidate Elimination algorithm finds all consistent hypothesis". 4

5. a) Explain Bayesian belief network using the following diagram. Support your answer with the given The Conditional Probability Table (CPT). 4



Node Probability		Train strike	
		True	False
Norman late	True	0.8	0.1
	False	0.2	0.9

b) When the feature space is larger, over fitting is more likely. Do you support or deny this statement. Justify with the help of an example.

c) Write short notes on **(any two)** : (6+6)

i) Gibbs Algorithm

ii) Bayes Optimal Classifier

iii) Weighted-Majority Algorithm.

GPP-A – 38

M.E. (Computer Science and Engineering) (Semester – I) Examination, November/December 2013
MACHINE LEARNING (Elective – I) (New Course)

Duration : 3 Hours | Total Marks : 100

Instructions : *i) Attempt **any five** questions.*
*ii) **Make** suitable assumptions **if required**.*

1. a) Discuss the design issues and approaches to machine learning. 5

b) For each of the following learning problems, identify the class of tasks, performance measure and the source of experience. Suggest an appropriate target function to model each : 12

i) Handwriting recognition

ii) Robot soccer player

iii) Autonomous Mars rover

iv) Internet book shopping.

c) What do you understand by the term inductive bias ? What are the limitations of an unbiased learner ? 3

2. a) Define a version space. Explain the list-then-eliminate algorithm to compute a version space. What is the main disadvantage of this method ? 5

b) Will the candidate elimination algorithm converge to the correct hypothesis if the training data contains error ? Justify your answer. 3

c) Consider the following data : 12

Datum	Temperature	Humidity	Cloudiness	Rain ?
a	high	high	high	yes
b	low	high	low	yes
c	high	low	low	no
d	low	low	low	no
e	high	low	high	yes

P.T.O.

i) Compute the probabilities needed to complete this Bayesian belief network. Assume that all the random variables assume binary values.

ii) Identify 3 conditional independences from this diagram.

iii) Give a formula for calculating the probability that the heater is on assuming that the temperature of the room is warn.

iv) Compute the probability of being not comfortable, having heater off, warm sensor reading is positive and the temperature of the room is warm.

5. a) Discuss the importance of Grammatical Inference. Give a scheme to learn a class of regular grammar from the set of strings. 6

b) Discuss the Probably Approximately Correct (PAC) learning model. 5

c) Define the PAC learnability. 3

d) Compute a general bound on the number of training examples sufficient for any consistent leaner to learn a target concept. 6

6. Write short notes on **any four** : (5×4=20)

a) VC dimension

b) Mistake bound model of learning

c) Application of learning to knowledge discovery

d) EM algorithm

e) Bayes concept learning.

ii) Construct a decision tree to learn a function that predicts the Result for any given testing example.

Qualification	Height	Dept.	Handsome	Result
PG	Tall	EECS	Yes	Good
PhD	Average	EECS	No	Good
PG	Average	Chemistry	Yes	Bad
PhD	Average	Physics	Yes	Good
PG	Tall	Chemistry	No	Good
PhD	Average	Physics	No	Bad
PhD	Tall	Physics	Yes	Good
PG	Tall	EECS	Yes	Bad
PG	Average	Chemistry	No	Bad

3. a) Give the decision trees to represent the following Boolean functions. 6

i) $A \vee [B \wedge C]$

ii) A XOR B

b) Give an outline of ID3 algorithm to learn a decision tree that correctly classifies the given examples. 6

c) Differentiate the restriction inductive bias and preference inductive bias. 3

d) Explain the overfitting in decision tree learning. What are the approaches to avoid overfitting ? 5

4. a) What is the difference between Maximum Likelihood (ML) estimation and Maximum a Posteriori (MAP) estimation ? 2

b) What do you understand by the minimum description length principle ? Does this prove that short hypotheses are best ? 4

c) Explain the Naive Bayes Classifier as applicable in learning to classify text documents. 8

d) Draw a belief network with two causal nodes A and B, and one symptom node C. What are the independence relationships in the network ? Give the Conditional Probability Table for each of the node. Write a joint probability distribution. What will happen if we reverse some arrows in the network ? 6

5. a) Explain the Probably Approximately Correct (PAC) learning model. 6

b) Show that conjunctions of Boolean literals are PAC learnable. 6

c) Discuss the applications of VC dimension of a hypothesis space H. 8

6. Write short notes on **any four** : (4×5=20)

a) EM Algorithm

b) Mistake bound model of learning

c) Grammatical Inference

d) Application of learning to Data Mining

e) Gibbs Algorithm.

In this set of exercises, you are to make sense of Figure 1.26, which presents the performance of selected processors and a fictional one (Processor X), as reported by *www.tomshardware.com*. For each system, two benchmarks were run. One benchmark exercised the memory hierarchy, giving an indication of the speed of the memory for that system. The other benchmark, Dhrystone, is a CPU-intensive benchmark that does not exercise the memory system. Both benchmarks are displayed in order to distill the effects that different design decisions have on memory and CPU performance.

1.12 [10/10/Discussion/10/20/Discussion] <1.7> Make the following calculations on the raw data in order to explore how different measures color the conclusions one can make. (Doing these exercises will be much easier using a spreadsheet.)

a. [10] <1.8> Create a table similar to that shown in Figure 1.26, except express the results as normalized to the Pentium D for each benchmark.

b. [10] <1.9> Calculate the arithmetic mean of the performance of each processor. Use both the original performance and your normalized performance calculated in part (a).

c. [Discussion] <1.9> Given your answer from part (b), can you draw any conflicting conclusions about the relative performance of the different processors?

d. [10] <1.9> Calculate the geometric mean of the normalized performance of the dual processors and the geometric mean of the normalized performance of the single processors for the Dhrystone benchmark.

e. [20] <1.9> Plot a 2D scatter plot with the *x*-axis being Dhrystone and the *y*-axis being the memory benchmark.

f. [Discussion] <1.9> Given your plot in part (e), in what area does a dual processor gain in performance? Explain, given your knowledge of parallel processing and architecture, why these results are as they are.

Chip	# of cores	Clock frequency (MHz)	Memory performance	Dhrystone performance
Athlon 64 X2 4800+	2	2,400	3,423	20,718
Pentium EE 840	2	2,200	3,228	18,893
Pentium D 820	2	3,000	3,000	15,220
Athlon 64 X2 3800+	2	3,200	2,941	17,129
Pentium 4	1	2,800	2,731	7,621
Athlon 64 3000+	1	1,800	2,953	7,628
Pentium 4 570	1	2,800	3,501	11,210
Processor X	1	3,000	7,000	5,000

Figure 1.26 Performance of several processors on two benchmarks.

Total No. of Printed Pages:02

M.E (Computer Science & Engineering) (Sem-I) (Revised Course)
EXAMINATION Nov/Dec 2019
Elective-II
Data Mining

[Duration : Three Hours] **[Total Marks:100]**

Instructions:
1) Answer **five** full questions.
2) Make **suitable** assumptions **wherever** necessary.

Q.1 a) Explain the characteristics that have to be applied to data sets in order to have a significant impact on the data mining techniques. Explain in detail with examples the various data sets. **(10)**

b) Consider the following data set for a binary class problem. **(10)**

A	B	C	Class Label
T	F	1.0	+
T	T	6.0	+
T	T	5.0	+
T	F	4.0	-
T	T	7.0	+
F	F	3.0	-
F	F	8.0	-
F	F	7.0	-
T	T	5.0	-
T	F	2.0	-

i) What is the entropy of the above data set with respect to the positive class?
ii) For C, which is a continuous attribute, compute the information gain for split 1 and split 2.
iii) Calculate the information gain when splitting on A and B. Which attribute would the decision tree induction algorithm choose?
iv) Calculate the gain in the Gini index when splitting on A and B. Which attribute would the decision tree induction algorithm choose?
v) What is the best split (between A and B) according to the classification error rate?

Q.2 a) Explain the methods that are commonly used to evaluate the performance of a classifier. **(10)**

b) Explain the characteristics of a rule – based classifier. **(03)**

c) Explain the following w.r.t. Indirect Method for Rule Extraction **(04)**
i) Rule Generation
ii) Rule ordering

d) Explain the k- Nearest Neighbor Classification Algorithm. **(03)**

i) Compute a maximally specific hypothesis using FIND-S algorithm.

ii) Compute a version space containing all hypotheses consistent with the above examples using Candidate Elimination Algorithm.

iii) Construct a decision tree to learn a function that predicts the rain.

3. a) Draw a decision tree for the problem of deciding whether to move forward at a road intersection, given that the light has just turned green.

b) We never test the same attribute twice along one path in a decision tree. Why not ?

c) Discuss the Occam's razor principle for deciding inductive bias in decision tree learning.

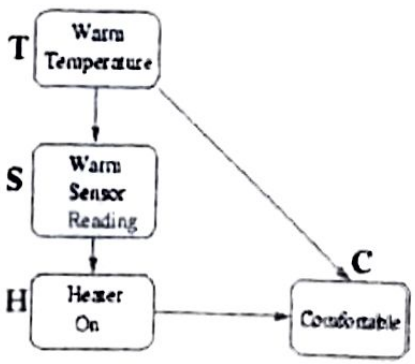
d) Explain the steps involved in Rule Post Pruning technique.

e) Suggest alternate way of selecting attributes based on some measure other than information gain in a decision tree learning algorithm.

4. a) Show that every consistent hypothesis in a version space is a MAP hypothesis.

b) Explain the Bayes optimal classifier. Suggest some alternate approaches to improve the performance of the classifier.

c) Consider the following Bayesian belief network.



PPP-A – 78

M.E. (Computer Science and Engineering) (Semester – I) Examination, Nov./Dec. 2012
MACHINE LEARNING (Elective – I)

Duration : 3 Hours

Max. Marks : 100

Instructions: i) *Attempt* ***any five*** *questions.*
ii) *Make suitable assumptions if* ***required.***

1. a) Consider the problem faced by an infant learning to speak and understand a language. Explain how this process fits into general learning model, identifying each of the components of the model as appropriate. 7

   b) What is the advantage of reducing a learning task to the problem of discovering a target function ? Suggest an appropriate target function to model a handwriting recognition problem. 7

   c) What do you understand by the term inductive bias ? Give a scheme to model and inductive system by equivalent deductive system. 6

2. a) Explain the concept learning task with the help of an example. 3

   b) Define a "more general than" relationship between two hypotheses. Show the use of this relationship to find a maximally specific hypothesis. 5

   c) Consider the following sequence of training examples. 12

      i) Give the sequence of S and G boundary sets computed by the Candidate Elimination Algorithm.

P.T.O.

C) Consider the following set of training examples : 4

Instance	Classification	A1	A2
1	+	T	T
2	+	T	T
3	–	T	F
4	+	F	F
5	–	F	T
6	–	F	T

a) What is the entropy of this collection of training examples with respect to the target function classification ?

b) What is the information gain of A2 relative to these training examples ?

2. A) Using the ID3 Algorithm construct a decision tree for the following data to decide what a person is likely to do over the weekend. 10

Weekend (Example)	Weather	Parents	Money	Decision (Category)
W1	Sunny	Yes	Rich	Cinema
W2	Sunny	No	Rich	Tennis
W3	Windy	Yes	Rich	Cinema
W4	Rainy	Yes	Poor	Cinema
W5	Rainy	No	Rich	Stay in
W6	Rainy	Yes	Poor	Cinema
W7	Windy	No	Poor	Cinema
W8	Windy	No	Rich	Shopping
W9	Windy	Yes	Rich	Cinema
W10	Sunny	No	Rich	Tennis

b) Construct a decision tree to infer a target concept "Wait for a table at the Restaurant". Represent the learned tree with the help of IF-THEN rules.

c) Prove the E-exhausting the version space theorem.

3. a) Consider the following set of training examples. (4+6)

Instance	Classification	a1	a2
1	+	T	T
2	+	T	T
3	–	T	F
4	+	F	F
5	–	F	T
6	–	F	T

i) What is the information gain of a2 relative to these training examples ? Provide the equation for calculating the information gain as well as the intermediate results.

ii) Write the ID3 decision tree algorithm. Point out how the information in i) is used.

b) Show that any learning algorithm that minimizes the squared error between the output hypothesis predictions and the training data will output a maximum likelihood hypothesis. 8

c) List various issues in Machine Learning. 2

4. a) Describe the list-then-eliminate algorithm of concept learning. 6

b) Explain the following : (6+6)

i) Gradient search to maximize likelihood in neural net.

ii) Hypothesis space search in decision tree learning.

c) What do you understand by inductive bias ? 2

b) Keys 1, 2, ... 2^k-1 are inserted in order into an initially empty AVL tree. Show that the resulting tree is perfectly balanced. 5

c) Given a B-Tree of minimum degree d : 9

i) Write an algorithm to perform insertion.

ii) Modify the insertion algorithm so that if an attempt is made to add into a node that already has 2d-1 entries a search is performed for a sibling with less than 2d-1 children before the node is split.

iii) Write an algorithm to perform deletion. When an item is deleted, is it necessary to update information in the internal nodes ? Why ?

3. a) Draw the binary min heap that results from inserting : 77, 22, 9, 68, 16, 34, 13, 8, 10, 26, 92, 6, 15 in that order into an initially empty binary min heap. Perform Extract_Min() and then Insert(7) operation on resultant binary heap. 6

b) Given two sorted lists L_1 and L_2. Write an algorithm to compute $L_1 \cap L_2$ using only the basic list operations. 4

c) Define a Red-Black Tree. Prove that the height of a Red-Black Tree is at most 2log N, and this bound cannot be substantially lowered. 5

d) Show that the k-d heap can be used to implement a double ended priority Queue. 5

4. a) What is the primary difference between Binomial Heaps and Fibonacci Heaps ? What conditions does a Binomial Tree satisfy ? 4

b) Perform the following operations in a Fibonacci Heap. Give the time complexity of each operation. 10

i) Insert the following into an empty Fibonacci Heap H1 : 33, 12, 25, 1, 66, 22, 45, 10, 6, 7, 14, 68.

ii) Insert the following into an empty Fibonacci Heap H2 : 8, 28, 11, 23, 9, 67, 5, 32, 88, 96, 13, 44, 38.

iii) Unite Fibonacci Heaps H1 and H2.

iv) Apply Extract-minimum operation two times to the Fibonacci Heap resulted in (iii).

v) Decrease the key value from 9 to 1 in the Fibonacci Heap resulted in (iv).

c) Explain the Binomial-Heap-Union algorithm with the help of an example. 6

PPP-A – 58

M.E. Computer Science and Engineering (Semester – I) (Revised Course) Examination, May/June 2014
DISTRIBUTED OPERATING SYSTEM

ration : 3 Hours | Total Marks : 100

Instructions : *1) Answer **any five full** questions.*
*2) Make suitable assumptions **wherever** necessary.*

a) Explain the Ricart-Agarwala algorithm to achieve mutual exclusion in a distributed system. 8

b) Explain the following edge-chasing algorithms for distributed deadlock detection. 8

i) Mitchell-Merritt algorithm

ii) Sinha-Natarajan algorithm.

c) Explain the two classes of hints supported by Mach operating system. 4

2. a) Explain the Schiper-Eggli-Sandoz protocol for causal ordering of messages in distributed system. 8

b) Explain the different forms of memory coherence and coherence protocols in a distributed shared memory system. 8

c) Explain the scalability issue to be addressed in the design and implementation of distributed file system. 4

3. a) Describe the architecture of a distributed file system. Also explain the data access actions in distributed system. 12

b) Explain the following processor scheduling in multiprocessor operating system :

i) Smart scheduling

ii) Affinity based scheduling. 8

4. a) Explain Singhal's token based heuristic algorithm to achieve mutual exclusion in a distributed system. 8

b) Explain the deadlock handling strategies for a distributed system. 4

c) Explain the components of load distributing algorithm in a distributed system. 8

P.T.O.

- Option 2 : Modifying the floating-point processor. The speedup of the floating-point processor is 100% faster than when it doesn't use it. Assume in this case that 60% of the computations can use the floating-point processor. Cost of this option is Rs. 600 K. Which option would you recommend ? Justify your answer quantitatively. 10

2. 1) Argue the statement as computer architect with numerical validation "Expecting the improvement of one aspect of a computer to increase overall performance by an amount proportional to the size of the improvement. 8

2) The following table shows the increase in clock rate and power of eight generations of intel processors over 28 years.

Processor	Clock rate	Power
80286 (1982)	12.5 MHz	3.3 W
80386 (1985)	16 MHz	4.1 W
80486 (1989)	25 MHz	4.9 W
Pentium (1993)	66 MHz	10.1 W
Pentium Pro (1997)	200 MHz	29.1 W
Pentium 4 Willamette (2001)	2GHz	75.3 W
Pentium 4 Prescott (2004)	3.6 GHz	103 W
Core 2 Ketsfield	2.667 GHz	95 W

a) What is the geometric mean of the ratios between consecutive generations for both clock rate and power ?

b) What is the largest relative change in clock rate and power between generations ?

c) How much larger is the clock rate and power of the last generation with respect to the first generations ?

2. 1) Argue the statement as computer architects with numerical validation "Expecting the improvement of one aspect of a computer to increase overall performance by an amount proportional to the size of the improvement.

2) The following table shows the increase in clock rate and power of eight generations of intel processors over 28 years.

Processor	Clock Rate	Power
80286 (1982)	12.5 MHz	3.3 W
80386 (1985)	16 MHz	4.1 W
80486 (1989)	25 MHz	4.9 W
Pentium (1993)	66 MHz	10.1 W
Pentium Pro (1997)	200 MHz	29.1 W
Pentium 4		
Willamette (2001)	2 GHz	75.3 W
Pentium 4		
Prescott (2004)	3.6 GHz	103 W
Core 2 Kets field	2.66 TGHz	95 W

a) What is the geometric mean of the ratios between consecutive generations for both clock rate and power ?

b) What is the largest relative change in clock rate and power between generations ?

c) How much larger is the clock rate and power of the last generation with respect to the first generations ?

4. a) What is meant by instruction level parallelism ? Explain various techniques to implement instruction level parallelism. 10

   b) Explain the different memory architectures supported by multicore systems. 10

5. a) What is dynamic scheduling ? Explain how it is used to reduce data hazards. 6

   b) Distinguish between hardware versus software speculation techniques. 4

   c) Briefly give two ways in which loop unrolling can increase performance and one in which it can decrease performance. 10

6. a) Discuss the features of any RISC Architecture you know, with its relative merits compared to CISC architecture. 8

   b) Give your view on "Multiprocessors are hard to build, but easy to program, whereas multicomputer are easy to build but hard to program". 12

7. a) How to evaluate cache performance ? Explain various cache optimization techniques. 10

   b) Consider a cache (M1) and memory (M2) hierarchy with the following characteristics :

      M1 : 32 K words, 100 ns access time

      M2 : 2 M words, 400 ns access time

      Assume eight-word cache blocks and a set size of 256 words with set-associative mapping.

      i) Show the mapping between M2 and M1.

      ii) Calculate the effective memory access time with a cache hit ratio of h = 0.95. 10

Assume that initial value of R3 is R2 + 396. Throughput this problem use the MIPS integer pipeline and assume all memory accesses are cache hits.

i) Show the timing of this instruction sequence for the MIPS pipeline without any forwarding but assuming a register read and a write in the same clock cycle "forwards" through the register file. Assume that a branch is handled by flushing the pipeline. If the memory references hit in the cache, how many cycles does this loop take to execute ? 8

ii) Show the timing of this instruction sequence for the MIPS pipeline with normal forwarding hardware. Assume that the branch is handled by predicting it as not taken. If the memory references hit in the cache, how many cycles does this loop take to execute ? 6

iii) Assume the MIPS pipeline with single-cycle delayed branch and forwarding hardware, schedule the instructions in the loop including the branch delay slot. You may re-order the instructions and modify the individual instruction operands. Show a pipeline timing diagram and compute no. of cycles needed to execute the entire loop. 6

4. a) Explain Tomasulo's algorithm. Also discuss how is it minimizes the data hazards. 10

b) Distinguish between true dependency, anti-dependency and output dependency of instructions. Give example of each of these dependencies.

i) Why should one be concerned about dependencies in pipelined execution of the programs ?

ii) If in pipelined execution of programs out of order completion of instructions is avoided, will these dependencies matter ? 10

5. a) What is cache coherence problem ? What is a snooping cache ? Explain, with example the write through and write once protocols for cache consistency. 10

b) Why are shared memory machines and distributed memory machines suited for fine grained and coarse grained problems respectively, in case of parallel computing ? 10

PPP – 51

M.E. Computer Science and Engineering (Semester – I) Examination, November/December 2016
ADVANCES IN COMPUTER ARCHITECTURE
(Revised Course)

Duration : 3 Hours

Total Marks : 100

Instructions : *a) Assume suitable data, **if** required.*
*b) Answer **any five** questions.*
*c) Draw **neat** diagrams, **whenever** necessary.*
*d) Write question numbers **legibly**, while answering.*
*e) **Use** of non-programmable calculators are **allowed.***

1. a) Explain the various techniques for reducing cache miss penalties. 10

   b) Compare VLIW versus super scalar architectures. List out merits and demerits of each architecture. 10

2. a) You have a system that contains a special processor for doing floating point operations. You have determined that 45% of your computations can use the floating point processor. The speed up of the floating point processor is 20.

      i) Over all speed up achieved by using the floating point processor.

      ii) Over all speed up achieved if you modify the compiler so that 80% of the computations can use floating point processor.

      iii) What fraction of the computations should be able to use the floating point processor in order to achieve an over all speed up of 3.25 ? 8

   b) Explain various dynamic branch prediction techniques with appropriate examples. Describe how dynamic branch prediction is different from static branch prediction technique. 12

3. Use the following code :

```
LOOP : LW   R1, 0(R2)
       Addi R1, R1, #1
       SW   R1, 0(R2)
       Addi R2, R2, #4
       Sub  R4, R3, R2
       Bnez R4, LOOP.
```

P.T.O.

b) What is Pattern Mining ? Is pattern mining is significant to data mining ? List and explain various pattern found in data warehouse. Comment on the quality of patterns found. 10

3. a) Suppose that the data for analysis include the attribute age. The age values or the data tuples are (in increasing order) : 10

13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25,

30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.

i) Use min-max normalization to transform the value 35 for age on to range [0.0, 1.0].

ii) Use Z-score normalization to transform the value 35 for age, were the standard deviation of age is 12.94 years.

iii) Use normalization by decimal scaling to transform the value 35 for age.

iv) Comment on which method you would prefer to use for the given data, giving reason as to why ?

b) What is frequent itemset mining ? Write a Apriori Algorithm and FP-Growth Algorithm. Compare above two algorithm with respect to time complexity. 10

4. a) Given a 1-dimensional value set [2, 4, 10, 12, 3, 20, 30, 11, 25]. Using K-mean algorithm cluster above data set such that k =2. 5

b) What is knowledge discovery in databases ? Explain in detail different challenges of data mining. 10

c) Explain the multi level association rule in detail. 5

PPP-A – 26

M.E. (Computer Science and Engineering) (Semester – I) Examination, November/December 2015
DATA MINING (Elective – II)
(Revised Course)

Duration : 3 Hours

Total Marks : 100

Instructions: *1) Answer **five** full questions.*
*2) Make **suitable** assumptions **wherever** necessary.*

1. a) State and explain the characteristics that have to be applied to data sets in order to have a significant impact on data mining techniques. Also explain the various data sets. 10

   b) Explain the indirect methods for rule extraction to build a rule based classifier. 10

2. a) Explain the following compact representations of frequent itemsets. 10

      i) Maximal frequent itemset

      ii) Closed frequent itemset.

   b) Explain the Apriori-Like Algorithm for sequential pattern discovery. 10

3. a) Explain the issues with the basic K-means clustering algorithm. 10

   b) Explain the key issues in hierarchical clustering. State the strengths and weaknesses of agglomerative hierarchical clustering. 10

4. a) Can we construct a data cube for multimedia data analysis ? Justify your answer. 5

   b) Explain the various text mining tasks 10

      i) Keyword based association analysis.

      ii) Document clustering analysis.

   c) "Suppose that a text retrieval system has retrieved a number of documents based on the inputs in the form of query, "how can the system be assessed for correctness and accuracy ? 5

P.T.O.

2. a) Explain the two strategies for avoiding model overfitting in context of decision tree induction. 5

b) Explain the potential causes of model overfitting in the decision tree induction. 5

c) Differentiate the following w.r.t. Rule-Based classifier 4

i) Ordered Rules v/s Unordered Rules

ii) Rule-Based Ordering Scheme v/s Class-Based Ordering Scheme.

d) Explain the RIPPER rule induction algorithm. 3

e) Explain the characteristics of Nearest Neighbor Classifier. 3

3. a) Explain the frequent itemset generation and rule generation in Apriori algorithm. 10

b) The RIPPER algorithm is an extension of an earlier algorithm called IREP. 5

Both algorithms apply the reduced-error pruning method to determine whether a rule needs to be pruned. Consider the following pair of rules :

R1 : $A \rightarrow C$

R2 : $A \wedge B \rightarrow C$

i) Suppose R1 is covered by 350 positive examples and 150 negative examples, while R2 is covered by 300 positive examples and 50 negative examples. Compute the FOIL's information gain for the rule R2 with respect to R1.

ii) Consider a validation set that contains 500 positive examples and 500 negative examples. For R1, suppose the number of positive examples covered by the rule is 200, and the number of negative examples covered by the rule is 50. For R2, suppose the number of positive examples covered by the rule is 100 and the number of negative examples is 5. Compute the v_{IREP} for both rules. Which rule does IREP prefer ?

iii) Compute v_{RIPPER} for the previous problem. Which rule does RIPPER prefer ?

c) Explain the Apriori-like algorithm for sequential pattern discovery. 5

4. a) Explain the unsupervised cluster evaluation using cohesion and separation. 10

b) Explain the Bisecting K-means clustering algorithm. 4

PPP – 55

M.E. (Computer Science and Engineering) (Semester – I) Examination, November/December 2016
DATA MINING (Elective – II)
(Revised Course)

Duration : 3 Hours | Total Marks : 100

Instructions : *1) Answer **five full** questions.*
*2) Make suitable assumptions **wherever** necessary.*

1. a) Discuss the specific challenges that motivated the development of data mining. **5**

   b) Explain with examples the different attribute types and the transformations that define the attribute levels. **5**

   c) Consider the following data set. Build a two-level decision tree. **10**

A	B	C	No. of instances	
			+	–
T	T	T	5	0
F	T	T	0	20
T	F	T	20	0
F	F	T	0	5
T	T	F	0	0
F	T	F	25	0
T	F	F	0	0
F	F	F	0	25

   i) According to the classification error rate, which attribute would be chosen as the first splitting attribute ? For each attribute, show the contingency table and the gains in the classification error rate.

   ii) Repeat for the two children of the root node.

   iii) How many instances are misclassified by the resulting decision tree ?

   iv) Repeat i), ii) and iii) using C as the splitting attribute.

P.T.O.

ME137

ME138

Total No. of Printed Pages: 02

M.E. (Computer Science & Engineering) (Sem-I) (Revised Course)
EXAMINATION Nov/Dec 2019
Advanced Data Structures and Algorithms

[Duration : Three Hours] **[Total Marks :100]**

Instructions:
1) Answer **five full questions.**
2) Make **suitable assumptions wherever necessary.**

Q.1 a) Into an empty two-pass **Min pairing Heap, insert** elements with priorities 200, 10, 5, 18, 6, 12, 14, 9, 8 and 22 **(in this order). Show the Min Pairing Heapfollowing each insert.** Delete the minimum element **from the final Min Pairing Heap. Show the resulting Heap.** **(08)**

b) Create a Red black Tree **from the set of values:** J, H, I, E, K, N, C, F, A, M, G, L. **(12)** Explain the various **rotations required** to construct the Red Black Tree. Write the algorithms to perform the **insert and delete** operations.

Q.2 a) **Write the recursive multiplication algorithm** to multiply two N- digit numbers X and Y. **(08)** Show how the recursive multiplication algorithm computes XY, where X=1234 and Y=4321. **Include all recursive computations.**

b) **Explain the following** **(08)**
i. **Probabilistic primality testing algorithm**
ii. $\alpha - \beta$ **pruning**

c) **Find the Eulerian tour** for the graph $G = (V, E)$ where $V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8\}$. **(04)** $E = \{< v_6, v_1 >, < v_6, v_3 >, < v_1, v_3 >, < v_1, v_5 >, < v_3, v_8 >, < v_5, v_2 >, < v_8, v_4 >, < v_2, v_4 >, < v_1, v_2 >, < v_3, v_4 >, < v_2, v_7 >, < v_4, v_7 >\}$.

Q.3 a) Write the johnson's algorithm to find the all pair shortest path, State its time complexity. **(10)** Use the Johnson's algorithm to find the shortest path between all pairs of vertices for the given weighted directed graph $G = (V, E, W)$ where $V = \{1, 2, 3, 4, 5, 6\}$, $E = \{< 1, 2 >, < 1, 4 >, < 1, 3 >, < 2, 3 >, < 3, 4 >, < 4, 5 >, < 5, 6 >, < 5, 3 >, < 3, 6 >, < 6, 2 >\}$ and $W = \{3, 4, 2, 4, -3, 2, 3, 6, 4, 6\}$.

b) Explain the need for external sorting algorithms. State their running time complexities. **(10)**

Q.4 a) Write the algorithms to compute the best ordering of matrix multiplication. Include the algorithm to print out the actual ordering. What is the optimal way to compute $A_1, A_2, A_3, A_4 A_5, A_6$ where the dimensions of the matrices are $A_1: 5 \times 10$, $A_2: 10 \times 8$, $A_3: 8 \times 10$, $A_4: 10 \times 2$, $A_5: 2 \times 6$, $A_6: 6 \times 9$? **(12)**

Total No. of Printed Pages:2

M.E (Computer Science & Engineering) (Sem-I)
(Revised Course) EXAMINATION Nov/Dec 2019
Distributed Operating Systems

[Duration : Three Hours] **[Total Marks :100]**

Instructions :
1) Answer any 05 questions.
2)Assume necessary detail if needed and state
3) Draw neat labeled diagrams using pencil.
4) Answer the questions in the same sequence.
5) Answer every new question on a fresh page.

Q.1 a) Discuss features of Tightly coupled multiprocessor system. Support your answer with a neat diagram. List its merits and demerits. Give one example. **10**

b) Explain features of Hypercube architecture. List its merits and demerits. Support your answer with a neat diagram. **06**

c) Discuss features of following threads. **04**
i) User-level threads.
ii) Kernel level threads.

Q.2 a) Discuss following uses with reference to processor scheduling **10**
i) Cache corruption
ii) Context switching overheads

b) List and discuss 05 design issues of multiprocessor operating systems. **10**

Q.3 a) Explain the architecture of distributed system. Support your answer with a neat diagram. List and explain its motivations. **10**

b) Explain features of any two broad categories distributed systems from the following: **10**
i) Minicomputer model
ii) Work station model
iii) Processor pool model

Q.4 a) Provide and explain LAMPORT's algorithm for distributed mutual exclusion. Also provide proof of correctness. Trade this algorithm and illustrate. **10**

b) Provide and explain SUZUKI-KASAMI's broadcast algorithm for distributed mutual exclusion. List its merits and demerits. **10**

F08749487D39F86420B23A89D6E5DED4

ME135

Total No. of Printed Pages:03

M.E (Computer Science & Engineering) (Sem-I) (Revised Course)
EXAMINATION Nov/Dec 2019
Elective-I
Machine Learning

[Duration : Three Hours] **[Total Marks :100]**

Instructions:
1) Answer any five questions.
2) Make suitable assumptions and clearly state them.

Q.1 a) Explain the list-then-eliminating algorithm to compute a version space. What is the main disadvantage of this method? 6

b) Grow the decision tree for predicting SPAM for the following dataset using Information Gain as the splitting criterion. The attributes "nigeria", "rubbish", and "learning" indicate whether that particular word occurs in the document. Show your calculations at each step. 10

Nigeria	Rubbish	Learning	SPAM
1	0	0	1
0	1	0	1
0	0	0	0
1	0	1	0
0	0	0	0
1	1	0	1
0	1	1	0
1	0	0	1
0	0	0	0
1	0	0	1

c) Why is pruning required in decision tree learning? What are the various ways to perform pruning? 4

Q.2 a) Give the computer applications for which machine learning approaches seem appropriate and three for which they seem inappropriate. For the applications for which machine learning seems appropriate, provide the performance measure and the training experience. 12

b) A spam filtering system has a probability of 0.95 to classify correctly a mail as spam and 0.10 probability of giving false positives. It is estimated that 0.5% of the mails are actual spam mails. 8
   a) Suppose that the system is now given a new mail to be classified as spam/ not-spam, what is the probability that the mail will be classified as spam?
   b) Find the probability that, given a mail classified as spam by the system, the mail actually being spam.

B) Explain with an example the list-then-eliminate algorithm of concept learning. 6

C) Give three computer applications for which machine learning approaches seem appropriate and three application for which they seem inappropriate. 4

3. A) Consider the following sequence of positive and negative training examples describing the concept. "Japanese Economy Car". Apply Candidate Elimination algorithm and derive a hypothesis that describes the concept. 8

Origin	Manufacturer	Color	Decade	Type	Example Type
Japan	Honda	Blue	1980	Economy	Positive
Japan	Toyota	Green	1970	Sports	Negative
Japan	Toyota	Blue	1990	Economy	Positive
USA	Chrysler	Red	1980	Economy	Negative
Japan	Honda	White	1980	Economy	Positive
Japan	Toyota	Green	1980	Economy	Positive
Japan	Honda	Red	1990	Economy	Negative

B) Explain the concept of inductive bias. Compare the inductive bias for rote learning, find S and candidate elimination algorithm. 8

C) List and explain the factors to be considered for choosing the training experience while designing the learning system. 4

4. a) Define Baye's theorem and show how leased squared line problem can be solved using Baye's theorem to find out most likely hypothesis. 10

b) Explain MAP and ML Hypothesis with relevant examples. 4

c) Explain Baye's Optimal Classifier (BOC) and show how Gibb's algorithm can be used for approximation of BOC. 6

5. a) Provide the formal definition of Baysian Belief Network and show with an example how Baysian networks can be used for representing joint probability distribution of variables under particular domain. 6

b) Why is naïve Baysian Classification called "naïve" ? Briefly outline the major ideas of naïve Baysian classification. 8

c) Explain in detail Weighted-Majority Algorithm. 6