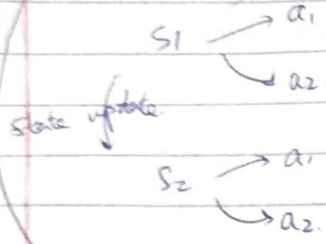


# Decision Process

## Q-Learning



## Q-table

| r  | a1 | a2 |
|----|----|----|
| s1 | -2 | 1  |

$$Q(s1, a1) < Q(s1, a2)$$

~~choose~~ choose  $a_2$

## Q-table

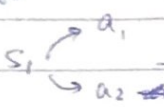
| r  | a1 | a2 |
|----|----|----|
| s2 | -4 | 2  |

$$Q(s2, a1) < Q(s2, a2)$$

~~choose~~ choose  $a_2$

How Q-table update and improve

→ estimated Q-table



In state  $s_1$ , choose  $a_2$

imaginary

$S_2 \rightarrow$  update Q-table (before making next move)

no real action  
imagine  $\rightarrow a_1$   
 $\rightarrow a_2$

Q-table

| r  | a1 | a2 |
|----|----|----|
| s1 | -2 | 1  |

according to previous Q-table,  $a_2$  better than  $a_1$

| r  | a1 | a2 |
|----|----|----|
| s1 | -2 | 1  |
| s2 | -4 | 2  |

$$R + \gamma \max_{a'} Q(s2, a')$$

$s_2$  最大的 Q 值

衰减系数 attenuation coefficient

reward of reaching  $s_2$

attenuated rate of future reward

real  $Q(s1, a2)$

recall estimated  $Q(s1, a2)$

$$d = |Q_{real}(s1, a2) - Q_{estimate}(s1, a2)|$$

$$new Q(s1, a2) = \underset{estimated}{Q(s1, a2)} + \alpha * d$$

learning efficiency: decide how much of the difference will be studied this time.

$$Q(s, a) \leftarrow Q(s, a) + \alpha [r + \gamma \max_{a'} Q(s', a') - Q(s, a)]$$

attenuated maximum estimated  $Q(s2)$

Basically, we are using the maximum estimated value of next step ~~to the reality~~ plus the reward as the reality value of this step

$$Q(s_1) = r_2 + \gamma Q(s_2) = r_2 + \gamma [r_3 + \gamma Q(s_3)] = r_2 + \gamma [r_3 + \gamma [r_4 + \gamma Q(s_4)]]$$

$$\Rightarrow Q(s_1) = r_2 + \gamma r_3 + \gamma^2 r_4 + \gamma^3 r_5 + \gamma^4 r_6 + \dots$$

$\therefore$  If  $\gamma = 1$ ,  $Q(s) = r_2 + r_3 + r_4 + \dots$

which means seeing the future clearly without any attenuation, i.e. it can perfectly predict the future.

If  $\gamma = 0$ ,  $Q(s) = r_2$ .

which means it cannot see any future step but only ~~it~~ cares about the most recent reward

If  $0 < \gamma < 1$

means it see ~~recent~~ near future steps more clearly than remote future steps.

~~As the training moves forward~~

As the training moves forward, more real future steps are used to update the Q-table, ~~the machine can better see~~  
the machine can better predict the ~~near future~~ <sup>see beyond short-term ~~future~~ benefit</sup>