

The Multi-Armed Bandit Problem

The Multi-Armed Bandit Problem

Machine Learning A-Z

© SuperDataScience



Machine Learning A-Z

© SuperDataScience

In this dog robot you can implement two algorithms: whether you implement a algorithms that you will show for each foot how to walk and there is second type of algorithms (reinforcement learning) that you will tell all the possible ways and the dog is given a reward for each step forward and given punishment by every fall. the reward and punishment are given in 0 and 1 formats (reward is 1 and punishment is 0). so the dog try all these random sets of actions to see where it leads to. so in here we don't actually implement a walking algorithms.

The Multi-Armed Bandit Problem

Machine Learning A-Z

© SuperDataScience

The Multi-Armed Bandit Problem



Machine Learning A-Z

© SuperDataScience

This is a one armed bandit

The Multi-Armed Bandit Problem



Machine Learning A-Z

© SuperDataScience

A multi-armed bandit problem is a challenge that a person faced when he comes up to a whole set of these machines

The Multi-Armed Bandit Problem



D1

D2

D3

D4

D5

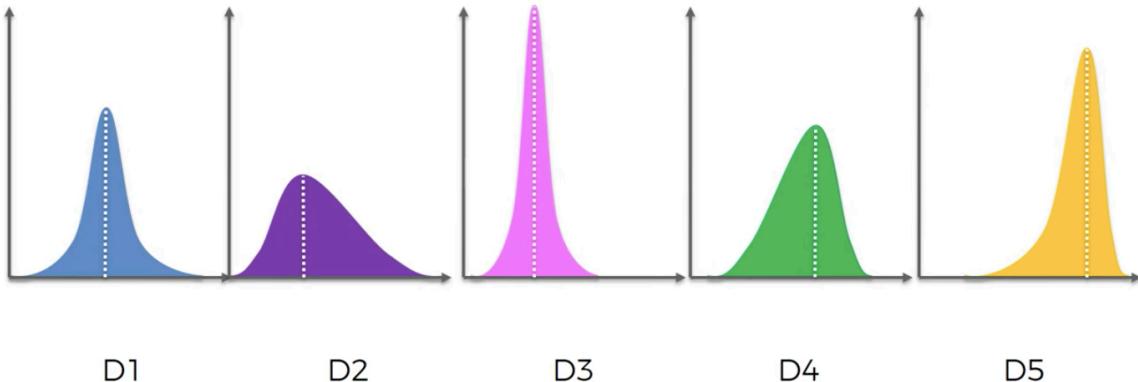
Machine Learning A-Z

© SuperDataScience

This is called multi-armed bandit and one of its application is these slot machines but this problem can applied to many other applications. In here we want to figure out which ones them to play in order to maximize our return

Each of these machines has a distribution behind out of which the machine picks result.

The Multi-Armed Bandit Problem



Machine Learning A-Z

© SuperDataScience

The fifth machine is the best one to bet on because it has the best distribution.

The Multi-Armed Bandit Problem

D1

D2

D3

D4

D5

Examples used for educational purposes. No affiliation with Coca-Cola

Machine Learning A-Z

© SuperDataScience

There are two choices here:

Exploitation Make the best decision given current information

Exploration Gather more information

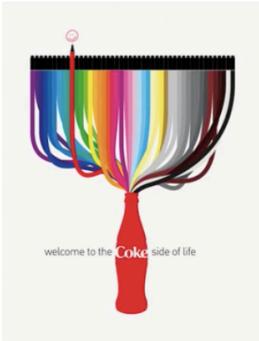
Exploration is good because it finds the best result but on the other hand we need the result fast (exploitation). The balance is the most important part in here

If we don't find the the result fast, it's going to cost us and in other hand, we want the best result.

For more information you can check out the article

Regret is when we use a non optimal machine. in here, if we explore other non-optimal machines the higher regret.
and also if don't explore long enough then we consider a non-optimal machine as a optimal machine.

The Multi-Armed Bandit Problem



D1

D2

D3

D4

D5

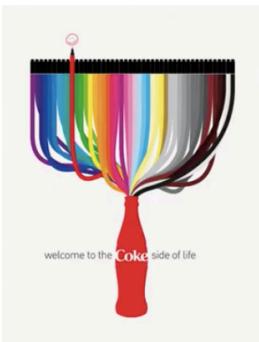
Examples used for educational purposes. No affiliation with Coca-Cola

Machine Learning A-Z

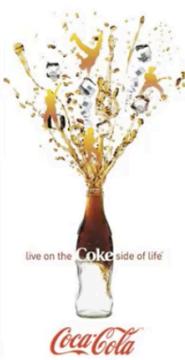
© SuperDataScience

One of the application of the Multi-armed Bandit is advertising, and in here we want to figure out which one is the best.

The Multi-Armed Bandit Problem



D1



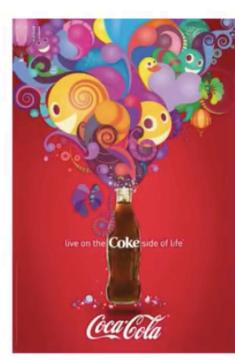
D2



D3



D4



D5

Examples used for educational purposes. No affiliation with Coca-Cola

Machine Learning A-Z

© SuperDataScience

This is used for advertising purpose