# AV499 - AVD871 - Programming Assgn. 4.

## Solution of finite horizon Markov decision processes.

In this assignment you have implement the recursive method discussed in class to solve finite horizon Markov decision processes. Suppose you are given a Markov decision process specified by the state space $\mathcal{S}$, action set $A$, transition probability matrices $p(a)$, reward functions $R_t, \forall t$ (assume that the expected rewards are given).

The finite horizon Markov decision process is solved by the following recursion equation, with $V_{-1}(s) = 0$.

$$V_t(s) = \max_a \left\{ \underset{T-t-1}{\mathbb{E}} R_t(s,a) + \mathbb{E}^a \cdot V_{t-1}(S) \right\}$$

here $\mathbb{E} R_t(s,a) (= r_t(s,a))$ is the expectation wrt to the intrinsic randomness of the reward; and

$\mathbb{E}^a V_{t-1}(S) = \sum_{s'} p_{ss'}^{(a)} V_{t-1}(s')$, and the recursion holds for $t \in \{0, 1, 2, \ldots, T-1\}$. You have to implement the above recursion and find out $V_t(s)$ as well as the policy $\pi_t(s)$.

Use your implementation to solve the MDP with the state space $\mathcal{S} = \{1, 2, 3\}$, action space $A = \{1, 2\}$, $p(a)$ given as

$$p^{(1)} = \begin{pmatrix} 0.5 & 0.2 & 0.3 \\ 0.6 & 0.2 & 0.2 \\ 0.1 & 0.8 & 0.1 \end{pmatrix} \qquad p^{(2)} = \begin{pmatrix} 0.3 & 0.3 & 0.4 \\ 0.9 & 0.05 & 0.05 \\ 0.7 & 0.2 & 0.1 \end{pmatrix}$$

The reward function $R_t$ is independent of time, with the expected reward $r_t(s,a) = r(s,a) = s + a^2$. Solve the MDP for the horizon $T = 10$.