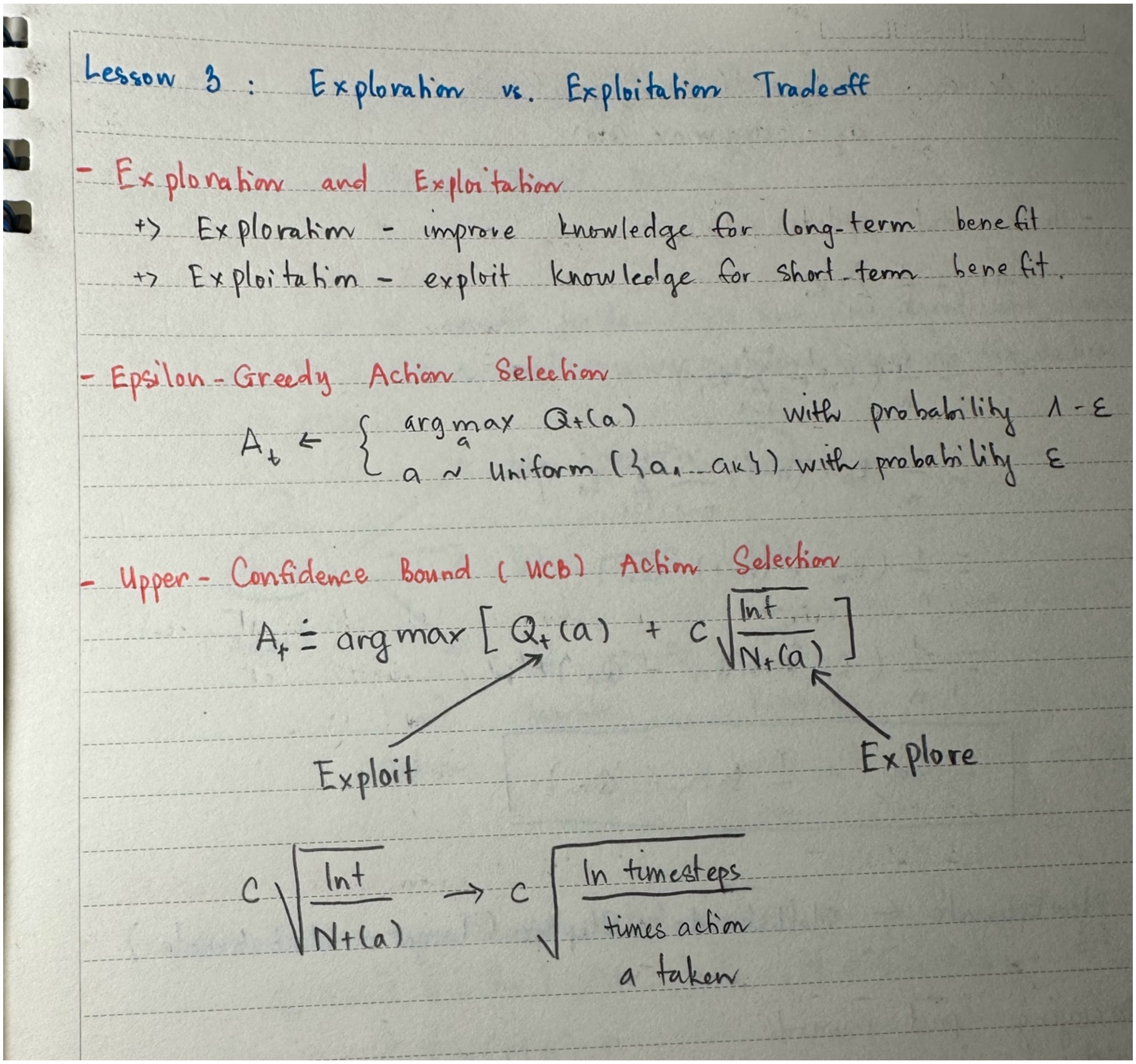
Exploration vs. Exploitation Tradeoff

Summary:



Lesson 3 : Exploration vs. Exploitation Tradeoff

- Exploration and Exploitation
  - +> Exploration - improve knowledge for long-term benefit
  - +> Exploitation - exploit knowledge for short-term benefit.

- Epsilon-Greedy Action Selection

$$A_t \leftarrow \begin{cases} \arg\max_a Q_t(a) & \text{with probability } 1-\varepsilon \\ a \sim \text{Uniform}(\{a_1 \ldots a_k\}) & \text{with probability } \varepsilon \end{cases}$$

- Upper-Confidence Bound (UCB) Action Selection

$$A_t \doteq \arg\max \Big[\underbrace{Q_t(a)}_{\text{Exploit}} + \underbrace{c\sqrt{\frac{\ln t}{N_t(a)}}}_{\text{Explore}} \Big]$$

$$c\sqrt{\frac{\ln t}{N_t(a)}} \rightarrow c\sqrt{\frac{\ln \text{ timesteps}}{\text{times action a taken}}}$$