ADELINE MAKOKHA 25/12/2024

ADM NO: 191199

DSA 8205: OPTIMIZATION FOR DATA SCIENCE

i) What is assignment problem?

Is a special type of linear programming problem where assignees are being assigned to perform tasks. It involves assigning a set of tasks in a way that minimizes (or maximizes) the total cost, time or effort, while ensuring that each task is assigned to exactly one assignee, and each assignee is assigned to exactly one task.

Describe how to formulate an assignment problem.

It involves representing the problem to find an optimal assignment of assignees to tasks.

1. Define the Decision variables:

$$x_{ij} = \begin{cases} 1 & \text{if assignee } i \text{ is assigned to task } j \\ 0 & \text{otherwise} \end{cases}$$

where $i = 1, 2, ... n$ represents assignees.

$j = 1, 2, ... n$ represents tasks

2. Define the objective function.

It represents the total cost (or profit) to be minimized (or maximized)

If c_{ij} is the cost of assigning assignee i to task j, the objective function is:

$$\text{Minimize: } Z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$$

3. Define the constraints.

a) Assignment constraints.

Each assignee to be assigned to exactly one task:

$$\sum_{j=1}^{n} x_{ij} = 1 \quad \text{for all } i = 1, 2 ... n$$

Each task is assigned to exactly one assignee

$$\sum_{i=1}^{n} x_{ij} = 1 \quad \text{for all } j = 1, 2, \cdots n$$

b) Binary Constraint.

The decision variable x_{ij} can only take binary values:

$$x_{ij} \in \{0, 1\} \quad \text{for all } i \text{ and } j$$

4 Matrix Representation.

The problem can be represented using a cost matrix $C = [c_{ij}]$, where rows represents assignee and columns represent tasks.

ii) Explain what a balanced assignment problem means.

It refers to a specific type of assignment problem where the number of assignee (e.g workers,) equals to the number of tasks, and each ~~agent~~ assignee must be assigned to exactly One task and vice versa.

iii) Use Hungarian method to solve both a maximization and minimization cases of an assignment problem. Get your own example to solve.

Minimization case

A small company is planning to assign three workers (W1, W2, W3) to three critical tasks (T1, T2, T3). Each task requires specific skills and the cost of assigning each worker to a particular task is estimated based on their expertise and efficiency. The management wants to minimize the total cost of assignment while ensuring that each task is completed by exactly one worker and each worker is assigned to only one task.

	T1	T2	T3
W1	8	7	9
W2	6	5	7
W3	4	6	8

Step 1: Subtract the Row minimum.

$$\begin{bmatrix} 8-7 & 7-7 & 9-7 \\ 6-5 & 5-5 & 7-5 \\ 4-4 & 6-4 & 8-4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 2 \\ 1 & 0 & 2 \\ 0 & 2 & 4 \end{bmatrix}$$

step 2: Subtract the column Minimum

$$\begin{bmatrix} 1-0 & 0-0 & 2-2 \\ 1-0 & 0-0 & 2-2 \\ 0-0 & 2-0 & 4-2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 2 & 2 \end{bmatrix}$$

Step 3: Using horizontal and vertical lines, cover all zeros in the matrix using the fewest lines possible. If the number of lines equals to size of the matrix, proceed to step 5. Otherwise, go to step 4.

Step 4: Adjust the matrix. Find the smallest uncovered value, subtract it from uncovered elements and add it to elements at intersections of lines.

Step 5: Assign Tasks. Assign workers to tasks using the zeros in the adjusted matrix. The solution is:

Worker 1 = Task 2 W2 = T1 , W3 = T3 ;

Total cost = 7 + 6 + 8 = 21.

Maximization Case

A manufacturing company is planning to assign three workers (W1, W2, W3) to three high priority tasks (T1, T2, T3). Each task offers a specific Profit based on the expertise and efficiency of the assigned worker. The company aims to maximize the total profit while ensuring that each task is assigned to exactly one worker, and no worker is assigned to more than one task. How should the company assign the workers to tasks such that the total profit is maximized?

	T1	T2	T3
W1	10	15	20
W2	20	25	30
W3	30	35	40

Step 1: Convert to a minimization problem.
Subtract each element from the largest element in the matrix

$$\begin{bmatrix} 40-10 & 40-15 & 40-20 \\ 40-20 & 40-25 & 40-30 \\ 40-30 & 40-35 & 40-40 \end{bmatrix} = \begin{bmatrix} 30 & 25 & 20 \\ 20 & 15 & 10 \\ 10 & 5 & 0 \end{bmatrix}$$

Step 2: Subtract the Row minimum.

$$\begin{bmatrix} 30-20 & 25-20 & 20-20 \\ 20-10 & 15-10 & 10-10 \\ 10-0 & 5-0 & 0-0 \end{bmatrix} = \begin{bmatrix} 10 & 5 & 0 \\ 10 & 5 & 0 \\ 10 & 5 & 0 \end{bmatrix}$$

Step 3: Subtract the column Minimum

$$\begin{bmatrix} 10-10 & 5-5 & 0-0 \\ 10-10 & 5-5 & 0-0 \\ 10-10 & 5-5 & 0-0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Step 4: Assign worker to tasks using the zeros in the matrix
W1 = T3 , W2 = T2 , W3 = T1
Total profit 20 + 25 + 30 = 75

ii) Explain with a help of an example how to obtain an optimal solution to an assignment problem.

A logistics company has four workers (W1, W2, W3, W4) and four tasks (T1, T2, T3, T4) to complete. Each task requires specific skills and the cost of assigning each worker to a particular task is calculated based on their efficiency and resource utilization. How should the logistics company assign the workers to tasks to minimize the total cost?

	T1	T2	T3	T4
W1	9	2	7	8
W2	6	4	3	7
W3	5	8	1	8
W4	7	6	9	4

Step 1. Subtract the Row minimum.

$$\begin{bmatrix} 9-2 & 2-2 & 7-2 & 8-2 \\ 6-3 & 4-3 & 3-3 & 7-3 \\ 5-1 & 8-1 & 1-1 & 8-1 \\ 7-4 & 6-4 & 9-4 & 4-4 \end{bmatrix} = \begin{bmatrix} 7 & \boxed{0} & 5 & 6 \\ 3 & 1 & \boxed{0} & 4 \\ 4 & 7 & \boxed{0} & 7 \\ 3 & 2 & 5 & 0 \end{bmatrix}$$

Step 2: Subtract the column minimum.

$$\begin{bmatrix} 7-3 & 0-0 & 5-0 & 6-0 \\ 3-3 & 1-0 & 0-0 & 4-0 \\ 4-3 & 7-0 & 0-0 & 7-0 \\ 3-3 & 2-0 & 5-0 & 0-0 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 5 & 6 \\ \boxed{0} & 1 & 0 & 4 \\ 1 & 7 & 0 & 7 \\ 0 & 2 & 5 & \boxed{0} \end{bmatrix}$$

Step 3: Cover All zeros with minimum number of lines.

Step 4: Optimal solution

W1 = T2, W2 = T3, W3 = T1, W4 = T4

Total cost = 2 + 3 + 5 + 4 = 14

v) Clearly describe a transportation problem.

Is a type of linear programming problem that focuses on finding the optimal way to transport goods from multiple suppliers (sources) to multiple consumers (destinations) while minimizing transportation costs or maximizing profits. It ensures that supply and demand constraints are satisfied.

Explain how to formulate a transportation problem.

1. Define the Decision variables

Let x_{ij} represent the number of goods transported from source i to destination j.

i ranges from 1 to m (number of sources)
j ranges from 1 to n (number to destinations).

2. Objective Function.

Minimize the transportation cost:

$$\text{Minimize } Z = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij}$$

where: c_{ij} is the transportation cost per unit from source i to destination j.

3. Constraints:

Supply constraint. The total goods transported from a source should not exceed its supply.

Demand Constraints. The total goods received at a destination should satisfy its demand.

Non-Negativity Constraints: The decision variables must be non-negative.

vi) Explain 3 main methods of solving transportation problem

- 1. Northwest corner rule: Begin by selecting x_{11} (that is, start in the northwest corner of the transportation simplex tableau). Thereafter, if x_{ij} was the last basic variable selected, then next select $x_{i,j+1}$ (that is, move one column to the right) if source i has any supply remaining. Otherwise, next select $x_{i+1,j}$ (that is, move one row down).

2. Vogel's approximation method: For each row and column remaining under consideration, calculate its difference, which is defined as the arithmetic difference between the smallest and next-to-the-smallest unit cost c_{ij} still remaining in that row or column. (If two unit costs tie for being the smallest remaining in a row or column, then the difference is 0.) In that row or column having the largest difference, select the variable having the smallest remaining unit cost. (Ties for the largest difference, or for the smallest remaining unit cost, may be broken arbitrarily.)

3. Russell's approximization method: For each source i remaining under consideration, determine its $\bar{u}_i$, which is the largest unit cost c_{ij} still remaining in that row. For each destination column j remaining under consideration, determine its $\bar{v}_j$, which is the largest unit cost c_{ij} still remaining in that column. For each variable x_{ij} not previously selected in these rows and columns, calculate $\Delta_{ij} = c_{ij} - \bar{u}_i - \bar{v}_j$. Select the variable having the largest (in absolute terms) negative value of Δ_{ij}.

vii) Explain with a help of an example how to obtain an optimal solution to a transportation problem.

A manufacturing company needs to transport products from 2 factories (sources) to three warehouses (destinations). Each factory has a limited supply of products, and each warehouse has a specific demand that must be met. The cost of transporting one unit of product from a factory to a warehouse is provided in the transportation cost matrix. The company aims to minimize the total transportation cost while meeting the demands at all warehouses & not exceeding the supplies of the factories.

	Warehouse 1	Warehouse 2	Warehouse 3	Supply
Factory 1	2	3	1	50
Factory 2	5	4	2	60
Demand	30	40	40	

Step 1: Finding an Initial Feasible solution.

1. Start with the cell that has the lowest cost ($a_{13} = 1$):
Allocate min $(50, 40) = 40$ to x_{13}. Update supply and demand: factory 1 has $50 - 40 = 10$ left & warehouse 3's demand is fully satisfied.
2. Move to the next lowest cost ($c_{11} = 2$):
Allocate min $(10, 30) = 10$ to x_{11}. Update supply & demand: factory 1's supply is fully used and warehouse 1 needs $30 - 10 = 20$

3. Move to the next lowest cost ($C_{21} = 5$):
Allocate min $(60,20) = 20$ to x_{21}. Update supply and demand: Factory 2 has $60 - 20 = 40$ left & warehouse 1's demand is satisfied.
4. Allocate the remaining good to x_{22} ($C_{22} = 4$):
Allocate min $(40,40) = 40$ to x_{22}. Supply and demand are fully satisfied

	Warehouse 1	Warehouse 2	Warehouse 3	Supply
Factory 1	10	0	40	50
Factory 2	20	40	0	60
Demand	30	40	40	

Total cost $= Z = (10\times2) + (0\times3) + (40\times1) + (20\times5) + (40\times4) + (0\times2)$
$= 20 + ~~0~~ + 40 + 100 + 160 = 320$

Step 2: Optimization of the solution.
- Compute opportunity costs:
Assign u_i to sources and v_j to destinations such that $u_i + v_j = C_{ij}$.
for all occupied cells ($x_{ij} > 0$)
- $u_i = 0$ $v_1 = 2$ $v_3 = 1$
$u_2 = 3$, $v_2 = 1$
- Calculate opportunity costs for unoccupied cells:
$C_{12} = 3$: $u_1 + v_2 = 0 + 1 = 1$, so Opportunity Cost $= C_{12} - (u_1 + v_2) = 3 - 1 = 2$
$C_{23} = 2$: $u_2 + v_3 = 3 + 1 = 4$ so Opportunity Cost $= C_{23} - (u_2 + v_3) = 2 - 4 = -2$
- Adjust the solution
The negative opportunity cost (-2) at x_{23} indicates potential for cost reduction.

Step 3 final optimal solution:
After optimization the final allocation achieves the minimum cost
$Z = 310$

viii) Describe what a network problem entails.

It involves optimizing a system of interconnected nodes (points) and arcs (lines connecting the nodes), where the objective is to determine the best way to move resources, flow or information across the network.

- Nodes: Represent points in the network, such as cities, factories.
- Arcs - connections or paths btn nodes eg roads, pipelines.
- Flows - quantity of resources, goods or data moving through the network
- Objective - often involves minimizing costs, time or distance or maximize profit or efficiency.
- Constraints - Capacity limitation, demand requirements

Explain the rules involved in formulating a network problem

1. Represent the network using a graph.
2. Define Decision Variables - flow along each arc.
   - x_{ij} = Amount of resource transported from node i to node j.
3. Specify the objective function

   For minimization problems (e.g minimizing transportation cost or travel time)

   $$\text{Minimize } Z = \sum C_{ij} x_{ij} \quad i_{j} = \text{Arcs}$$

   For maximization problems (e.g maximizing network flow or profit)

   $$\text{Maximize } Z = \sum P_{ij} x_{ij} \quad (i,j) \in \text{Arcs}$$

4. Define constraints.

   Capacity constraints - Each arc may have a capacity limit.
5. Non-negativity. Ensure non-negativity for all decision variables

ix) How do we get a critical path and project completion time.

The Critical Path method is a project management technique used to identify the sequence of tasks that determine the minimum project completion time. The critical path is the longest path through a project network, considering task durations and dependencies.

Steps to obtain the Critical Path and Project Completion Time

1. Create a Project Network Diagram
   Represent project tasks as nodes or activities.
   Represent dependencies between tasks as arrows.
2. Determine Task Duration
   Assign estimated duration to each activity.
3. Perform Forward Pass (Earliest Start and Finish Times).
4. Perform Backward Pass (Latest start and finish times)
5. Calculate slack (Float) for Each Activity.
6. Identify the critical path. Trace the path through the network where all activities have zero slack.

Explain how to obtain the three floats of a project and their significance. Use an Example.

1. Total Float [TF]
   Represents the amount of time a task can be delayed without affecting the project's overall completion time.
2. Free float [FF)
   Indicates the amount of time a task can be delayed without delaying the start of any subsequent task.
3. Independent float (IF)
   Is the amount of time a task can be delayed without affecting its earliest start or delaying any subsequent tasks.

Example.

Task	Duration (days)	Predecessor(s)
A	4	None
B	3	A
C	5	A
D	2	B, C

step 1. Forward Pass (ES and EF calculation)

Task	Duration	ES	EF
A	4	0	4
B	3	4	7
C	5	4	9
D	2	9	11

step 2: Backward Pass (LS and LF calculation)

Task	Duration	LS	LF
A	4	0	4
B	3	4	7
C	5	4	9
D	2	9	11

step 3. Calculate floats.

1. Total floats = TF = LS − ES = LF − EF
   For all tasks: TF=0, meaning no task can be delayed without affecting the project completion.
2. free float (FF): FF = ES next − EF current
   Task A : FF = 4−4=0
   Task B = FF = 9−7=2
   Task C = FF = 9−9=0
   Task D = FF = 11−11=0
3. Independent float (IF): IF = FF − (TF − FF)
   Since TF=0 for all task IF = FF

Significance of floats
- Total float: helps identify the flexibility in the schedule and determine which tasks are critical (TF=0)
- Free float: shows the freedom of delaying a task without impacting its immediate successor(s).
- Independent float: Useful for assessing individual task flexibility without affecting other tasks.