Lista 2 MAC0425

Fernanda Itoda 10740825

Exercício 1

Sim, o resultado do Q-learning corresponde ao esperado intuitivamente. Foram necessárias 6 rodadas para esse resultado, no mínimo. Devido ao fator aleatório, em algumas execuções o resultado não é o esperado, mas essa parcela corresponde a minoria das execuções. Com 50 execições o trajeto sempre alcança o estado final de recompensa. O trajeto intuitivo considerado foi (2,0) - (2,1) - (2, 2) - (2,3) - (1,3) - (0, 3).

Exercício 2

Não, o resultqado do Q-learning não corresponde ao esperado intuitivamente. No programa, o trajeto tende a entrar em loop no estate com recompensa +10, visto que é mais benéfico para ele se permanecer nesse local e receber a recompensa infinitamente. O esperado seria que o trajeto passasse por esse estado e seguisse para cima, em direção ao estado final.

Uma forma de contornar esse problema é aumentando o valor de α, no código, self.lr na classe Agent, que corresponde a aversão ao loop.