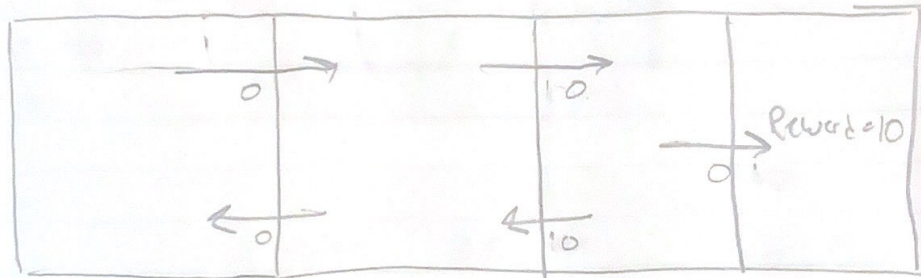
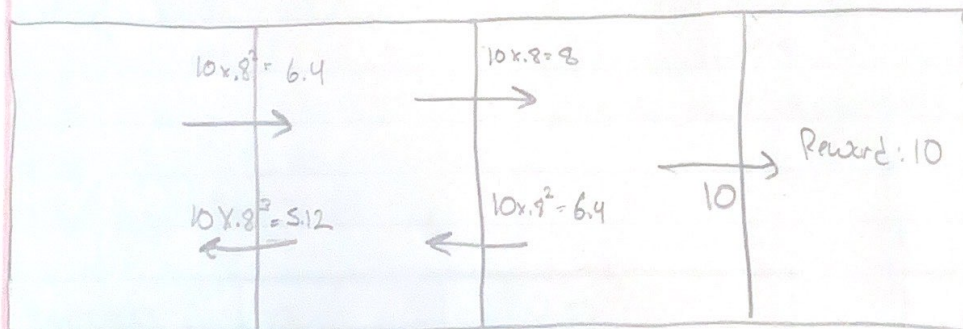


Q-learning HW

Initial model

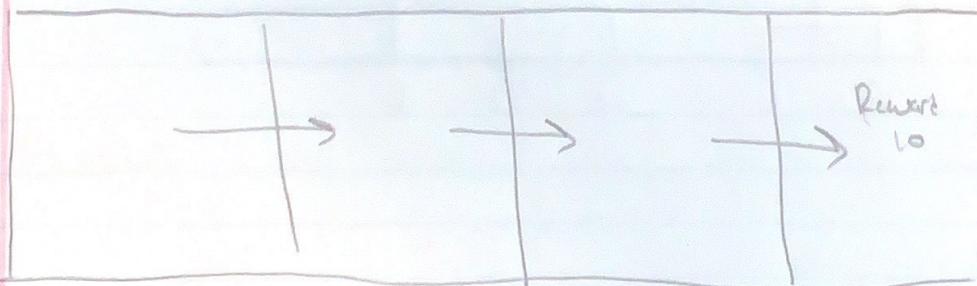


Discount = .8 , Initial Q's = 0



Model with final Q-values

Optimal policy



The optimal policy would be to go right from any state beside the final one since it will always increase the reward value.