Homework 3
Due November 1, Tuesday

1. Consider a model with $S = \{s_1, s_2\}$, $A_{s_1} = \{a_{11}, a_{12}\}$, $A_{s_2} = \{a_{21}, a_{22}, a_{23}\}$, $p\{s_1|s_1, a_{11}\} = 1$, $p\{s_1|s_1, a_{12}\} = 0.5$, $p\{s_1|s_2, a_{21}\} = 1$, $p\{s_1|s_2, a_{22}\} = 0$ and $p\{s_1|s_2, a_{23}\} = 0.75$.
   a. Determine the chain structure of each deterministic stationary policy.
   b. Compute the long average optimal gain. What is the long run average optimal policy?

2. Show that if all stationary deterministic policies are unichain, then all stationary randomized policies are unichain.

3. A decision maker observes a discrete time system which moves between states $\{s_1, s_2, s_3, s_4\}$ according to the following transition probability matrix:

\[
P = \begin{bmatrix}
0.3 & 0.4 & 0.2 & 0.1 \\
0.2 & 0.3 & 0.5 & 0.0 \\
0.1 & 0.0 & 0.8 & 0.1 \\
0.4 & 0.0 & 0.0 & 0.6 \\
\end{bmatrix}
\]

At each point in time the decision may leave the system and receive a reward of $R = 20$ units or alternatively remain in the system and receive a reward of $r(s_i)$ units if the system occupies state $s_i$. If the decision maker decides to remain in the system, its state at the next decision epoch is determined by $P$. Assume a discount rate of $\lambda = 0.9$ and $r(s_i) = i$. Formulate this problem as a Markov decision process and use the policy iteration algorithm to find a stationary policy which maximizes the expected total discounted reward.