**Naïve Strategies**:

Explore only:

Spend entire time exploring, visit a random point each time, equal split over all points

Exploit only:

Visit each point once and then the best one you keep going to for the rest of the time

E-Greedy:

Set E value, something like 10%. So, there is a 90% chance that we’re going to the point that has historically given us the best score so far and a 10% chance to visit a random point.

**Upper Confidence Bound (UCB)**:

Works off E-Greedy method.

The strategy is that:

At each time T, pick point R, such that

mean of R + √ 2 \* current time T / number of times visited R

**Bibliography**:

[ritvikmath](https://www.youtube.com/@ritvikmath)(2020) Multi-Armed Bandit: Data Science Concepts. Available at: <https://www.youtube.com/watch?v=e3L4VocZnnQ>

[ritvikmath](https://www.youtube.com/@ritvikmath)(2020) Best Multi-Armed Bandit Strategy? (feat: UCB Method). Available at: <https://www.youtube.com/watch?v=FgmMK6RPU1c>

[Nikolay Manchev](https://domino.ai/blog/author/nikolay-manchev)(2023) Reinforcement learning: The K-armed bandit problem. Available at: <https://domino.ai/blog/k-armed-bandit-problem>

[Nikolay Manchev](https://domino.ai/blog/author/nikolay-manchev)(2022) Semi-uniform strategies for solving K-armed bandits. Available at: <https://domino.ai/blog/semi-uniform-strategies-for-solving-k-armed-bandits>

<https://www.youtube.com/watch?app=desktop&v=RPbtzWgzD9M>

<https://www.youtube.com/watch?v=s6UHInwoqb0>

<https://www.geeksforgeeks.org/upper-confidence-bound-algorithm-in-reinforcement-learning/>