Upper Confidence Bound (UCB)

Imagine you are in a casino and you want to play a slot machine. You are faced with many slot machines and you don't know which one will give you the biggest payout. You decide to use the UCB algorithm to help you make your decision.

The UCB algorithm works by keeping track of how much reward you have received from each slot machine and how many times you have played it. It then calculates a score for each slot machine based on the reward it has given you so far and how many times you have played it.

The score is calculated using a formula that balances exploration and exploitation. The algorithm wants to explore new slot machines to see if they will give bigger rewards, but also wants to exploit the slot machines that have given bigger rewards in the past.

Once the algorithm has calculated the scores for each slot machine, it chooses the one with the highest score and plays it. This process is repeated over and over again, with the algorithm updating the scores and choosing the slot machine with the highest score each time.

The UCB algorithm is useful in many real-life applications, such as online advertising where you want to show ads to users that are most likely to click on them, or in robotics where you want to learn the optimal sequence of actions to achieve a task.

In summary, the UCB algorithm is a way to balance exploration and exploitation in decision-making. It is useful in many real-life applications and works by keeping track of rewards and the number of times an action has been taken, and using a formula to calculate a score that guides the decision-making process.