EE807 Mathematical Foundations of Reinforcement Learning

Homework 1

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

1. Consider the system

$$x_{k+1} = x_k + u_k + w_k, \qquad k = 0,1,2,3,$$

with initial state $x_0 = 5$, and the cost function

$$\sum_{k=0}^{3} (x_k^2 + u_k^2).$$

Apply the DP algorithm for the following three cases:

(a) The control constraint set $U_k(x_k)$ is $\{\mathrm{u} | 0 \leq x_k + \mathrm{u} \leq 5, \mathrm{u} : \mathrm{integer}\}$ for all x_k and k, and the disturbance w_k is equal to zero for all k.

(b) The control constraint and the disturbance w_k are as in part (a), but there is in addition a constraint x_4=5 on the final state. *Hint:* For this problem you need to define a state space for x_4 that consists of just the value x_4 =5, and also to redefine $U_3(x_3)$. Alternatively, you may use a terminal cost $g_4(x_4)$ equal to a very large number for $x_4 \neq 5$.

(c) The control constraint is as in part (a) and the disturbance w_k takes the value -1 and 1 with equal probability 1/2 for all x_k and u_k, except if $x_k + u_k$ is equal to 0 or 5, in which case $w_k = 0$ with probability 1.

2. Assume that we have a vessel whose maximum weight capacity is z and whose cargo is to consist of different quantities of N different items. Let v_i denote the value of the *i* th type of item, w_i the weight of *i* th type of item, and x_i the number of items of type *i* that are loaded in the vessel. The problem is to find the most valuable cargo, i.e., to maximize $\sum_{i=1}^{N} x_i v_i$ subject to the constraints $\sum_{i=1}^{N} x_i w_i \leq z$ and $x_i = 0,1,2, \ldots$ Formulate this problem in terms of DP.

3. A repairman must service n sites, which are located along a line and are sequentially numbered 1,2,…,n. The repairman starts at a given site s with 1<s<n, and is constrained to service only sites that are adjacent to the ones serviced so far, i.e., if he has already serviced sites i, i+1,…,j, then he may service next only site i-1 (assuming 1<i) or site j+1 (assuming j<n). There is a waiting cost c_i for each time period that site I has remained unserviced and there is a travel cost t_{ij} for servicing

site j immediately after servicing site i. Formulate a DP algorithm for finding a minimum cost service schedule.

4. An energetic salesman works every day of the week. He can work in only one of two towns A and B on each day. For each day he works in town A (or B) his expected reward is r_A (or r_B , respectively). The cost for changing towns is c. Assume that $c > r_A > r_B$ and that there is a discount factor. $\alpha < 1$

(a) Show that for α sufficiently small, the optimal policy is to stay in the town he starts in, and that for α sufficiently close to 1, the optimal policy is to move to town A (if not starting there) and stay in A for all subsequent times.

(b) Solve the problem for $c = 3$, $r_A = 2, r_B = 1$ and $\alpha = 0.9$ using policy iteration

(c) Use a computer to solve the problem of part (b) by value iteration.

5. A person has an umbrella that she takes from home to office and vice versa. There is a probability p of rain at the time she leaves home or office independently of earlier weather. If the umbrella is in the place where she is and it rains, she takes the umbrella to go to the other place (this involves no cost.) If there is no umbrella and it rains, there is a cost W for getting wet. If the umbrella is in the place where she is but it does not rain, she may take the umbrella to go to the other place (this involves an inconvenience cost V) or she may leave the umbrella behind (this involves no cost.) Costs are discounted at a factor $\alpha < 1$

(a) Formulate this as an infinite horizon total cost discounted problem. *Hint* : Try to use as few states as possible.

(b) Characterize the optimal policy as best as you can.

6. An unemployed worker receives a job offer at each time period, which she may accept or reject. The offered salary takes one of n possible values $\omega^1, \dots, \omega^n$ with given probabilities, independently of preceding offers. If she accepts the offer, she must keep the job for the rest of her life at the same salary level. If she rejects the offer, she receives unemployment compensation c for the current period and is eligible to accept future offers. Assume that income is discounted by a factor $\alpha < 1$

(a) Show that there is a threshold $\bar{\omega}$ such that it is optimal to accept an offer if and only if its salary is larger than $\bar{\omega}$, and characterize $\bar{\omega}$.

(b) Consider the variant of the problem where there is a given probability p_i that the worker will be fired from her job at any one period if her salary is ω^i. Show that the result of part (a) holds in the case where p_i is the same for all i. Analyze the case where p_i depends on i.