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.

$S = \{cold, hot\}$

$A = \{cooler, heater\}$

Draw transition matrices:

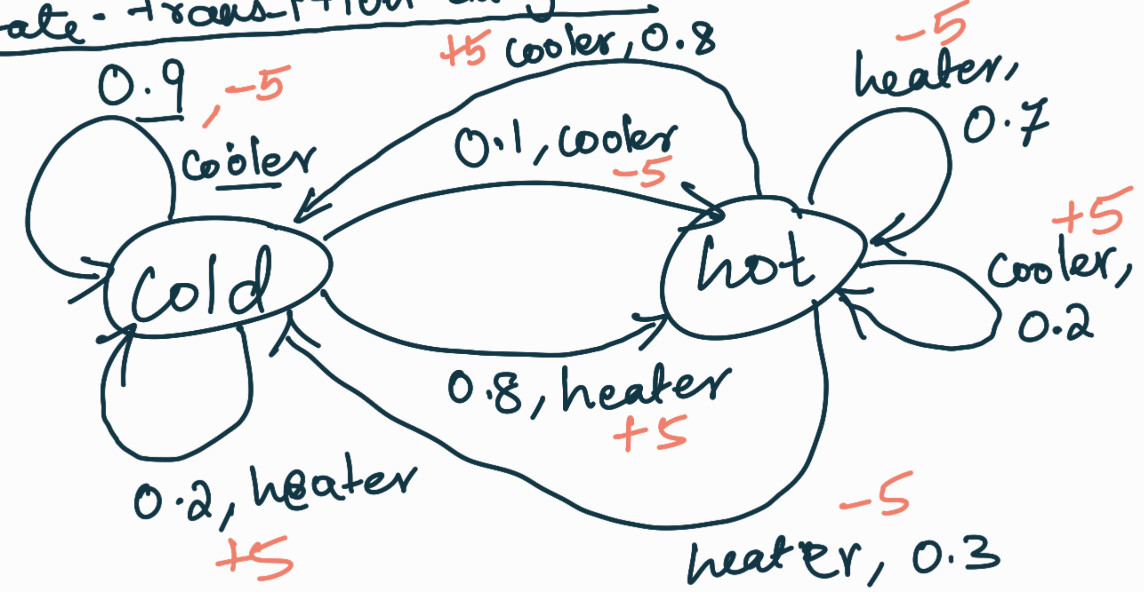
A = cooler

	cold	hot
cold	0.9	0.1
hot	0.8	0.2

A = heater

	cold	hot
cold	0.2	0.8
hot	0.3	0.7

State-transition diagram



Rewards cold → turn on the cooler ⇒ -5

One-step rewards

hot → turn on the heater ⇒ -5

cold → heater ⇒ +5

hot → coole ⇒ +5

\# $P(\text{hot} \mid \text{hot}, \text{cooler}) = 0.2$

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

$$= T(\text{hot}, \text{cooler}, \text{hot})$$

$p(\text{hot} \mid \text{hot}, \text{heater}) = ?$

$p(\cancel{\text{ho}}\ \text{cold} \mid \text{hot}, \text{heater}) = ?$

$p(\text{cold} \mid \text{cold}, \text{heater}) = ?$

$p(\text{cold} \mid \text{hot}, \text{cooler}) = ?$

$p(\text{cold} \mid \text{cold}, \text{cooler}) = ?$

$p(\text{hot} \mid \text{cold}, \text{cooler}) = ?$

$p(\text{hot} \mid \text{cold}, \text{heater}) = ?$

One step rewards :

* $r(\underline{S_t}, \underline{A_t}, \underline{S_{t+1}})$

$r(\underline{\text{hot}}, \underline{\text{cooler}}, \underline{\text{hot}}) = +5$

⋮

Discounted return

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

$$\gamma \left[R_{t+2} + \gamma R_{t+3} + \ldots \right]$$

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

$\gamma \in [0, 1]$

$= 0 \leq \gamma \leq 1$

chess $\Rightarrow \gamma = ?$

$\gamma = 1$

$G_{t+2} = 10$

$R_{t+1} = 5$

$\gamma = 0.9$

$G_t = 5 + 0.9(10)$

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

$A = \{ \underset{U}{\text{umbrella}}, \underset{N.U}{\text{no umbrella}} \}$

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

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

If you carry an umbrella on a rainy day => +5 reward

sunny day => −5 reward

No umbrella, rainy day => −10

No umbrella, sunny day => +10

* Draw state transition diagram.
* write the probability transition matrix
* $P(S_{t+1} \mid S_t, A_t)$
* $r(S_t, A_t; S_{t+1})$

R_{t+1}

G_t

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 $G_0, G_1, \ldots, G_5$? Hint: Work backwards. □

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

$G_4 = R_5 + \gamma G_6$

$= 2$

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

$\vdots$

G_2

$\vdots$

G_1

$\vdots$

G_0