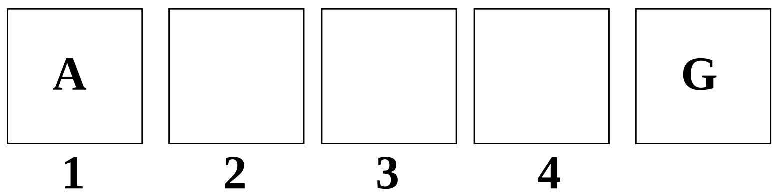
CS 6300 | # MDP Practice | **February 16, 2021**

A soccer robot A is on a fast break toward the goal, starting in position 1. From positions 1 through 3, it can either shoot (S) or dribble the ball forward (D). From 4 it can only shoot. If it shoots, it either scores a goal (state G) or misses (state M). If it dribbles, it either advances a square or loses the ball, ending up in M. When shooting, the robot is more likely to score a goal from states closer to the goal; when dribbling, the likelihood of missing is independent of the current state.



In this MDP, the states k are 1, 2, 3, 4, G and M, where G and M are terminal states. The transition model depends on the parameter y, which is the probability of dribbling success. Assume a discount of $\gamma = 1$.

$$
\begin{aligned}
T(k, S, G) &= \frac{k}{6} \\
T(k, S, M) &= 1 - \frac{k}{6} \\
T(k, D, k+1) &= y \text{ for } k \in \{1, 2, 3\} \\
T(k, D, M) &= 1 - y \text{ for } k \in \{1, 2, 3\} \\
R(k, S, G) &= 1
\end{aligned}
$$

Rewards are 0 for all other transitions.

1. Using $y = 3/4$, compute the first two iterations of value iteration. The equations for value iteration with $\gamma = 1$ are:

$$
\begin{aligned}
Q^*_{i+1}(s, a) &= \sum_{s'} T(s, a, s')[R(s, a, s') + V^*_i(s')] \\
V^*_{i+1}(s) &= \max_{a_i} Q^*_{i+1}(s, a)
\end{aligned}
$$

i	$Q_i(1,S)$	$Q_i(2,S)$	$Q_i(3,S)$	$Q_i(4,S)$
0	0	0	0	0
1				
2				

i	$Q_i(1,D)$	$Q_i(2,D)$	$Q_i(3,D)$
0	0	0	0
1			
2			

i	$V_i(1)$	$V_i(2)$	$V_i(3)$	$V_i(4)$
0	0	0	0	0
1				
2				

Below is the workspace for your answers. For iteration 1,

$Q_1(1,S) =$

$Q_1(2,S) =$

$Q_1(3,S) =$

$Q_1(4,S) =$

$Q_1(1,D) =$

$Q_1(2,D) =$

$Q_1(3,D) =$

$V_1(1) =$

$V_1(2) =$

$V_1(3) =$

$V_1(4) =$

For iteration 2,

$Q_2(1, S) =$

$Q_2(2, S) =$

$Q_2(3, S) =$

$Q_2(4, S) =$

$Q_2(1, D) =$

$Q_2(2, D) =$

$Q_2(3, D) =$

$V_2(1) =$

$V_2(2) =$

$V_2(3) =$

$V_2(4) =$

2. After two iterations, perform policy extraction.
3. Do two iterations of policy iteration for the initial policy $\pi_0^*(s) = S$.