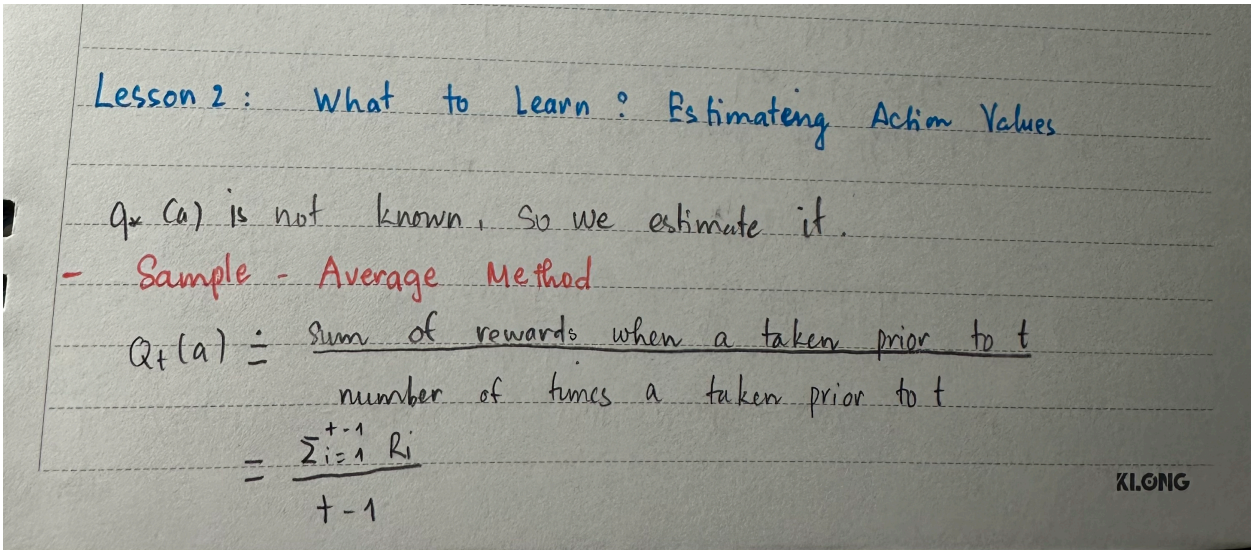
What to Learn? Estimating Action Values

Summary:



Action Selection (greedy action)

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

- Incremental update rule

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

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

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

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

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

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

New Estimate ← Old Estimate + Step Size (Target − Old Estimate)

- Decaying past rewards

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

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

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

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

$$\Rightarrow Q_{n+1} = \alpha R_n + (1-\alpha)\alpha R_{n-1} + (1-\alpha)^2\alpha R_{n-2} + \dots + (1-\alpha)^{n-1}\alpha R_1 + (1-\alpha)^n Q_1$$

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

Q_1 → initial action value