CS449/CS549 Computational Learning Optional Quiz 5 Fall 2023 (take home).
Out: 12/06/23. *You may consult class materials and other additional references, but your submitted answers must be your own work.*

Partially Observable Markov Decision Process (POMDP).
Let $S = \{s_0, s_1\}$ be a set of states, $A = \{a, b\}$ be a set of actions, $\{P_a, P_b\}$ be a set of Markov chains, $O = \{0, 1\}$ be a set of outputs, $\{Q_{s_0}, Q_{s_1}\}$ be a set of stochastic output maps corresponding to each state where

$$P_a = \begin{bmatrix} 3/4 & 1/3 \\ 1/4 & 2/3 \end{bmatrix}, \qquad P_b = \begin{bmatrix} 1/5 & 1/2 \\ 4/5 & 1/2 \end{bmatrix}, \qquad Q_{s_0} = \begin{bmatrix} 0.1 \\ 0.9 \end{bmatrix} \qquad Q_{s_1} = \begin{bmatrix} 0.8 \\ 0.2 \end{bmatrix}$$

Let $R = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ be the state-based reward function and $\gamma = \frac{1}{2}$. Our POMDP is

$$M = (S, A, \{P_a, P_b\}, O, \{Q_{s_0}, Q_{s_1}\}, R, \gamma).$$

We assume M is equally likely to start in s_0 or s_1. Assume $(Q_s)_i$ is the probability of output i from state s, $i = 0, 1$. A policy is a map $\pi : S \to A$.

1. State the initial belief state vector β_0 of a Bayesian agent playing against M.
2. What is the Bayes-updated belief vector β_1 after observing output of 1?
3. What is the *expected* (average) reward at this point?
4. Suppose the agent takes action a based on its belief β_1 (is this rational?). What is the revised belief after this action (according to the appropriate Markov chain)?
5. If the second output observed is 0, what is the belief now?