

What to Learn? Estimating Action Values

Summary:

Lesson 2: What to Learn? Estimating Action Values

$q_*(a)$ is not known, so we estimate it.

- Sample - Average Method

$Q_t(a) \doteq \frac{\text{sum of rewards when } a \text{ taken prior to } t}{\text{number of times } a \text{ taken prior to } t}$

$$= \frac{\sum_{i=1}^{t-1} R_i}{t-1}$$

KLONG

Action Selection (greedy action)

$$a_g = \arg \max Q(a)$$

- Incremental update rule

$$Q_{n+1} = \frac{1}{n} \sum_{i=1}^n R_i$$

$$= \frac{1}{n} \left(R_n + \sum_{i=1}^{n-1} R_i \right)$$

$$= \frac{1}{n} \left(R_n + (n-1) \frac{1}{n-1} \sum_{i=1}^{n-1} R_i \right)$$

$$= \frac{1}{n} (R_n + (n-1) Q_n)$$

$$= \frac{1}{n} (R_n + nQ_n - Q_n)$$

$$= Q_n + \frac{1}{n} (R_n - Q_n)$$

$$\text{New Estimate} \leftarrow \text{Old Estimate} + \text{Step Size} (\text{Target} - \text{Old Estimate})$$

- Decaying past rewards

$$Q_{n+1} = Q_n + \alpha_n (R_n - Q_n)$$

$$= \alpha_n R_n + (1 - \alpha_n) Q_n$$

$$= \alpha R_n + (1 - \alpha) [\alpha R_{n-1} + (1 - \alpha) Q_{n-1}]$$

$$= \alpha R_n + (1 - \alpha) \alpha R_{n-1} + (1 - \alpha)^2 Q_{n-1}$$

$$\rightarrow Q_{n+1} = \alpha R_n + (1 - \alpha) \alpha R_{n-1} + (1 - \alpha)^2 \alpha R_{n-2} + \dots$$

$$+ (1 - \alpha)^{n-1} \alpha R_1 + (1 - \alpha)^n Q_1$$

$$= (1 - \alpha)^n Q_1 + \sum_{i=1}^n \alpha (1 - \alpha)^{n-i} R_i$$

$Q_1 \rightarrow$ initial action value

