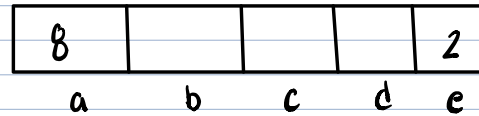
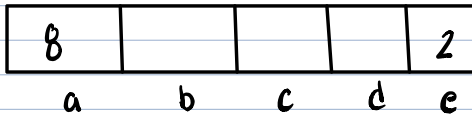


1. Discounting



Actions: W, E, X - "a" or "e"

→ X W W W W

Transitions: deterministic

Rewards: -0.01 W or E
8 for X at "a"
2 for X at "e"

$$-0.01 + (1 \cdot 8) + -0.01(0.95^1 \cdot 8) + -0.01(0.95^2 \cdot 8) + -0.01(0.95^3 \cdot 0) + -0.01(0.95^4 \cdot 2)$$

$$7.99 + 0.918 + 0.84256 + 0.84740 + 1.6190 \approx 12.2889$$

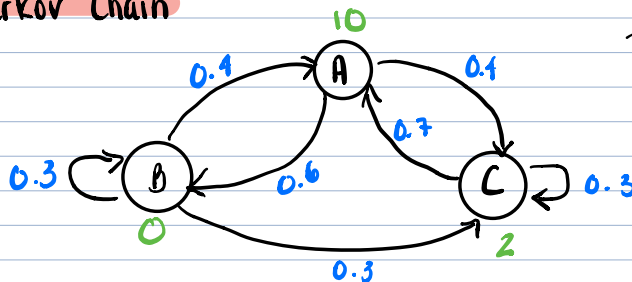
$\gamma = 0.95$, optimal policy?

→ E E E E X

$$-0.01(1 \cdot 2) + -0.01(0.95^1 \cdot 0) + -0.01(0.95^2 \cdot 0) + -0.01(0.95^3 \cdot 8) + -0.01(0.95^4 \cdot 8)$$

$$1.99 + 0.940 + 0.89256 + 0.84740 + 6.50605 \approx 11.17595$$

2. Markov Chain



$\gamma = 0.9$

$$V(A) = 10 + 0.9 [(0.6V(B)) + (0.1V(C))]$$

$$x = 10 + 0.9 (0.6y + 0.1z)$$

$$10 + 0.54y + 0.36z - x = 0$$

$$V(A) = x$$

$$V(B) = y$$

$$V(C) = z$$

$$V(B) = 0 + 0.9 [(0.4V(A)) + (0.3V(B)) + (0.3V(C))]$$

$$y = 0.36x + 0.27y + 0.27z$$

$$0.36x + 0.27y + 0.27z - y = 0$$

$$0.36x - 0.73y + 0.27z = 0$$

$$V(C) = 2 + 0.9 [(0.7V(A)) + (0.3V(C))]$$

$$2 + 0.63x + 0.27z - z = 0$$

$$2 + 0.63x - 0.73z = 0$$

plugged into Symbolab's systems of equations calculator.

$$V(A) = 16.01182$$

$$V(B) = 38.39335$$

$$V(C) = 42.45115$$

3. Value Iteration

V^0	Cool 0	Warm 0	Overheated 0	$\gamma = 0.8$
V^1	1.0	1.0	0	
V^2	7.2	1.80	0	

s	a	s'	$T(s, a, s')$	$R(s, a, s')$
cool	slow	cool	1.0	+1
cool	fast	cool	0.3	+1
cool	fast	warm	0.7	+4
warm	slow	cool	0.4	+1
warm	slow	warm	0.6	+1
warm	fast	over	1.0	-10

Math:

$$\max_{s'} \sum_{s'} T(s, a, s') [R(s, a, s') + \gamma V^e(s')]$$

$$\max \begin{cases} a = \text{slow} & s' = \text{cool} & 1[1 + 0.8(0)] \rightarrow 1[1] = 1 \\ a = \text{fast} & s' = \text{cool} & 0.3[1 + 0.8(0)] \rightarrow 0.3[1] = 1.2 + = \textcircled{4.0} \\ & s' = \text{warm} & 0.7[4 + 0.8(0)] \rightarrow 0.7[4] = 2.8 \end{cases}$$

$$\max \begin{cases} a = \text{slow} & s' = \text{cool} & 0.4[1 + 0.8(0)] \rightarrow 0.4[1] = 0.4 = 1 \uparrow \\ & s' = \text{warm} & 0.6[1 + 0.8(0)] \rightarrow 0.6[1] = 0.6 \\ a = \text{fast} & s' = \text{over} & 1[1 + -10(0)] \rightarrow 1[-9] = -9 \end{cases}$$

$$\max \begin{cases} a = \text{slow} & s' = \text{cool} & 1[1 + 0.8(4.0)] \rightarrow 1[4.2] = 4.2 \\ a = \text{fast} & s' = \text{cool} & 0.3[4 + 0.8(4.0)] \rightarrow 0.3[7.2] = 2.16 + = \textcircled{6.2} \\ & s' = \text{warm} & 0.7[4 + 0.8(4.0)] \rightarrow 0.7[7.2] = 5.04 \end{cases}$$

$$\max \begin{cases} a = \text{slow} & s' = \text{cool} & 0.4[1 + 0.8(1)] \rightarrow 0.4[1.8] = 0.72 + = \textcircled{1.80} \\ & s' = \text{warm} & 0.6[1 + 0.8(1)] \rightarrow 0.6[1.8] = 1.08 \\ a = \text{fast} & s' = \text{over} & 1[1 + -10(1)] \rightarrow 1[-9] = -9 \end{cases}$$