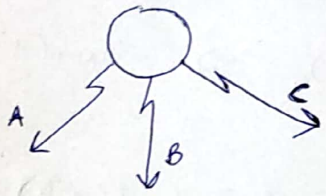
Quiz# 1 : E1 277: Reinforcement Learning

Name: Aman Singh
S.R.No.: 22027

Department: RBCCPS
Programme: PHD.

1. Consider a multi-armed bandit with three arms A, B, and C. The decision maker pulls these arms in the order A, B, C, B, A, C, B and gets the rewards $5, 4, 1, 6, 3, 7, 2$, respectively. Assuming initial Q-values for each arm as zero, find the final values of $Q(A)$, $Q(B)$ and $Q(C)$? (2 marks)



Actions \ time →	0	1	2	3	4	5	6	7
A	0	5	5	5	5	4	4	4
B	0	0	4	4	5	5	5	4
C	0	0	0	1	1	1	4	4.
Action done	A	B	C	B	A	C	B	
Reward	5	4	1	6	3	7	2	

$Q_0(A) = 0$, $Q_1(A) = \frac{5}{1} = 5$, $Q_2(A) = Q_3(A) = Q_4(A) = 5$, $Q_5(A) = \frac{5+3}{2} = 4$.

$Q_6(A) = 4$

$\boxed{Q_7(A) = 4.}$

$Q_0(B) = Q_1(B) = 0$, $Q_2(B) = \frac{4}{1} = 4$, $Q_3(B) = 4$, $Q_4(B) = \frac{4+6}{2} = 5$,

$Q_5(B) = Q_6(B) = 5.$,

$\boxed{Q_7(B) = \frac{5\times 2 + 2}{3} = \frac{12}{3} = 4.}$

$Q_0(C) = Q_1(C) = Q_2(C) = 0$, $Q_3(C) = \frac{1}{1} = 1.$, $Q_4(C) = Q_5(C) = 1.$

$Q_6(C) = \frac{1\times 1 + 7}{2} = 4$,

$\boxed{Q_7(C) = 4}$

(2)

2. Consider a finite horizon MDP for which the single stage costs at each of the times $0, 1, \ldots, N-1$ are the same, i.e., the functions $g_0 = g_1 = \cdots = g_{N-1} \stackrel{\triangle}{=} g$. Let the terminal cost be g_N and it only depends on the terminal state as before. Assume now that $J_{N-1}(x) \geq g_N(x)$, $\forall x \in S$ (where S denotes the state space). Show that this implies $J_k(x) \geq J_{k+1}(x)$, for all $k = 0, 1, \ldots, N-1$ and all $x \in S$. (3 marks)

$J_N(x) = g_N(x)$.

$J_{N-1}(x) = g(x_{N-1}) + J_N(x)$

$J_{N-1}(x) > g_N(x). \Rightarrow J_{N-1}(x) > J_N(x).$ —(1).

$J_k(x_k) = J_{k+1}(x_{k+1}) + g(x_k)$

$\Rightarrow J_{N-1}(x) = J_N(x) + g(x)$

$J_{N-2}(x) = J_{N-1}(x) + g(x)$.

J_K

@ $K = N-1$.

$J_{N-1}(x_{N-1}) = J_N(x_N) + g(x_N)$

$\Rightarrow J_{N-1}(x_{N-1}) - J_N(x_N) = g(x_N)$

$\Rightarrow g(x) \geq 0.$

$\Rightarrow J_k(x_k) = J_{k+1}(x_{k+1}) + g(x_k).$

$\Rightarrow J_k(x) \geq J_{k+1}(x_{k+1}) + J(x_N)]$

$J_{N-1}(x) = \min_{a \in A(x)} \sum_j p(x,a,j)[g(x,a) + J(j)]$

$J_{N-1}(x_{N-1}) = \min_{a \in A(x_{N-1})} \sum_{j \geq 0} p(x_{N-1}, a, j)[g(x_{N-1}, a) + J(j)]$

$(g + g_N(x_N))$

J_{N-1}