Note: The Section 3 questions should be done with pencil and paper. If you'd like to take a picture of your work and convert it to a pdf before emailing it, we recommend CamScanner

1 Discount Factor

1. If we're looking for a policy with maximum discounted cumulative reward, what does a high discount factor encourage?
2. If we're looking for a policy with maximum discounted cumulative reward, what does a low discount factor encourage?
3. What is the range of values a discount factor can be?

2 MDPs in Perspective

Recall: An MDP with deterministic transitions means that taking a certain action from a certain state always leads to the same unique state, but an MDP with stochastic transitions means that taking that action from that state can result in multiple possible next states. On the other hand, a deterministic policy is a 1-1 mapping from states to actions (defining the agent's behavior), and a stochastic policy is a mapping from states to a probability distribution of actions.

1. If we have a markov decision process with deterministic transitions and we pick a specific deterministic policy to follow, what kind of system do we end up with?
2. If we have a markov decision process with deterministic transitions and we pick a specific stochastic/nondeterministic policy to follow, what kind of system do we end up with?
3. If we have a markov decision process with stochastic/nondeterministic transitions and we pick a specific deterministic policy to follow, what kind of system do we end up with?

3 An MDP Example

Consider an MDP defined as follows:

- States: S = a, b, c, e, f
- Actions: A = Left, Right
- Transition Probabilities:

  T(a, Left, b) = 0.5, T(a, Left, c) = 0.5, T(a, Right, b) = 0.25, T(a, Right, c) = 0.75
  T(b, Left, e) = 1.0, T(b, Right, e) = 0.5, T(b, Right, f) = 0.5
  T(c, Left, f) = 1.0

  Only the transitions with nonzero probabilities are given. If a state has no listed actions, it means it is a terminal state.

- Reward function:

  R(a, Left, b) = 0, R(a, Left, c) = 0, R(a, Right, b) = 0, R(a, Right, c) = 0
  R(b, Left, e) = +3, R(b, Right, e) = +6, R(b, Right, f) = +2
  R(c, Left, f) = +8

- Discount Factor:

  1.0

1. Draw the MDP given by the above definitions
2. Calculate the value (discounted expected cumulative reward) of each state under a policy that gives Left for all states