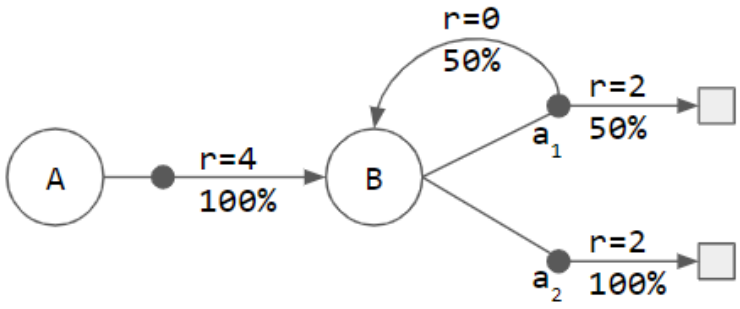
Deep Learning and Practice - Final (RL)

1. (18 pts) Consider a MDP shown below.



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

   (a) (8 pts) Given $\pi(B, a_1) = 25\%$, $\pi(B, a_2) = 75\%$.
      i. (4 pts) What is $V_\pi(A)$ when $\gamma = 1$?
      ii. (4 pts) What is $V_\pi(A)$ when $\gamma = 0.5$?

   (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 importance of the target network in DQN.