

Tribhuvan University
Institute of Science and Technology
SCHOOL OF MATHEMATICAL SCIENCES
Second assessment 2079

Subject: Monte Carlo Methods
Course No: MDS 607
Level: MDS/II Year/III Semester

Full Marks: 45
Pass Marks: 22.5
Time: 2hrs

Candidates are required to give their answer in their own words as far as practicable.

Attempt ALL questions.

Group A [5×3=15]

1. What is Gibbs Sampling? Explain briefly.
2. Distinguish between frequentists and Bayesian inferences.
3. Describe the meaning of "Burn-in" period.
4. Illustrate meaning of "Reparametrization".
5. Define prior distribution in light of "Bayesian Statistics".

Group B [5×6=30]

6. What is the Metropolis-Hastings algorithm and how does it differ from the Metropolis algorithm?

OR

Justify Metropolis-Hastings Algorithm to carry on any MCMC.

7. Discuss convergence criteria in Gibbs sampling.

OR

Consider bivariate normal distribution. Discuss necessary theory and Algorithm to solve this problem using Gibbs sampling.

8. Discuss the significance of Bayesian inference over other methods of statistics.
9. How do we check for convergence of a Metropolis-Hastings sampler? Discuss with an example.

10. A manufacturer claims that the shipment contains only 5% of defective items, but the inspector feels that in fact it is 10%. We have to decide whether to accept or to reject the shipment based on θ , the proportion of defective parts. Before we see the real data, let's assign a 50-50 chance to both suggested values of θ i.e. $\pi(0.05) = \pi(0.10) = 0.5$. A random sample of 20 parts has 3 defective ones. Calculate the posterior distribution of θ .

$$F(3/\theta=0.05) - F(2/\theta=0.05) = 0.0596$$

$$F(3/\theta=0.10) - F(2/\theta=0.10) = 0.1901$$

$$0.190 \times 0.15 / 0.1748 = 0.1625$$

Tribhuvan University
Institute of Science and Technology
SCHOOL OF MATHEMATICAL SCIENCES
Second assessment 2079

Subject: Research Methodology
Course No: MDS 601
Level: MDS/II Year/III Semester

Full Marks: 45
Pass Marks: 22.5
Time: 2hrs

Candidates are required to give their answer in their own words as far as practicable.

Attempt ALL questions.

Group A [5×3=15]

1. What is case study? Describe the characteristics of case study.
2. Describe the nature and purpose of historical research.
3. What is ratio scale? What are its superiorities over other scale of measurement
4. Explain how a stratified sample can reduce the standard error?
5. What is research report? What purposes does it serve?

Group B [5x6=30]

6. A manager notices the number of grievances in the organization is increasing. The manager wishes to investigate this occurrence. What research design seems appropriate for the study? Give appropriate explanation.
7. Under what conditions might be a laboratory experiment be better than a field experiment.
8. Define the term "reliability" and "validity". Describe the various types of validity.
9. What is a questionnaire? What are the most critical problems or concerns in questionnaire design?

OR

What are the sampling error and non-sampling error? How non-sampling error can be reduced in the process of data collection?

10. Explain the conventions of academic writing.

OR

What are the various methods of documenting sources? Describe the **APA** method of citation with examples

Handwritten note:
Handwritten text in the bottom right corner, possibly a signature or date.

Tribhuvan University
Institute of Science and Technology
SCHOOL OF MATHEMATICAL SCIENCES
Second assessment 2079

Subject: Techniques for Big Data
Course No: MDS 603
Level: MDS/II Year/III Semester

Full Marks: 45
Pass Marks: 22.5
Time: 2hrs

Candidates are required to give their answer in their own words as far as practicable.

Attempt ALL questions

Group A [5x3=15]

1. What is NoSQL? How is it different than SQL?
2. / Compare and Contrast HBase with HDFS.
3. What is Spark? How is it different than MapReduce?
4. / Compare Pig Latin with SQL.
5. Differentiate between Hive QL and SQL

Group B [5x6=30]

6. Explain the CAP Theorem with example.

OR

What are the different types of NoSQL Database? Explain with example.

7. Can Mongo DB be considered as an alternate of RDBMS? Compare it against SQL with example.
8. How could SQL queries be executed in Spark. Explain with example.
9. What are the different execution modes in Apache Pig? Explain.
10. Explain about Hive Shell, Hive Services and Hive Metastore.

OR

Explain the user defined functions in Apache Pig and Apache Hive.

Tribhuvan University
Institute of Science and Technology
SCHOOL OF MATHEMATICAL SCIENCES
Second assessment 2079

Subject: Advanced Data Mining
Course No: MDS 602
Level: MDS/II Year/III Semester

Full Marks: 45
Pass Marks: 22.5
Time: 2hrs

Candidates are required to give their answer in their own words as far as practicable.

Attempt ALL questions.

Group A [5×3=15]

1. What is a support vector machine (SVM)? Why does SVM work fast and well?
2. Why does ensemble method classifier give better result? Explain main types of ensemble method.
3. How is the cluster quality measured? What are the factors affecting cluster quality. Explain with an example.
4. What is anomaly? Why it is important? Explain different types of outliers that occur in a dataset.
5. Explain distance-based outlier detection approach.

Group B [5×6=30]

6. Consider the 5 datapoints shown below: P1:(1,2,3), P2:(0,1,2), P3:(3,0,5), P4:(4,1,3), P5:(5,0,1). Apply the K-means clustering algorithm, to group those data points into 2 clusters, using the Manhattan distance. Also calculate the SSE (Sum of Square Error). Suppose that the initial centroids are C1:(1,0,0) and C2:(0,1,1).

OR

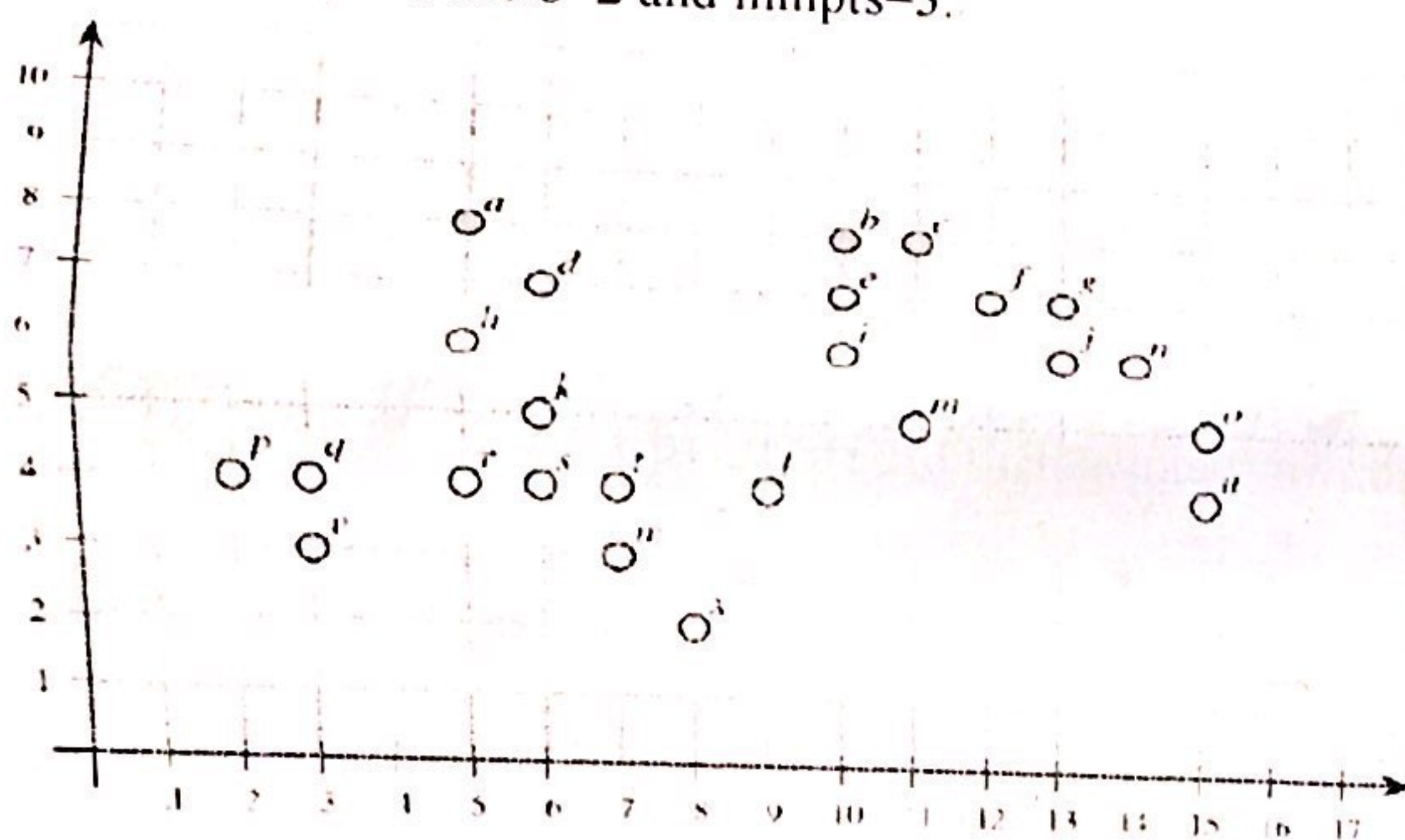
Cluster the following data and cluster the data using fuzzy C-means clustering algorithms. Assume $k=2$ and $P=2$.

X1	1	2
X2	2	3
X3	9	4
X4	10	1

7. Assume that database D is given by the table below. Follow single link technique to find clusters in D. Use any distance measure method.

	A	B	C	D
A	0	1	4	5
B		0	2	6
C			0	3
D				0

8. Consider the figure below and answer the following questions, we use the Euclidean distance between points, and the $\epsilon=2$ and $\text{minpts}=3$.



- List all the core points.
 - Is a directly reachable from d?
 - Is o density reachable from i? Show the intermediate points on the chain or the point where the chain breaks.
9. Explain different statistical based clustering method.
10. What are the approaches for density-based outlier detection. Explain Local Outlier Factor approach for outlier detection.

OR

A (0,0), B (1,0), C (1,1) and D (0,3) and $K=2$. Use LOF to detect one outlier among these 4 points. Use Manhattan distance method to measure the distance between the points.

Tribhuvan University
Institute of Science and Technology
SCHOOL OF MATHEMATICAL SCIENCES
Second assessment 2079

Subject: Decision Analysis
Course No: MDS 606
Level: MDS/II Year/III Semester

Full Marks: 45
Pass Marks: 22.5
Time: 2hrs

Candidates are required to give their answer in their own words as far as practicable.

Attempt ALL questions.

Group A [5×3=15]

✓ What is goal programming problem? Distinguish between LPP and GP.

2. Determine the saddle point solution, the associated pure strategies and the value of the game for the following game. The payoff matrix for player A is given by

Player A	Player B			
	B ₁	B ₂	B ₃	B ₄
A ₁	8	6	2	8
A ₂	8	9	4	5
A ₃	7	5	3	5

3. An office equipment manufactures produces two kinds of products chair and lamp. Production of either requires one hour of production capacity in the plant. The plant has a maximum production capacity of 50 hours per week because of the limited sales capacity, the maximum number of chairs and lamps that can be sold are 6 and 8 respectively. The gross margin from sale of chair is 90 and 60 for a lamp. The plant manager desires to determine the number of units of each product that should be produced per week in consideration of the following equally ranked goals.

Goal 1: Available production capacity should be utilized as much as possible but not exceeded.

Goal 2: Sales of two products should be as much as possible.

Goal 3: Overtime should not exceed 20% of available production time.

Formulate the problem as a Goal programming.

4. Distinguish between risk appetite and risk tolerance.

5. Explain about decision making philosophies.

Group B [5×6=30]

6. Find the optimal strategies using the dominance rule for player A and B in the following game. Also obtain the value of the game.

Player A's Strategy	Player B's Strategy		
	B ₁	B ₂	B ₃
A ₁	-3	-2	4
A ₂	-1	4	2
A ₃	2	2	6

7. A company manufactures two products radio and transistors which must be processed through assembly and finishing departments. Assembly has 90 hours available; finishing can handle up to 72 hours of work. Manufacturing one radio requires 6 hours in assembly and 3 hours in finishing. Each transistor requires 2 hours in assembly and 4 hours in finishing. If profit is 1.20 per radio and 1.90 per transistor, determine the best combination of radios and transistors to realize a maximum profit of 2000. Formulate a problem as a GP problem and solve it.

OR

- A company manufactures two products radio and transistors which must be processed through assembly and finishing departments. Assembly has 90 hours available; finishing can handle up to 72 hours of work. Manufacturing one radio requires 6 hours in assembly and 3 hours in finishing. Each transistor requires 2 hours in assembly and 4 hours in finishing. If profit is 1.20 per radio and 1.90 per transistor, determine the best combination of radios and transistors to realize a maximum profit of 2000. Formulate a problem as a GP problem and solve it.
8. What is decision analysis? Describe the various steps of decision making processes.

OR

Describe about the different types of decision theories with suitable examples.

9. What is Enterprise Risk Management? Explain briefly about COSO: Integrated Framework and their components.
10. Use dominance rule to reduce the size of the following game to 2×2 game and hence find the optimal strategies and the value of the game.

Player A	Player B		
	B ₁	B ₂	B ₃
A ₁	2	-20	43
A ₂	3	4-1	21
A ₃	5	2	6-1

Man kin
just identity