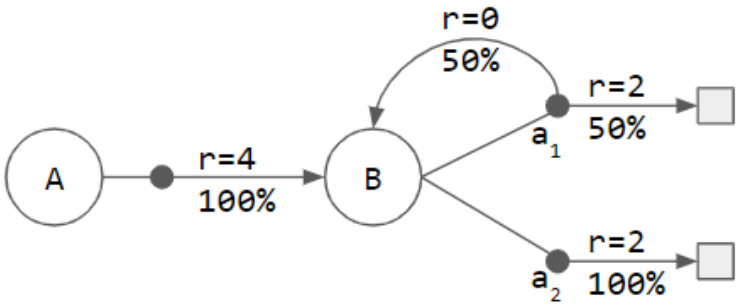
Deep Learning and Practice - Final (RL)

1. (17 pts) Consider the MDP shown below.



The non-terminal states are $S = \{A, B\}$, and the terminal state is the shaded square in the figure. There are two actions, $\{a_1, a_2\}$, at state B.

   (a) (7 pts) Given $\pi(B, a_1) = 25\%$, $\pi(B, a_2) = 75\%$, and $\gamma = 1$. What is $V_\pi(A)$?

   (b) (10 pts) Given $\gamma = 0.5$,

      i. (7 pts) What is the optimal value $V^*(A)$? (Hint: Bellman optimality equation)

      ii. (3 pts) Give an example of optimal policy and justify.

2. (12 pts) Answer the following questions related to DQN and DDPG.

   (a) (6 pts) What techniques are used for exploration in DQN and DDPG respectively?

   (b) (6 pts) Explain the necessity and the importance of the target network in DQN.