Epsilon greedy policy

Advantage of epsilon greedy policy ($\epsilon = 0.4$)

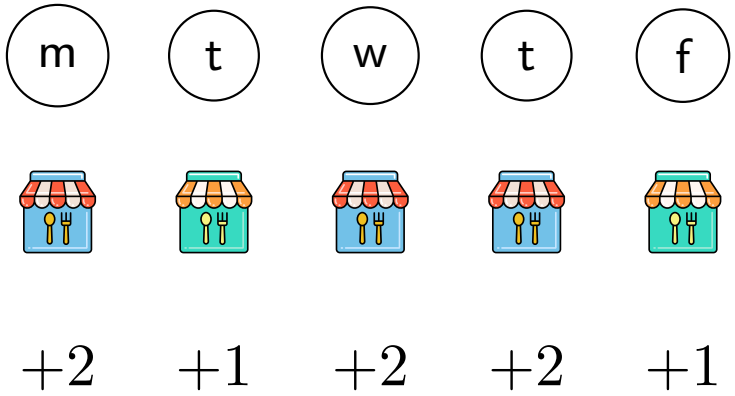

Figure: Week 1

- In the random phase, we haven't seen the following state-action pair with Q-value 3.
  - $Q($ m ,  $) = ?$

Advantage of epsilon greedy policy ($\epsilon = 0.4$)

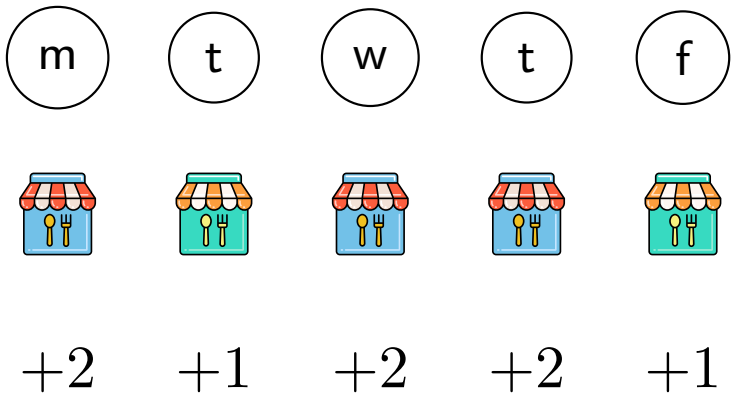

Figure: Week 1

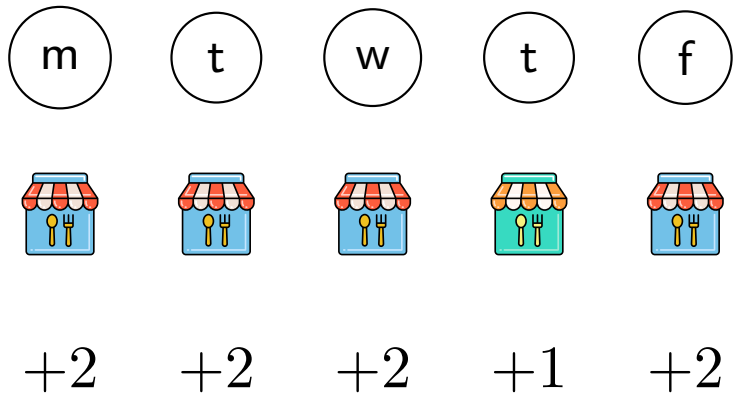

Figure: Week 2

- In the random phase, we haven't seen the following state-action pair with Q-value 3.
  - $Q($ m , $) = ?$

Advantage of epsilon greedy policy ($\epsilon = 0.4$)

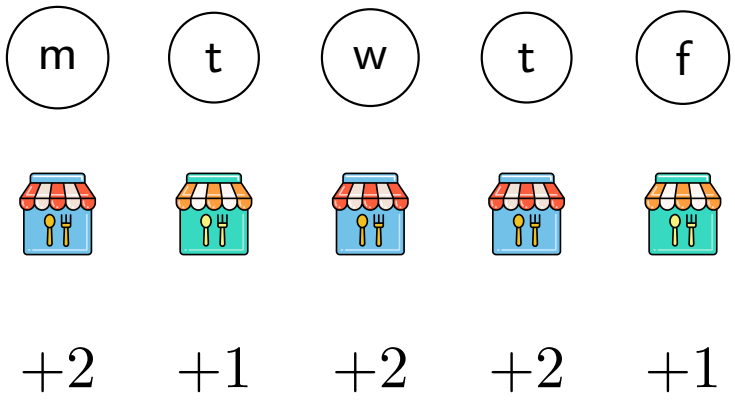

Figure: Week 1

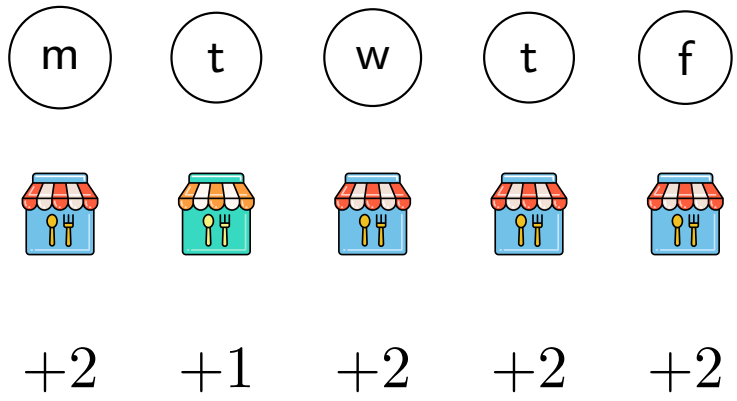

Figure: Week 3

- In the random phase, we haven't seen the following state-action pair with Q-value 3.
  - $Q($ m , $) = ?$

Advantage of epsilon greedy policy ($\epsilon = 0.4$)

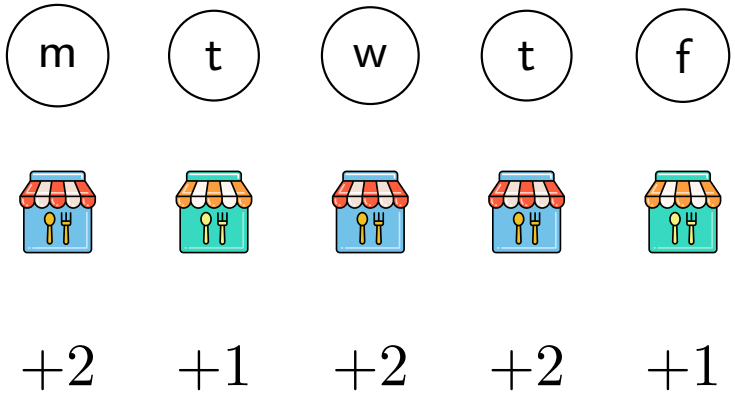

Figure: Week 1

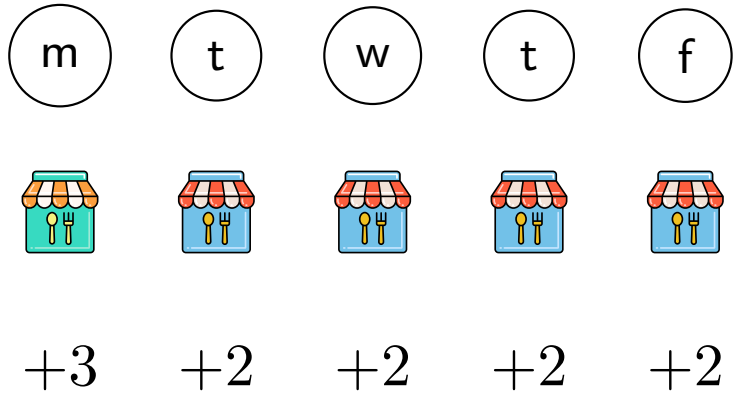

Figure: Week 4

- In the random phase, we haven't seen the following state-action pair with Q-value 3.
  - $Q($ m , $) = ?$
- $\epsilon - \text{greedy}$ policy finds the never-before-seen state-action pair with Q-value 3!
  - $Q($ m , $) = 3$

"Your assumptions are your windows on the world. Scrub them off every once in a while, or the light won't come in." (Isaac Asimov)

Does $\epsilon - \text{greedy}$ policy have the policy improvement aspect of greedy policies?

- Is it a better policy?

$$\epsilon - \text{greedy}(\pi) \stackrel{?}{\geq} \pi$$

Does $\epsilon - \text{greedy}$ policy have the policy improvement aspect of greedy policies?

- Is it a better policy?

$$\epsilon - \text{greedy}(\pi) \overset{?}{\geq} \pi$$

- Under what condition will iterative $\epsilon - \text{greedy}$ policy improvement lead to the optimal policy?

$$\pi_1 \leq \epsilon - \text{greedy}(\pi_1) = \pi_2 \leq \epsilon - \text{greedy}(\pi_2) = \pi_3 \leq \cdots \pi_*$$