IE5571/8571 Assignment 2

You can discuss problems with other students but must write up your own solution.

1. From Sutton and Barto exercises 3.7-3.9, 3.12, 3.13, 3.15, 3.17, 3.25-3.27, 4.5, 4.7
2. **(IE 8571 Only:)** Consider a Markov decision process $\{(S_t, R_{t+1}); t \geq 0\}$ and denote the discounted value function under policy π by
$$v_\pi(s,\gamma) = E_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{k+1} | S_0 = s\right].$$
   (a) Let T be a geometric random variable with pmf $\gamma^{k-1}(1-\gamma)$ for positive integer k, independent of the MDP process and define
$$\hat{v}_\pi(s,\gamma) = E_\pi\left[\sum_{k=0}^{T} \gamma^k R_{k+1} | S_0 = s\right].$$
   Show that $\hat{v}_\pi(s,\gamma) = v_\pi(s,\gamma)$ for all s and γ.

   (b) Now assume that T is a negative binomial random variable with pmf $(k-1)\gamma^{k-2}(1-\gamma)^2$ for integer $k \geq 2$ independent of the MDP and define
$$\tilde{v}_\pi(s,\gamma) = E_\pi\left[\sum_{k=0}^{T} \gamma^k R_{k+1} | S_0 = s\right].$$
   Show that
$$\tilde{v}_\pi(s,\gamma) = v_\pi(s,\gamma) + (1-\gamma)\frac{\partial v_\pi(s,\gamma)}{\partial \gamma}.$$
3. **(IE 8571 Only:)** One of the simpler MDP's is the so-called optimal stopping problem. In this problem there is an underlying Markov chain $\{X_n; n \geq 0\}$ taking values in a state space $\mathcal{S}$. The Markov chain is defined by a transition probability kernel
$$P(X_{n+1} \in A | X_n = x, X_{n-1}, \ldots, X_0) = P(A|x).$$
   We can also evaluate expected values with the transition probability kernel,
$$E[f(X_{n+1}) | X_n = x, X_{n-1}, \ldots, X_0] = \int_{\mathcal{S}} f(y) P(dy|x).$$
   Let $c : \mathcal{S} \to \mathbb{R}$ and $g : \mathcal{S} \to \mathbb{R}$ be given. The objective of the optimal stopping problem is to minimize the expected cost
$$E\left[g(X_\tau) + \sum_{j=0}^{\tau-1} c(X_j)\right]$$
   over stopping times τ. Here stopping time means an integer valued random variable that is not allowed to look into the future of the Markov chain, i.e., the event $\{\tau = n\}$ is decided based on information available at time n. Therefore, the goal of the optimal stopping problem is to choose an optimal time to stop the Markov chain $\{X_n; n \geq 0\}$. We will denote the class

of finite stopping times (i.e., finite with probability 1) by $\mathcal{T}$. The optimal stopping problem is then to find

$$v(x) = \inf_{\tau \in \mathcal{T}} E\left[g(X_\tau) + \sum_{j=0}^{\tau-1} c(X_j) \big| X_0 = x \right].$$

It can be shown that $v(x)$ is the maximal solution to the dynamic programming equation

$$v(x) = \min\left\{ g(x), c(x) + \int_{\mathcal{S}} v(y) P(dy|x) \right\},$$

and if the time

$$\tau^* = \inf\{n \geq 0; v(X_n) = g(X_n)\}$$

is finite then it is the optimal stopping rule. If τ^* is infinite there is no optimal stopping time.

Suppose now that $\{X_n; n \geq 0\}$ is a simple symmetric random walk on the integers, i.e., $X_{n+1} = X_n + 1$ w.p. $1/2$ and $X_{n+1} = X_n - 1$ w.p. $1/2$. We then model the price of a financial asset with $S_n = \exp(X_n)$. We are interested in finding a stopping time to price an option with discounting, i.e., solving the optimal stopping

$$v(x) = \sup_{\tau \in \mathcal{T}} E\left[\gamma^\tau \left(1 - \exp(X_\tau)\right)^+ \big| X_0 = x \right].$$

(a) Write down the DPE for this problem.

(b) It turns out that the optimal stopping time has the form

$$\inf\{n \geq 0 : v(X_n) = (1 - \exp(X_n))^+\} = \inf\{n \geq 0 : X_n \leq z\},$$

for some unknown z. Should z be positive or negative?

(c) Based on part (b), do you have any ideas about the structural form of $v(x)$? This is hard, and it's okay if you don't have an exact answer.

(d) Show that if $\gamma = 1$ then $v(x) = 1$ for all x, and argue why this implies there is no optimal stopping time.