\[ Y = 0.9 \]

\[ p(1,2 | 1,1, U) = 0.8 \]
\[ p(2,1 | 1,1, U) = 0.1 \]
\[ p(1,1 | 1,1, U) = 0.1 \]

\[ u \]
\[ u' \]
\[ 0 \]
\[ 0 \]
\[ 0.789 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
\[ 0 \]
Policy-Iteration $(S, A, P, R, V)$

For all $s \in S$  \( \Pi(s) = \text{random legal action} \), $U(s) = 0$

Repeat

\[ U \leftarrow \text{Policy Evaluation} \left( S, A, P, R, V, \Pi, U \right) \]

For each $s \in S$

\[ \text{if } \max_{a \in A} \sum_{s' \in S} \pi(s'|s,a) U(s') > \sum_{s' \in S} \pi(s'|s,a) U(s') \]

then \( \Pi(s) \leftarrow \text{argmax} \sum_{s' \in S} \pi(s'|s,a) U(s') \)

Until $\Pi$ doesn't change

Partially Observable MDPs (POMDPs)

Still enumerate states

Don't observe states, get information on possible states you may be in

Maintain belief state - prob. distribution over states

Treat these as states, solve MDP over these states