Reinforcement Learning Algorithms

Value Iteration

First step towards optimal policy…

Assumptions:

1. Reward at every state is known.
2. Transition probability is known.

In a way, the dynamics of the environment is completely known.

Algorithm:

Initialize $V^*(s) = 0 \quad \forall s$

For each episode:

For all states 's' in S:

$V_k^*(s) \leftarrow \max_a \sum_{s'} P(s'|s,a)[r + \gamma V_{k-1}^*(s')]$

$\pi_k^*(s) \leftarrow \operatorname{argmax}_a \sum_{s'} P(s'|s,a)[r + \gamma V_{k-1}^*(s')]$

Bellman's Optimality Equation

$$V^*(s) = \max_a \sum P(s'|s,a)\big(r + \gamma V^*(s')\big)$$

$$V_k^*(s) = \max_a \sum_{s'} P(s'|s,a)[r + \gamma V_{k-1}^*(s')]$$

#Episode

The value iteration function **runs through all possible actions** at once to find the maximum action value.

~ This makes it computationally heavy.

https://www.profrkumar.co.in/