Reinforcement Learning Algorithms: Value Based

SARSA Learning: Maths & Algorithm

The beginning of **Reinforcement Learning** algorithms

- The dynamics of the environment are unknown.
- However, this algorithm requires access to:
  - S_t
  - A_t
  - R_t
  - S_{t+1}
  - A_{t+1}

$$Q^*(s,a) = \sum_{s'} p(s',a,s)\left[r + \gamma \max_{a'} Q^*(s',a')\right]$$

$$Q_{k+1}(s_t,a_t) \leftarrow Q_k(s_t,a_t) + \alpha \cdot [r_t + \gamma Q_k(s_{t+1},a) - Q_k(s_t,a_t)]$$

Algorithm:

Initialize $Q(s,a) \in \mathbb{R}$ & $\pi(s) \in A(s)\ \forall s,a$

$G(s,a) \leftarrow empty\ list\ \forall s,a$

For each episode:

Choose $s_0 \in S, a_0 \in A(s_0)$ randomly

Generate an episode starting with s_0, a_0 and following π:

episode: $s_0, a_0, r_1, \dots, s_T, a_T, r_T$

$g \leftarrow 0$

For each step: $t = T, T-1, \dots 0$

$g \leftarrow \gamma g + R_{t+1}$

Unless the pair s_t, a_t appears in $s_0, a_0, s_1, \dots, s_{t-1}, a_{t-1}$:

Append g to $G(s_t, a_t)$

$Q(s_t, a_t) \leftarrow avg\big(G(s_t, a_t)\big)$

$\pi(s_t) \leftarrow \text{argmax}_a Q(s_t, a)$

https://www.profkumar.co.in/

Reinforcement Learning Algorithms: Value Based

Q – Learning: Maths & Algorithm

The beginning of **Reinforcement Learning** algorithms

- The dynamics of the environment are unknown.
- This algorithm requires access to:
  - S_t
  - A_t
  - R_t
  - S_{t+1}

$$Q^*(s,a) = \sum_{s'} p(s'|s,a)\left[r + \gamma \max_{a'} Q^*(s',a')\right]$$

Algorithm:

Initialize $Q(s,a) \in \mathbb{R}$ & $\pi(s) \in A(s)$ $\forall s \in S^+, a \in A(s)$

$Q(s_T,.) = 0$

For each episode:

Choose $s_0 \in S$

For each step:

Choose a from s using policy derived from Q (e.g. $\epsilon - greedy$)

Take action A, observe R, S' $Q(S,A) \leftarrow$

$Q(S,A) + \alpha\left[R + \gamma \max_a Q(S',a) - Q(S,A)\right]$

$S \leftarrow S'$

until S is terminal

$$Q_{k+1}(s_t,a_t) \leftarrow Q_k(s_t,a_t) + \alpha \cdot \left[r_t + \gamma \max_a Q_k(s_{t+1},a) - Q_k(s_t,a_t)\right]$$

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