ELL729: Stochastic Control & Reinforcement Learning
Coding Assignment I

Maximum Marks: 6

October 13, 2020

Nuclear Power Plant Management

You are the manager at a private nuclear facility with the simple aim of maximising profits over a period of 50 days. Assume each day to be a discrete interval of time. Each day, you have to make two decisions: τ, the raise in temperature of the core reactor and ω, the increase in units of heavy water. You can make the decisions such that $\tau, \omega \in [-5, -4, \ldots, 4, 5]$. Based on the decisions taken by you, $\alpha \in [0, 1]$ i.e. the productivity rate of the plant varies such as

$$\alpha_{next} = \begin{cases} \alpha + \frac{\tau\omega}{125} + \epsilon & \text{if } \tau \geq 0, \omega \geq 0 \\ \alpha - \frac{\tau\omega}{125} + \epsilon & \text{if } \tau < 0, \omega < 0 \\ \alpha - \frac{|\tau - \omega|}{125} + \epsilon & \text{otherwise} \end{cases}$$

where τ, ω and α_{next} represent the decisions taken today and today's productivity rate. For tractability, assume $\alpha \in [0, 0.1, 0.2, \ldots, 1]$. ϵ is a categorical random variable taking values -0.05, 0 & 0.05 with equal probabilities.

Assume that for each non zero unit of temperature increase i.e. $\tau > 0$, the cost per unit increase in temperature is c_τ and for $\omega > 0$, the cost per unit increase in water is c_ω. Additionally, the equilibrium productivity of the plant is 0.3 i.e. it costs $\lambda|\alpha - 0.3|$ for maintenance of the plant at α productivity throughout the day. The plant working at full productivity can earn the plant M units of currency. Terminal cost at the end of 50 days is zero.

Technical Subtleties

- For a given α, τ and ω, if $\alpha_{next} \notin [0, 0.1, \ldots, 1]$, round to nearest value.
  Eg. if $\alpha + \frac{\tau\omega}{125} + \epsilon = 0.56$, then $\alpha_{next} = 0.6$
- Round of middle value to lower discrete value.
  Eg. if $\alpha + \frac{\tau\omega}{125} + \epsilon = 0.55$, then $\alpha_{next} = 0.5$

To Do List

- Formulate the MDP, the state transitions and the cost per stage.
- How does the noise factor affect the solution to the MDP?
- Use dynamic programming to solve the given finite horizon, finite state MDP with complete information and write down the optimal policy for at least one $c_\tau, c_\omega, \lambda$ and M such that the set of parameters do not yield a trivial solution.
- Vary parameters such as $c_\tau, c_\omega, \lambda$ and M and run your code. What is the change in total cost? Is there any change in the decisions made?
- Make a comprehensive report containing answers to all the above.

Evaluation Criteria

- Correctness of code: 3 marks
- Report: 3 marks
- Code without comments and proper indentation may not be evaluated.

Logistics

- **Deadline** is 31^{st} October midnight
- **Only** MATLAB or Python will be accepted
- Any plagiarism detected will lead to a zero in the entire programming assignments section i.e. a zero on twenty.
- Switching to project will not be allowed if plagiarism is detected.
- Libraries for DP are not allowed. Numpy is allowed and encouraged for python.
- Try to ensure that all of your code is contained in one .py or .m file only.
- All discussions pertaining to the assignment to be done on Piazza.
- Make one single zip file containing all the files and name it as 2017MTabcde_1.zip