Assignment 10

April 1, 2024 1:56 PM

MIE 376 Assignment 10 Winter 2024

In this assignment, you solve problems for Markov Decision Processes (MDP) - Average. You need to formulate the mathematical model and solve it with Python and Gurobi.

1. Submission Instructions

Submit a PDF file describing a Markov Decision Processes (MDP) model and reporting the solution to the problem instance. Also, submit a program (a Python script or a Jupyter notebook) using Gurobi to solve the problem instance. In the MDP formulation, clearly define the stages, states, and actions and formulate the Bellman equations. In the Linear Programming (LP) formulation, clearly define decision variables and state the objective function and constraints. For the problem instance, report the values of the objective and solution.

2. Problem

A patient is suffering from a chronic disease that is non-lethal but incurable. Each stage not only impacts the patient's quality of life but also incurs increasing costs for symptom relief per cycle: \$100 for mild (M), \$300 for intermediate (I), and \$500 for severe (S). To manage this condition, two treatment options are available: basic medication (B) and advanced medication (A). Basic one costs \$100 per cycle, while the advanced costs \$400 but is more effective.

Treatment Effects:

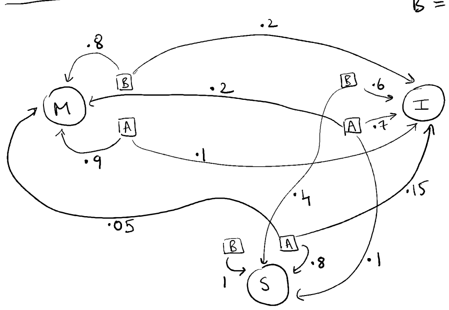
- **Mild Stage:**
  Basic Medication: 80% chance of remaining mild, 20% chance of worsening to intermediate.
  Advanced Medication: 90% chance of remaining mild, 10% chance of worsening to intermediate.
- **Intermediate Stage:**
  Basic Medication: 60% chance of staying intermediate, 40% chance of worsening to severe.
  Advanced Medication: 70% chance of staying intermediate, 20% chance of improving to mild, 10% chance of worsening to severe.
- **Severe Stage:**
  Basic Medication: 100% chance of remaining severe.
  Advanced Medication: 80% chance of staying severe, 15% chance of improving to intermediate, and 5% chance of improving to mild.

Question 1: draw the State Transition Diagram for MDP, formulate an Average Reward MDP as LP and then solve it using Python and Gurobi.] SOLUTION BELOW AND ON GUROBI ↓

Question 2: formulate the Bellman equations for Discounted Reward MDP (discount factor is 0.9), perform value iteration with Python (100 iterations to converge) and report the optimal policy and value. Note that the solution to Problem 2 is NOT necessarily the same as Problem 1.] SOLUTION BELOW AND ON PYTHON ↓

Problem 1 (Average Reward MDP)

State Transition Diagram : M = mild, I = intermediate, S = Severe
B = basic, A = advanced.



General Formulation

$$\min \sum_s \sum_x r_{sx} \pi_{sx}$$

$$\sum_s \sum_x \pi_{sx} = 1 \quad , \pi_{sx} \geq 0$$

$$\sum_x \pi_{sx} = \sum_x \sum_i \pi_{ix} P_{isx} \quad \forall s$$

↳ transition state probabilities : see diagram

π_{sx} is the state and choice

$s \in \{m, i, s\}$, ↳ $x \in \{b, a\}$

mild ↳ intermediate ↳ severe ; basic ↳ advanced

Decision variables : ($\pi_{SA}, \pi_{SB}, \pi_{MA}, \pi_{MB}, \pi_{IA}, \pi_{IB}$)

to understand which states form optimal policy.

r_{sx} = fixed cost for symptom relief + treatment

$r_{Mb} = 200$, $r_{Ib} = 400$, $r_{SB} = 600$

$r_{Mb} = 200$, $r_{Ib} = 400$, $r_{SB} = 600$

$r_{ma} = 500$, $r_{IA} = 700$, $r_{SA} = 900$

Objective

$$\min\left(r_{mb}\pi_{mb} + r_{ma}\pi_{ma} + r_{ib}\pi_{ib} + r_{ia}\pi_{ia} + r_{sb}\pi_{sb} + r_{sa}\pi_{sa}\right)$$ minimize avg cost

Constraints

$\pi_{MB} + \pi_{MA} + \pi_{iB} + \pi_{iA} + \pi_{SB} + \pi_{SA} = 1$
$\pi_{MB}, \pi_{MA}, \pi_{iB}, \pi_{iA}, \pi_{SB}, \pi_{SA} \geq 0$
ensure state probabilities sum to 1, ≥ 0

$\pi_{mb} + \pi_{ma} = 0.8\pi_{mb} + 0.9\pi_{ma} + 0.2\pi_{ia} + 0.05\pi_{sa}$
$\pi_{sb} + \pi_{sa} = \pi_{sb} + 0.8\pi_{sa} + 0.1\pi_{ia} + 0.4\pi_{ib}$
Prob of being in state = prob of going to state

$\pi_{ib} + \pi_{ia} = 0.6\pi_{ib} + 0.7\pi_{ia} + 0.15\pi_{sa} + 0.1\pi_{ma} + 0.2\pi_{mb}$ Not needed extra constraint.

Problem Instance Solution (using Gurobi)

$\pi_{MB} = 0.428571$
$\pi_{MA} = 0$
$\pi_{iB} = 0$
$\pi_{iA} = 0.380952$
$\pi_{SB} = 0$
$\pi_{SA} = 0.190476$

Decision variable values.

optimal policy must include states that take values
so π_{MB}, π_{iA}, π_{SA}
Thus, $x^* = (0,1,1)$, ie. $x^*(m) = b$ (basic), $x^*(i) = a$ (advanced), $x^*(s) = a$

Objective minimized: \$523.809524 ~ avg cost.
optimal policy $= \{0,1,1\} = \{b,a,a\}$
i.e basic treatment if mild condition, advanced treatment if intermediate or severe condition.

optimal solution.

NOTE: in code: $\pi[0] = \pi_{MB}$, $\pi[1] = \pi_{MA}$
$\pi[2] = \pi_{iB}$, $\pi[3] = \pi_{iA}$
$\pi[4] = \pi_{SB}$, $\pi[5] = \pi_{SA}$
the values are mapped accordingly.

Problem 2: Discounted Reward MDP

Stages: time period (need to make decision at each stage)
States: $s \in \{m, i, s\}$ (mild, intermediate, or severe condition)
Decision: $x = \{b, a\}$ (basic or advanced treatment)
General Bellman

$$V_s = \min_x \{ r_{sx} + \beta \sum_j P_{sjx} V_j \} \quad \forall s$$

β: 0.9 discount rate
P_{sjx}: transition probabilities given

$$V_s = \min_x \{ r_{sx} + \beta \sum_j p_{sjx} V_j \} \quad \forall s$$

↳ transition probabilities given

Fixed symptom relief + treatment cost

r_{sx} : $r_{MB} = 200$, $r_{MA} = 500$, $r_{iB} = 400$, $r_{iA} = 700$, $r_{SB} = 600$, $r_{SA} = 900$

Bellman Equations

$$V_M = \min_{x \in \{b,a\}} \{ r_{MB} + 0.9(0.8 V_m + 0.2 V_I),\ r_{MA} + 0.9(0.9 V_m + 0.1 V_I) \}$$

$$V_I = \min_{x \in \{b,a\}} \{ r_{iB} + 0.9(0.6 V_I + 0.4 V_S),\ r_{iA} + 0.9(0.7 V_I + 0.2 V_m + 0.1 V_S) \}$$

$$V_S = \min_{x \in \{b,a\}} \{ r_{SB} + 0.9(V_S),\ r_{SA} + 0.9(0.8 V_S + 0.15 V_I + 0.05 V_M) \}$$

Problem Instance Solution (in code basic = 0, advanced = 1)

After 100 iterations of value iteration,

Initial guess: $V_M^{(0)}, V_I^{(0)}, V_S^{(0)} = 0$.] → Action for initial guess.

[b, b, b] ↳ ie. $x^*(m) = b$, $x^*(i) = b$, $x^*(s) = b$

Converged Values

$V_M^{(100)} = \$4173.9982$

$V_I^{(100)} = \$5381.8633$

$V_S^{(100)} = \$5999.8406$

converged values after 100 iterations

Optimal Policy

[b, a, b] → basic treatment for mild condition

$x^*(m) = b$, advanced treatment for intermediate condition

$x^*(i) = a$, basic treatment for severe condition.

$x^*(s) = b$

optimal treatment plan.