EE807 Mathematical Foundations of Reinforcement Learning

Homework 2

Due date : 2:30 PM, April 9, 2018

1. Consider a problem of operating a machine that can be in any one of n states, denoted 1, 2, …, n. We denote by g(i) the operating cost per period when the machine is in state i, and we assume that

$$g(1) \leq g(2) \cdots \leq g(n)$$

The implication here is that state i is better than state i+1, and state 1 corresponds to a machine being in the best condition. The transition probabilities during one period of operation satisfy

$$p_{i(i+1)} > 0 \; if \; i < n,$$
$$p_{ij} = 0 \;\; if \;\; j \neq i, j \neq i+1$$

We assume that at the start of each period we know the state of the machine and we must choose one of the following two options:

(1) Let the machine operate one more period in the state it currently is

(2) Repair the machine and bring it to the best state 1 at a cost R.

We assume that the machine, once repaired, is guaranteed to stay in state 1 for one period. In subsequent periods, it may deteriorate to states j>1.

(a) Assume an infinite horizon and a discount factor $\alpha \in (0,1)$, and show that there is an optimal policy which is a threshold policy; i.e., it takes the form

replace if and only if $i \geq i^*$,

where i^* is some integer.

(b) Show that the policy iteration method, when started with a threshold policy, generates a sequence of threshold policies.

2. Consider a problem with perfect state information involving the n-dimensional linear system.

$$x_{k+1} = A_k x_k + B_k u_k + w_k \,, \qquad k = 0, \dots, N-1,$$

and a cost of function of the form

$$\underset{\substack{w_k \\ k=0,1,\ldots N-1}}{E}\left\{g_N(c'x_N) + \sum_{k=0}^{N-1} g_k(u_k)\right\},$$

where $c \in R^n$ is a given vector. Show that the DP algorithm for this problem can be carried out over a one-dimensional state space.