

MDP with an Example

There is an agent that is training to regulate the temperature of a room. The room can either be cold or hot. The agent (thermostat) can either decide to turn on the cooler or the heater.

- * Given that the room is cold, by turning on the cooler there is a 90% chance of room remaining cold. However, if heater is turned on, there is 80% chance that the room gets hot.
- * Given that the room is hot, by turning on the cooler there is a 80% chance of room becoming cold. However, if heater is turned on, there is 70% chance that the room gets hot.

$$S = \{\text{cold, hot}\}$$

$$A = \{\text{cooler, heater}\}$$

Draw transition matrices:

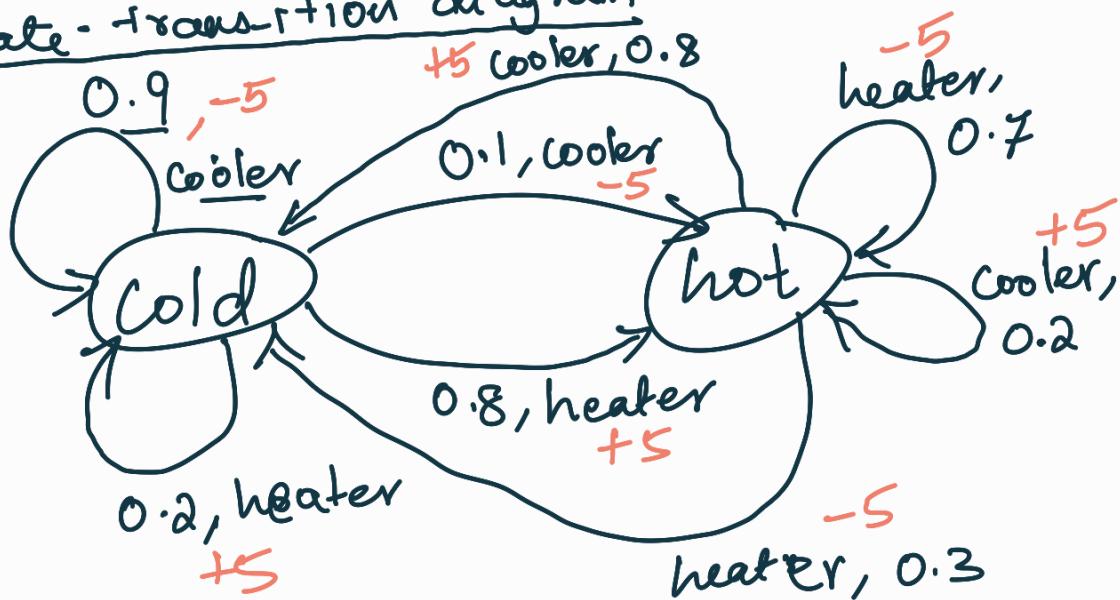
$A = \text{cooler}$

	cold	hot
cold	0.9	0.1
hot	0.8	0.2

$A = \text{heater}$

	cold	hot
cold	0.2	0.8
hot	0.3	0.7

State-transition diagram



Rewards cold \rightarrow turn on the cooler $\Rightarrow -5$

One-step rewards hot \rightarrow turn on the heater $\Rightarrow -5$

cold \rightarrow heater $\Rightarrow +5$
hot \rightarrow cooler $\Rightarrow +5$

* $P(\underbrace{\text{hot}}_{\text{hot}} | \text{hot, cooler}) = 0.2$

$$P(S_{t+1} = \text{hot} | S_t = \text{hot}, A_t = \text{cooler}) \\ = T(\text{hot, cooler, hot})$$

- $P(\text{hot} | \text{hot, heater}) = \frac{2}{3}$
 $P(\text{cold} | \text{hot, heater}) = ?$
 $P(\text{cold} | \text{cold, } \cancel{\text{heater}}) = ?$
 $P(\text{cold} | \text{hot, cooler}) = ?$
 $P(\text{cold} | \text{cold, cooler}) = ?$
 $P(\text{hot} | \text{cold, cooler}) = ?$
 $P(\text{hot} | \text{cold, heater}) = ?$
-

One step rewards :

* $\gamma(\underline{s_t}, \underline{a_t}, \underline{s_{t+1}})$
 $\gamma(\underline{\text{hot}}, \underline{\text{cooler}}, \underline{\text{hot}}) = +5$

⋮

Discounted return

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots$$

$$G_t = R_{t+1} + \gamma G_{t+1}$$

$$\gamma \in [0, 1]$$

$$= 0 \leq \gamma \leq 1$$

$$G_{t+2} = 10$$

$$R_{t+1} = 5$$

$$\gamma = 0.9$$

$$G_t = 5 + 0.9(10)$$

$$\text{chess} \Rightarrow \gamma = ?$$

$$\gamma = \underline{\underline{1}}$$

$$S = \{ \text{Sunny, Rainy} \}$$

$$A = \{ \text{Umbrella, no umbrella} \}$$

v
N.U

Given that ~~at~~ today is rainy there is 10% chance that tomorrow is sunny.

Given that today is sunny there is 20% chance that tomorrow is rainy.

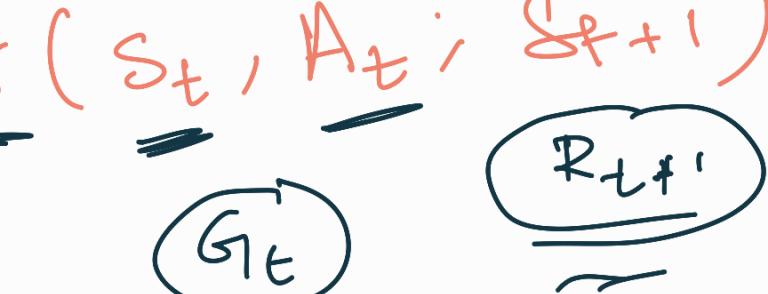
If you carry an umbrella
on a rainy day $\Rightarrow +5$ reward

Sunny day $\Rightarrow -5$ reward

No umbrella, rainy day $\Rightarrow -10$

No umbrella, sunny day $\Rightarrow +10$

- * Draw state transition diagram: probability.
- * Write the transition matrix

$$P(S_{t+1} | S_t, A_t)$$
$$\gamma(S_t, A_t; S_{t+1})$$


Exercise 3.8 Suppose $\gamma = 0.5$ and the following sequence of rewards is received $R_1 = -1$, $R_2 = 2$, $R_3 = 6$, $R_4 = 3$, and $R_5 = 2$, with $T = 5$. What are G_0, G_1, \dots, G_5 ? Hint: Work backwards. \square

$$G_5 = R_{5+1} = R_6 = 0$$

$$G_4 = R_5 + \gamma G_5$$
$$= 2$$

$$G_3 = R_4 + \gamma^2 G_4$$

$$G_2$$

$$G_1$$

$$G_0$$