Exercises : Markov Decision Process

February 10, 2022

Exercise 1 (A simple MDP). *Let $\mathcal{X} = \{0, 1, 2, 3\}$, $\mathcal{A} = \{0, 1\}$. Let $(\boldsymbol{X}_t)_{t \in [\![1,5]\!]}$ be a controlled Markov chain, such that, if $a = 0$, it stays in its state, and if $a = 1$ it has a probability 0.5 of going 1 up (if possible, otherwise stay in place), and 0.5 of going 1 down (if possible, otherwise stay in place).*
Solve by Dynamic Programming the following optimization problem.

$$\text{Max} \qquad \mathbb{E}\Big[\sum_{t=0}^{4} \boldsymbol{X}_t^2 \mid \boldsymbol{X}_0 = 0\Big]$$

You can represent the cost-to-go and the optimal policy as matrices, each column representing one time-step.

Exercise 2. *Consider a unit that have 3 possible states : New, Working, Broken. When the unit is New at the beginning of one year, it will be Working at the beginning with probability 0.75 and Broken with probability 0.25. If it is in Working condition it can be either maintained or not. If maintained, for a cost of 2, it stay in Working condition with probability 1. If not maintainted, there is a probability of 0.5 of staying in the same condition, and of 0.5 of being Broken. If broken you can either stay this way, for a cost of 5, or repair it for a cost of 10, making it new for the next step.*
We want to manage the unit over an horizon of $T = 5$ steps, starting with a new unit. Find the policy with minimal expected cost.

Exercise 3 (Optimal stopping time). *Consider the following "push your luck" game. At turn t the player gain 1 point with probability $0 < p < 1$, and loose everything with probability $1 - p$. At the end of the turn she chooses to stop, earning her current points or continue - with the risk of loosing all. Solve the problem of maximizing expected earned points directly and by dynamic programming.*