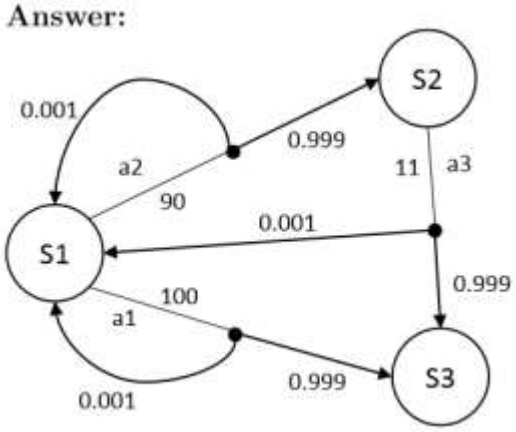
Assignment 2

Write c++ / python code that implement a simple Markov decision process (MDP) with the following properties:

a) Its immediate action costs are all positive
b) All of its actions can result with some probability in the start state
c) The optimal policy without discounting differs from the optimal policy with discounting and a discount factor of 0.9. Prove
d) Use value iteration algorithm

Answer:



With no discount factor, action a1 is preferred over action a2 in state s1:

	i	0	1	2	3	4	...
State1	a1	0	100	**100.09**	**100.10009**	**100.1001001**	
	a2	0	**90**	101.079	101.179	101.18909	
State2	a3	0	11	11.09	11.10009	11.10010009	
State3		0	0	0	0	0	

With discount factor = 0.9, action a2 is preferred over action a1 in state s1:

	i	0	1	2	3	4	...
State1	a1	0	100	100.081	100.089974	100.0900476	
	a2	0	**90**	**99.9711**	**100.0529011**	**100.0610432**	
State2	a3	0	11	11.081	11.08997399	11.09004761	
State3		0	0	0	0	0	

- You should work individually.
- Your code should produce the same result as shown in the answer section.
- Make sure you understand the problem first; solve it by hand initially before writing your code.
- You must write the code yourself.
- Identical solutions will be considered cheating, and both assignments will receive a grade of zero.