TPC5

1.

1. TODO Add steps and formula

[0.25 0.32055 0.32 0.25]

1. TODO Add steps and formula

[0.25 0.3272 0.32 0.25]

1. The difference between the on-policy (SARSA) and off-policy (Q-learning) is that the first one learns by improving the policy that is used to make decisions, i.e, learns the value of the policy that it follows. The second one learns the value of the policy independently of the agent’s action, i.e, learns the value of one policy while following another. Using the previous questions as exemple, on question a) we could see by the Q-learning formula that the agent chose the minimum Q-value for the next state while following the given policy, showing that it only follows the best estimated policy. For question b) we could see by the SARSA formula that the agent chose the Q-value corresponding to the policy that the agente is following.

Grupo nº 37

João Silvestre, nº 80996

João Pestana, nº 79515