Bonus Assignment

April 9, 2024 7:44 PM

In this assignment you solve a problem regarding MDP. You need to formulate the mathematical model and solve it with Python and Gurobi.

Problem

A student is concerned about her car and does not like dents. When she drives to school, she has a choice of parking it on the street in one space, parking it on the street and taking up two spaces, or parking in the lot. If she parks on the street in one space, her car gets dented with probability 0.1. If she parks on the street and takes two spaces, the probability of a dent is 0.02 and the probability of a $15 ticket is 0.3. Parking in a lot costs $5, but the car will not get dented. If her car gets dented, she can have it repaired, in which case it is out of commission for 1 day and costs her $50 in fees and cab fares. She can also drive her car dented, but she feels that the resulting loss of value and pride is equivalent to a cost of $9 per school day. She wishes to determine the optimal policy for where to park and whether to repair the car when dented in order to minimize her (long-run) expected average cost per school day.

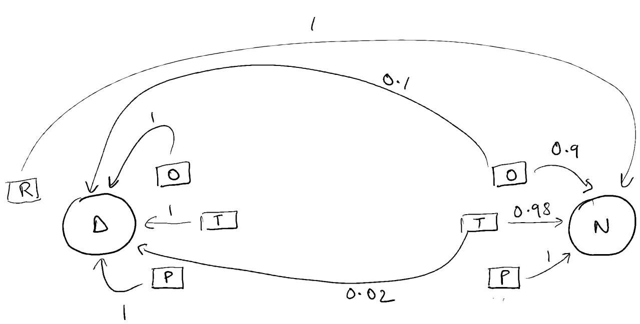
1. Draw the State Transition Diagram for MDP, formulate an Average Reward MDP as LP and then solve it using Python and Gurobi.
2. Formulate the Bellman equations for Discounted Reward MDP (discount factor is 0.9), perform value iteration with Python (100 iterations to converge) and report the optimal policy and value. Note that the solution to Problem 2 is NOT necessarily the same as Problem 1.

Instructions

Submit a PDF file describing a MDP model and a solution to the problem instance. Also, submit a program (a Python script or a Jupyter notebook) using Gurobi to solve the problem instance (if needed). For the problem instance, report the values of the objective and decision variables.

Problem 1 (Average Reward MDP

State Transition Diagram : O = one space, T = two space
P = parking lot, D = dented
N = not-dented.



General Formulation

$$\text{Min} \sum_s \sum_x r_{sx} \pi_{sx}$$

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

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

↳ transition state probabilities : see diagram

π_{sx} is the state and choice

$s \in \{D, N\}$ $\qquad$ $x \in \{O, T, P, R\}$ ↳ repair

Decision variables : $(\pi_{do}, \pi_{dt}, \pi_{dp}, \pi_{dr}, \pi_{no}, \pi_{nt}, \pi_{np})$

to understand which states form optimal

r_{sx}

$s \in \{D, N\}$ — D: dented, N: not dented

$x \in \{O, T, P, R\}$ — O: one space, T: two space, P: parking lot, R: repair

Decision variables : $(\pi_{do}, \pi_{dt}, \pi_{dp}, \pi_{dr}, \pi_{no}, \pi_{nt}, \pi_{np})$
to understand which states form optimal policy.

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

$r_{do} = 9$

$r_{dt} = 9 + 0.3(15) = 13.5$

$r_{dp} = 9 + 5 = 14$

$r_{dr} = 50$

$r_{no} = 0$

$r_{nt} = 0.3(15) = 4.5$

$r_{np} = 5$

(using expected value of ticket)
0.3 probability of \$15 ticket. $\mathbb{E}(\text{ticket cost}) = 0.3 \times \$15 = \$4.5$

(this is uncorrelated from denting)

Objective

$$\min(r_{do}\pi_{do} + r_{dt}\pi_{dt} + r_{dp}\pi_{dp} + r_{dr}\pi_{dr} + r_{no}\pi_{no} + r_{nt}\pi_{nt} + r_{np}\pi_{np})$$ minimize avg cost

Constraints

$$\pi_{do} + \pi_{dt} + \pi_{dp} + \pi_{dr} + \pi_{no} + \pi_{nt} + \pi_{np} = 1$$
$$\pi_{do}, \pi_{dt}, \pi_{dp}, \pi_{dr}, \pi_{no}, \pi_{nt}, \pi_{np} \geq 0$$
ensure state probabilities sum to 1, ≥ 0

$$\pi_{no} + \pi_{nt} + \pi_{np} = \pi_{np} + \pi_{dr} + 0.98\pi_{nt} + 0.9\pi_{no}$$
Prob of being in state = prob of going to state

Problem Instance Solution (using Gurobi)

$\pi_{do} = 0$

$\pi_{dt} = 0$

$\pi_{dp} = 0$

$\pi_{dr} = 0.0909$

$\pi_{no} = 0.9090$

$\pi_{nt} = 0$

$\pi_{np} = 0$

Decision variable values.
optimal policy must include states that take values
so π_{dr}, π_{no}
Thus, $x^*(d) = r$ (repair), $x^*(n) = 0$ (one-space).

Objective minimized : \$4.545 ~ avg cost.
optimal policy = $\{r, 0\}$
i.e repair the car if dented and park in one-space if not dented.

optimal solution.

NOTE : in code : $\pi[0] = \pi_{do}$, $\pi[1] = \pi_{dt}$
$\pi[2] = \pi_{dp}$, $\pi[3] = \pi_{dr}$
$\pi[4] = \pi_{no}$, $\pi[5] = \pi_{nt}$
$\pi[6] = \pi_{np}$

the values are mapped accordingly.

Problem 2 : Discounted Reward MDP

... to make decision at each stage)

Problem 2: Discounted Reward MDP

Stages: school days (need to make decision at each stage)

States: $s \in \{d, n\}$ (dented or not-dented)

Decision: $x = \{o, t, p, r\}$ {one-space, two-space, parkinglot, repair}

General Bellman

$\beta \to$ 0.9 discount rate

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

$P_{sjx} \hookrightarrow$ transition probabilities given

fixed symptom relief + treatment cost

r_{sx} : $r_{do} = 9$, $r_{dt} = 13.5$, $r_{dp} = 14$, $r_{dr} = 50$, $r_{no} = 0$, $r_{nt} = 4.5$, $r_{np} = 5$

Bellman Equations

assume car is repaired in the morning and out of commision for that day only, ie. usable next day.

$$V_D = \min_{x \in \{o,t,p,r\}} \{ 9 + 0.9(V_D), 13.5 + 0.9(V_D), 14 + 0.9(V_D), 50 + 0.9(V_N) \}$$

$$V_N = \min_{x \in \{o,t,p\}} \{ 0 + 0.9(0.9V_N + 0.1V_D), 4.5 + 0.9(0.98V_N + 0.02V_D), 5 + 0.9(V_N) \}$$

Problem Instance Solution (in code one-space = 0 , two-space = 1, parkinglot = 2, repair = 3)

After 100 iterations of value iteration,

Initial guess : $V_D^{(0)}, V_N^{(0)} = 0$.] $\to$ Action for initial guess .

$\hookrightarrow$ ie. $x^*(d) = 0$, $x^*(n) = 0$

Converged values

$V_D^{(100)} = \$87.1540$

$V_N^{(100)} = \$41.2824$

converged values after 100 iterations

$x^*(d) = 3 = r \to$ repair

$x^*(n) = 0 = 0 \hookrightarrow$ one-space

Optimal Policy

[r , 0] $\to$ repair car when dented and park in one-space when not-dented.

$x^*(d) = r$,

$x^*(n) = 0$,

optimal parking plan.