1. Consider the following Markov decision problem. Suppose a discount factor $\gamma = 0.5$.

\[
\begin{align*}
  s (P) & : s (r = 6, p = 0.6) ; a (r = 7, p = 0.4), \\
  a (P) & : s (r = 2, p = 0.1) ; b (r = -3, p = 0.9), \\
  b (D) & : b (r = 3) ; c (r = 0), \\
  c (D) & : a (r = 4) ; b (r = -5) ; c (r = -1).
\end{align*}
\]

(1)

Here, “$P$” indicates a probabilistic node, and “$D$” a decision node.

(a) Give a state transition diagram, including probabilities and rewards.

(b) Give the probability transition matrix, $P$, and the immediate reward matrix, $R$.

(c) Express the vector of optimal values $v^* = (s, a, \ldots, c)$ as a solution of a system of non-linear equations. It is not necessary to simplify this system of non-linear equations.

(d) Express the vector of optimal values $v^* = (s, a, \ldots, c)$ as a solution of a linear program. (It is not necessary to simplify the linear program.)

(e) Perform policy iteration with a convergence tolerance of $\epsilon = 0.01$. Take $\pi_0$ to be the policy that puts probability 1 on the first successor of every decision node. (The order of the successor nodes can be read off from Table 1.)

(f) Perform value iteration with a convergence tolerance of $\epsilon = 0.01$. If values lead to a new (better) policy, write down the new policy.