Week 1

Sunday, August 25, 2019 8:51 AM

Reinforcement Learning is a subfield under machine learning. In RL, the agent collects its own data by trial and error. Every time, an agent is asked to make a decision, it is given a few options. The agent gets a reward for selecting an option. The agent's job is to maximize the reward for its actions.

If the reward for actions was known, the problem of RL would be trivial. The agent can greedily select the actions with the highest reward. However, the reward most often is not known. What we know are the estimates of these rewards. The estimates are known as action values.

Theoratically, the expected reward is given as :-

$$q^*(a) \doteq \mathbb{E}[R_t \mid A_t = a] \quad \forall a \in \{1, \ldots, k\}$$

ie the expected reward is the summation of expected rewards received if an action a is taken at times 1 .... k.

This can be found by

This can be found by

$$q^*(a) = p(r|a) \cdot r$$

Again the goal is to maximize this expected reward which can also be represented as

$$\arg\max_a q_*(a)$$ is the selection of actions yielding a max expected reward.

In practical terms, we can find this by doing an experiment multiple times and plotting the results or using other statistical methods to find the kind of distribution. Then just find the expected value for that distribution.

↓ mean of the distribution

This problem is called the problem of bandits. It is a smaller problem under RL and is a good intro to the field of RL.