

## hw5\_rl\_q7\_exploration\_and\_exploitation

### Question 7: Exploration and Exploitation

0.0/10.0 points (graded)

For each of the following action-selection methods, indicate which option describes it best. A: With probability  $p$ , select  $\operatorname{argmax}_a Q(s, a)$ . With probability  $1 - p$ , select a random action.  $p = 0.99$

☐ Mostly exploration

☒ Mostly exploitation ✓

☐ Mix of both

99% of the time, it will choose the greedy action with respect to the current Q values, which is exploitation, so this is mostly exploitation.

B: Select action  $a$  with probability

$$P(a|s) = \frac{e^{Q(s,a)/\tau}}{\sum_{a'} e^{Q(s,a')/\tau}}$$

where  $\tau$  is a temperature parameter that is decreased over time.

☐ Mostly exploration

☐ Mostly exploitation

☒ Mix of both ✓

When  $\tau$  is high this method results in mostly exploration, and as  $\tau$  is decreased, it results in more exploitation. Consider when  $\tau$  is  $\infty$ , in this case all actions are selected uniformly at random. Then, as  $\tau$  decreases, actions with higher Q values start to get selected with higher probability resulting in more exploitation.

C: Always select a random action.

☒ Mostly exploration ✓

☐ Mostly exploitation

☐ Mix of both

By not considering the policy at all, it is doing no exploitation, and is thus mostly exploration.

D: Keep track of a count,  $K_{s,a}$ , for each state-action tuple, (s,a), of the number of times that tuple has been seen and select  $\operatorname{argmax}_a [Q(s,a) - K_{s,a}]$ .

☐ Mostly exploration

☐ Mostly exploitation

☒ Mix of both ✓

This method initially does mostly exploitation, but as the  $K_{s,a}$  counts increase for commonly seen tuples, it begins trying new tuples with lower Q values, which is more exploration.

Which method(s) would be advisable to use when doing Q-Learning?

☐ A

☒ B ✓

☐ C

☒ D ✓

In general, it is best to use methods that mix exploration and exploitation when doing Q-learning.

Submit

---

**i** Answers are displayed within the problem