Savitribai Phule Pune University
SY M.Sc. Data Science (2023)
Course Code : DS-605-MJP
Lab Course On Optimization Techniques and Predictive Analysis (2023-Pattern)-(Semester III)

[Duration: 3 Hours] **[Marks: 35]**

Slip 1

Q1 Perform following using Rstudio. [15 M]

a. Import demo.csv & find its dimension
b. Find the mean, median, standard deviation and variance of math, hindi and english subject
c. Find max,min and range of Math,hindi and English subject.
d. How many observations are there having age 20 to 25
e. How many observations are there having age - 24 to 45 and gender=boy

Q2) Solve the following [15 M]

a. Solve the following LPP by using graphical method

$$\text{Maximize } Z = 4x_1 + 4x_2$$

$$\text{Subject to : } x_1 + x_2 \leq 4$$
$$x_1 \leq 3$$
$$x_2 \leq 2, \quad x_j \geq 0, j = 1, 2.$$

b. Write the dual of the following primal problem.

$$\text{Minimize } Z = 5x_1 + 2x_2 + x_3$$
$$\text{subject to, } \quad 2x_1 + 3x_2 + x_3 \geq 20$$
$$-2x_1 + 6x_2 - x_3 \geq 40$$
$$12x_1 + 13x_2 + 8x_3 \geq 30$$
$$2x_1 - x_2 - 8x_3 \geq 80$$
$$\text{And } x_1, x_2, x_3 \geq 0$$

c. Solve the following game graphically. The payoff is for player *A*

d. Player *B*

Player *A*		B_1	B_2	B_3	B_4
	A_1	8	5	-7	9
	A_2	-6	6	4	-2

Q3) Viva [5M]

Savitribai Phule Pune University

SY M.Sc. Data Science (2023)

Course Code : DS-605-MJP

Lab Course On Optimization Techniques and Predictive Analysis (2023-Pattern)-(Semester III)

[Duration: 3 Hours] **[Marks: 35]**

Slip 2

Q1 Perform following using Rstudio.[15 M]

a. Import salesdata.csv
b. Find the dimension of sales data
c. Find mean, median, variance and standard deviation of order quantity
d. Find the summary of sales data
e. Find max,min,range and quantile of order quantity

Q2 Solve the following [15 M]

a. A college Principal administrating infrastructures and classes for admissions by quoting tender from several agencies. The data is given as under:

Activity	Predecessor(s)	Duration (Days)
A	-	3
B	-	4
C	A	5
D	A	6
E	C	7
F	D	8
G	B	9
H	E, F, G	3

i) Draw a project network
ii) Number the events using Fulkerson's rule
iii) Help the Principal to find the project completion time to participate in the tender.

b. Find the initial basic feasible solution of the following transportation problem:

	I	II	III	Demand
A	1	2	6	7
B	0	4	2	12
C	3	1	5	11
Supply	10	10	10	

Using (i) North West Corner rule

(ii) Least Cost method

(iii) Vogel's approximation method

c. Find the optimum solution of the following LPP by Simplex method

$$\text{Maximize} \quad Z = 10x_1 + 6x_2$$

Subject to

$$5x_1 + 3x_2 \leq 30$$

$$x_1 + 2x_2 \leq 18$$

$$x_j \geq 0, \quad j = 1, 2.$$

Q3) Viva **[5 M]**

Savitribai Phule Pune University

SY M.Sc. Data Science (2023)

Course Code : DS-605-MJP

Lab Course On Optimization Techniques and Predictive Analysis (2023-Pattern)-(Semester III)

[Duration: 3 Hours] **[Marks: 35]**

Slip 3

Q1 Perform following using Rstudio. [15 M]

a. create a dataset in R containing study hours and exam scores
b. Perform simple Linear Regression

Q2) Solve the following [15 M]

a. Write the dual of the following primal

$$\text{Minimize } Z = x_1 + \tfrac{1}{2}x_2$$

subject to, $6x_1 - 2x_2 \geq 24$

$3x_1 + 2x_2 \geq 18$

$x_1 + 3x_2 \geq 12$ and $x_1, x_2 \geq 0$

c. Consider the following transportation problem. Find IBFS by VAM

	D_1	D_2	D_3	D_4	Availability
O_1	5	8	3	6	30
O_2	4	5	7	4	50
O_3	6	2	4	6	20
Requirement	30	40	20	10	

d. Solve the following game using graphical approach.

		B			
		B_1	B_2	B_3	B_4
A	A_1	2	2	3	-1
	A_2	4	3	2	6

Q3) Viva **[5 M]**

Savitribai Phule Pune University
SY M.Sc. Data Science (2023)
Course Code : DS-605-MJP
Lab Course On Optimization Techniques and Predictive Analysis (2023-Pattern)-(Semester III)
[Duration: 3 Hours] **[Marks: 35]**

Slip 4

Q1 Perform following using Rstudio. [15 M]

a. Load iris dataset
b. Build decision tree model
c. Visualize the decision tree
d. Make Predictions

Q2) Solve the following [15 M]

a. A small project has seven activities and the time in days for each activity is given below:

Activity	A	B	C	D	E	F	G
Duration in days	6	8	3	4	6	10	3

Given that activities A and B can start at the beginning of the project. When A is completed C and D can start. An activity E can start only when B and D are finished. F can start when B, C and D are completed and it is the final activity. G can start when E is finished and it is the final activity. Draw the network and find the project completion time.

b. The productivity of operators A, B, C, D, and E on different machines P, Q, R, S, and T are given in the matrix below. Assign machine to operators of maximum productivity

Operators	P	Q	R	S	T
A	9	14	10	7	12
B	8	11	12	–	13
C	10	10	8	11	–
D	12	14	11	10	7
E	13	10	12	13	10

c. A furniture company manufactures desks and chairs. The sawing department cuts the lumber for both products, which is then sent to separate assembly departments. Assembled items are sent to the painting department for finishing. The daily capacity of the sawing department is 200 chairs or 80 desks. The chair assembly department can produce 120 chairs daily, and the desk assembly department 60 desks daily. The paint department has a daily capacity of either 150 chairs or 110 desks. Given that the profit per chair is Rs. 50 and that of a desk is Rs. 100, determine the optimal production mix for the company.

Q3) Viva [5 M]

Savitribai Phule Pune University

SY M.Sc. Data Science (2023)

Course Code : DS-605-MJP

Lab Course On Optimization Techniques and Predictive Analysis (2023-Pattern)-(Semester III)

[Duration: 3 Hours] **[Marks: 35]**

Slip 5

Q1 Perform following using Rstudio.[15 M]

a. Create dataset of "Buying behavior of computer data" that includes demographic information (such as age and income) and behavior data (such as browsing duration and whether they visited the store).
b. Build decision tree model
c. Visualize the decision tree
d. Make predictions

Q2) Solve the following [**15 M]**

a. Solve the following LPP graphically.

$$\text{Maximize } Z = 6x_1 - 4x_2$$
$$\text{Subject to } x_1 + x_2 \leq 4$$
$$x_1 + x_2 \geq 6$$
$$x_j \geq 0, \quad j = 1, 2.$$

b. A company has five jobs V, W, X, Y and Z and five machines A, B, C, D and E. The given matrix shows the return in Rs. of assigning a job to a machine. Assign the jobs to machines so as to minimize the total returns.

Machines (time in hours)

Jobs	A	B	C	D	E
V	5	11	10	12	4
W	2	4	6	3	5
X	3	12	5	14	6
Y	6	14	4	11	7
Z	7	9	8	12	5

c. Draw a network for a house construction project. The sequence of activities with their predecessors is given in following table:

Activity	Activity Code	Predecessor(s)	Duration (days)
Prepare the house plan	A	-	4
Construct the house	B	A	58
Fix the door/windows	C	B	2
Electrification	D	B	2
Paint the house	E	C	1
Polish the doors/windows	F	D	1

Q3) Viva **[5 M]**

Savitribai Phule Pune University

SY M.Sc. Data Science (2023)

Course Code : DS-605-MJP

Lab Course On Optimization Techniques and Predictive Analysis (2023-Pattern)-(Semester III)

[Duration: 3 Hours] **[Marks: 35]**

Slip 6

Q1 Perform following using Rstudio.[15 M]

a. Create a golf game dataset.
b. Perform data analysis
c. Apply ANNOVA test

Q2) Solve the following: [**15 M]**

a. Reddy Mikks produces both, interior and exterior paints from two raw materials M_1 and M_2 . The following table provides, the basic data of the problems:

	exterior paint	Tons of row materials per ton of interior paint	Maximum daily availability(tons)
M1	6	4	24
M2	1	2	6
profit (Rs.)	5	4	

A market survey indicates that, the daily demand for interior paint can not exceed that of exterior paint by more than 1 ton. Also, the Maximize imum daily demand of interior paint is 2 tons. Formulate the LPP model.

b. Solve of the following LPP by Simplex method[10 M]

Maximize $Z = 3x_1 + 4x_2$

Subject to: $x_1 - x_2 \geq 1$

$x_1 + x_2 \geq 3, x_1, x_2 \geq 0$

Q3) Viva **[5 M]**

Savitribai Phule Pune University

SY M.Sc. Data Science (2023)

Course Code : DS-605-MJP

Lab Course On Optimization Techniques and Predictive Analysis (2023-Pattern)-(Semester III)

[Duration: 3 Hours] **[Marks: 35]**

Slip 7

Q1 Perform following using Rstudio.[15 M]

a. Load mtcars dataset.
b. Perform data analysis
c. Prepare scatter plot

Q2) Solve the following: **[15 M]**

a. Explain Vogel's approximation method by obtaining initial feasible solution of the following transportation problem and optimize it [10 M].

	D_1	D_2	D_3	D_4	Supply
O_1	2	3	11	7	6
O_2	1	0	6	1	1
O_3	5	8	15	9	10
Demand	7	5	3	2	

b. Draw a network for a house construction project. The sequence of activities with their predecessors is given in following table:

Activity	Activity Code	Predecessor(s)	Duration (days)
Prepare the house plan	A	-	4
Construct the house	B	A	58
Fix the door/windows	C	B	2
Electrification	D	B	2
Paint the house	E	C	1
Polish the doors/windows	F	D	1

Q3) Viva **[5 M]**

Savitribai Phule Pune University

SY M.Sc. Data Science (2023)

Course Code: DS-605-MJP

Lab Course On Optimization Techniques and Predictive Analysis (2023-Pattern)-(Semester III)

[Duration: 3 Hours] **[Marks: 35]**

Slip 8

Q1 Perform following using Rstudio. [15 M]

a. Import demo.csv & find its dimension
b. Find the mean, median, standard deviation and variance of math, hindi and english subject
c. Find max,min and range of Math,hindi and English subject.
d. How many observations are there having age 20 to 25
e. How many observations are there having age - 24 to 45 and gender=boy

Q2) Solve the following [**15 M]**

a. Solve the following game using dominance rule

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

b. Determine an initial basic feasible solution of the following transportation prob- lem by north west corner method and optimize it.

	Bangalore	Nasik	Bhopal	Delhi	Capacity
Chennai	6	8	8	5	30
Madurai	5	11	9	7	40
Trichy	8	9	7	13	50
Demand (Units/day	35	28	32	25	

Q3) Viva **[5 M]**

Savitribai Phule Pune University
SY M.Sc. Data Science (2023)

Course Code : DS-605-MJP

Lab Course On Optimization Techniques and Predictive Analysis (2023-Pattern)-(Semester III)

[Duration: 3 Hours] **[Marks: 35]**

Slip 9

Q1 Perform following using Rstudio.[15 M]

d. Import salesdata.csv
e. Find the dimension of sales data
f. Find mean, median, variance and standard deviation of order quantity
g. Find the summary of sales data
h. Find max,min,range and quantile of order quantity

Q2) Solve the following [**15 M]**

a. Solve the following problem with x_3 and x_4 as the starting basic variables and without using any artificial variables

$$\text{Maximize} \quad Z = 2x_1 + 4x_2 + 4x_3 - 3x_4$$

$$\text{Subject to:} \quad x_1 + x_2 + x_3 = 4$$
$$x_1 + 4x_2 + x_4 = 8$$
$$x_j \geq 0, \quad j = 1, 2, 3, 4.$$

b. Find an optimum solution of the following problem.[10 M]

	D_1	D_2	D_3	D_4	Supply
O_1	5	3	6	2	19
O_2	4	7	9	1	37
O_3	3	4	7	5	34
Demand	16	18	31	25	

Q3) Viva **[5 M]**

Savitribai Phule Pune University

SY M.Sc. Data Science (2023)

Course Code : DS-605-MJP

Lab Course On Optimization Techniques and Predictive Analysis (2023-Pattern)-(Semester III)

[Duration: 3 Hours] **[Marks: 35]**

Slip 10

Q1. Perform following using Rstudio.[15 M]

a. create a dataset in R containing study hours and exam scores
b. Perform simple Linear Regression

Q2. Solve the following [**15 M]**

a. Obtain an initial basic feasible solution to the following transportation problem by using least- cost method. Find the Optimum Solution [10 M]

	D_1	D_2	D_3	Supply
O_1	9	8	5	25
O_2	6	8	4	35
O_3	7	6	9	40
demand	30	25	45	

b. Write the Dual of the following primal:

Maximize $Z = 3x_1 - x_2$

Subject to: $2x_1 + x_2 \geq 2$

$x_1 + 3x_2 \leq 2$

$x_2 \leq 4$

$x_j \geq 0, j = 1, 2.$

Q3) Viva **[5 M]**

Savitribai Phule Pune University

SY M.Sc. Data Science (2023)

Course Code : DS-605-MJP

Lab Course On Optimization Techniques and Predictive Analysis (2023-Pattern)-(Semester III)

[Duration: 3 Hours] **[Marks: 35]**

Slip 11

Q1 Perform following using Rstudio.[15 M]

a. Load iris dataset
b. Build decision tree model
c. Visualize the decision tree
d. Make Predictions

Q2) Solve the following:[15 M]

a. Write the dual of the following primal
   Maximize Z= 2x1+3x2+6x3
   Subjected to: x1+x2 ≥ 3
   x2+x3 ≥ 12
   x1 –x2+x 3≥ 10
   x1,x2,x3≥0
b. Consider the following transportation problem [10 M]

	D_1	D_2	D_3	D_4	Availability
O_1	5	8	3	6	30
O_2	4	5	7	4	50
O_3	6	2	4	6	20
Requirement	30	40	20	10	

Determine initial basic feasible solution by VAM and obtain the optimum solution.

Q3) Viva **[5 M]**

Savitribai Phule Pune University

SY M.Sc. Data Science (2023)

Course Code : DS-605-MJP

Lab Course On Optimization Techniques and Predictive Analysis (2023-Pattern)-(Semester III)

[Duration: 3 Hours] **[Marks: 35]**

Slip 12

Q1 Perform following using Rstudio.[15 M]

a. Create dataset of "Buying behavior of computer data" that includes demographic information (such as age and income) and behavior data (such as browsing duration and whether they visited the store).
b. Build decision tree model
c. Visualize the decision tree
d. Make predictions

Q2) Solve the following [15 M]

a. Solve the LPP graphically

Maximize $Z = 6x_1 - 4x_2$

Subject to: $x_1 + x_2 \leq 4$

$x_1 + x_2 \geq 6$

$x_j \geq 0, \quad j = 1, 2.$

b. Find the optimum solution of the following LPP by Simplex method

Maximize $Z = 10x_1 + 6x_2$

Subject to: $5x_1 + 3x_2 \leq 30$

$x_1 + 2x_2 \leq 18$

$x_j \geq 0, j = 1, 2$

c. There are 3 jobs A, B, and C and three machines X, Y, and Z. All the jobs can be processed on all machines. The time required for processing job on a machine is given below in the form of matrix. Make allocation to minimize the total processing time.

Machines (time in hours)

Jobs	X	Y	Z
A	11	16	21
B	20	13	17
C	13	15	12

Q3) Viva **[5 M]**

Savitribai Phule Pune University

SY M.Sc. Data Science (2023)

Course Code : DS-605-MJP

Lab Course On Optimization Techniques and Predictive Analysis (2023-Pattern)-(Semester III)

[Duration: 3 Hours] **[Marks: 35]**

Slip 13

Q1 Perform following using Rstudio.[15 M]

a. Create a golf game dataset.
b. Perform data analysis
c. Apply ANNOVA test

Q2) Solve the following [**15 M]**

a. Solve the dual of the following problem then find its optoimal solution from the solution of the dual. Does the solution of the dual offer computational advantages over solving the primal directly? [10 M]

$$\text{Minimize } Z = x_1 + \tfrac{1}{2}x_2$$

subject to,
$$6x_1 - 2x_2 \geq 24$$
$$3x_1 + 2x_2 \geq 18$$
$$x_1 + 3x_2 \geq 12$$

and $x_1, x_2 \geq 0$

b. Solve the following game graphically. The payoff is for player A

		Player B			
		B_1	B_2	B_3	B_4
Player A	A_1	8	5	-7	9
	A_2	-6	6	4	-2

Q3) Viva **[5 M]**

Savitribai Phule Pune University
SY M.Sc. Data Science (2023)
Course Code : DS-605-MJP
Lab Course On Optimization Techniques and Predictive Analysis (2023-Pattern)-(Semester III)

[Duration: 3 Hours] **[Marks: 35]**

Slip 14

Q1 Perform following using Rstudio.[15 M]

a. Load mtcars dataset.
b. Perform data analysis
c. Prepare scatter plot

Q2) Solve the following [**15 M]**

a. For the following game apply dominance rule first and then graphical method to get the solution of the reduced game.

		B				
		b_1	b_2	b_3	b_4	b_5
A	a_1	-5	5	0	8	-1
	a_2	8	-4	-1	-5	6

b. Solve the following Assignment Problem for the Maximization

	A	B	C	D
P	50	60	70	65
Q	54	69	81	61
R	89	44	69	76

c. Obtain an initial basic feasible solution to the following transportation problem by using least- cost method.

	D_1	D_2	D_3	Supply
O_1	9	8	5	25
O_2	6	8	4	35
O_3	7	6	9	40
demand	30	25	45	

Q3. Viva [5 M]