CS 520 Final: Question 2 - Markov Decision Processes

16:198:520

> You cannot be graded on what you don't write down. This problem is to be completed individually, without any coordination with others. Complete each question to the best of your ability. If you write code to solve one of the problems, you must submit that code along with your final answers. Explain your process and algorithms. Be explicit. Answers, without evidence of the thought and process that led to them, will not earn credit.

Question 2 - Markov Decision Processes

You are a robot. One of your parts is a machine that can be in any one of ten possible states:

New, $\textbf{Used}_1$, $\textbf{Used}_2$, $\textbf{Used}_3$, $\textbf{Used}_4$, $\textbf{Used}_5$, $\textbf{Used}_6$, $\textbf{Used}_7$, $\textbf{Used}_8$, **Dead**

- In states **New** and $\textbf{Used}_1, \ldots, \textbf{Used}_8$, you can **USE** the machine. In states $\textbf{Used}_1, \ldots, \textbf{Used}_8$, **Dead**, you can **REPLACE** the machine.
- In any state, the action **REPLACE** sends the machine to state **New** with probability 1, at cost 250.
- In state **New**, **USE** gives a reward of 100, and transitions to state $\textbf{Used}_1$ with probability 1. For $i = 1, \ldots, 7$, in state $\textbf{Used}_i$, taking action **USE** yields a reward of $100 - 10 * i$, and with probability $0.1 * i$ the machine transitions to state $\textbf{Used}_{i+1}$, otherwise stays in state $\textbf{Used}_i$. In state $\textbf{Used}_8$, action **USE** yields a reward of 20, and transitions to state **Dead** with probability 0.8, otherwise stays in $\textbf{Used}_8$ with probability 0.2.
- The only action available in state **Dead** is **REPLACE**.
- At every time step, all future rewards (and costs) are discounted at a factor of $\beta = 0.9$.

a) For each of the 10 states, what is the optimal utility (long term expected discounted value) available in that state (i.e., $U^*(\textbf{state})$)?

b) What is the optimal policy that gives you this optimal utility - i.e., in each state, what is the best action to take in that state?

c) Instead of buying a new machine, a MachineSellingBot offers you the following option: you could buy a used machine, which had an equal chance of being in $\textbf{Used}_1$ and $\textbf{Used}_2$. Intuitively:

   - If the MachineSellingBot were offering you this option for free, you would never buy a new machine.
   - If the MachineSellingBot were offering you this option at a cost of 250, you would never take this option over buying a new machine.

   What is the highest price for which this used machine option would be the rational choice? i.e., what price should MachineSellingBot be selling this option at?

d) For different values of β (such that $0 < \beta < 1$), the utility or value of being in certain states will change. However, the **optimal policy** may not. Compare the optimal policy for $\beta = 0.1, 0.3, 0.5, 0.7, 0.9, 0.99$, etc. Is there a policy that is optimal for all sufficiently large β? Does this policy make sense? Explain.

Bonus: The cost of a new machine is 250. *What is the (long term discounted) value of a new machine? Determine the break-even cost of a new machine - the price above which you are operating at a net loss, and below which you are operating at a net gain, i.e., when does a new machine become so expensive this game isn't even worth playing?*