



[Course](#) > [Home...](#) > [Home...](#) > [hw5_rl...](#)

hw5_rl_q7_exploration_and_exploitation

Question 7: Exploration and Exploitation

10/10 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



B: Select action a with probability

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

where τ is a temperature parameter that is decreased over time.

☐ Mostly exploration

☐ Mostly exploitation

☒ Mix of both



C: Always select a random action.

☒ Mostly exploration☐ Mostly exploitation☐ Mix of both

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

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

☐ A☒ B☐ C☒ D

✓ Correct (10/10 points)