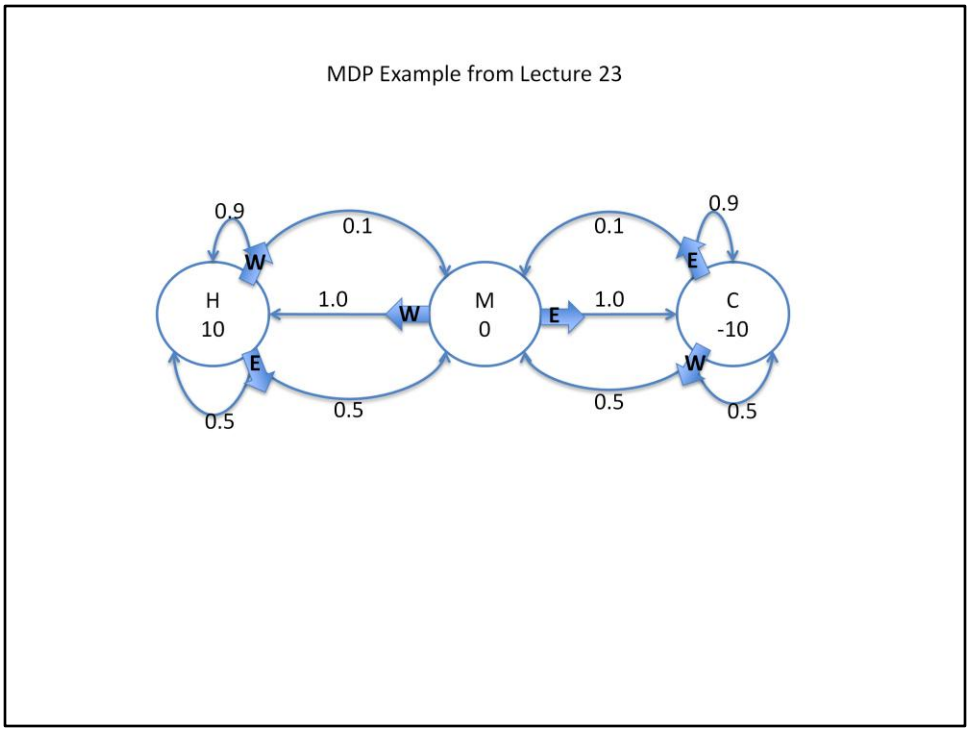

States:
Hot (reward 10)
Mild (reward 0)
Cold (reward -10)

Actions: East, West

Transitions:
From Hot go West (.9 -> Hot, .1 -> Mild)
From Hot go East (.5 -> Hot, .5 -> Mild)
From Mild go West (1.0 -> Hot)
From Mild go East (1.0 -> Cold)
From Cold go West (.5 -> Mild, .5 -> Cold)
From Cold go East (.1 -> Mild, .9 -> Cold)

VALUE ITERATION

$V^0(H) = 10$
$V^0(M) = 0$
$V^0(C) = -10$
Gamma = 0.5

$V^1(H) = 10+.5 *\max(W, E) = 10 + .5 * \max(.9*10 + .1*0, .5*10 + .5*0) = 10 + .5 * \max(9,5) = 14.5$

POLICY ITERATION
$\Pi^0(H) = E$, $\Pi^0(M) = E$, $\Pi^0(C) = W$

Check these numbers (They are right)
V(H) = 10 + 0.5(0.5*V(H) + 0.5*V(M))
V(M) = 0 + 0.5(1.0*V(C))
V(C) = -10 + 0.5(0.5*V(M) + 0.5*V(C))

V(H) = 10.67
V(M) = -8
V(C) = -16

$\Pi^1(H) = \text{argmax a } \{W, E\} = \{10+.5(.9*10.67 + .1*-8), 10+.5(.5*10.67 + .5*8)\} \Rightarrow \Pi^1(H) = W$