(1) Epsilon-Greedy

$$a_t = \begin{cases} \arg\max_i Q_{t-1}(i), & \text{with probability } 1-\varepsilon, \\ \text{a random arm in } \{1,\ldots,K\}, & \text{with probability } \varepsilon, \end{cases}$$

$$N_t(a_t) = N_{t-1}(a_t) + 1, \quad Q_t(a_t) = Q_{t-1}(a_t) \; + \; \frac{1}{N_t(a_t)}\bigl(r_t - Q_{t-1}(a_t)\bigr).$$

(2) UCB (Upper Confidence Bound)

$$a_t = \arg\max_i \Bigl[\, Q_{t-1}(i) \; + \; c\sqrt{\frac{\ln t}{N_{t-1}(i)}}\,\Bigr],$$

$$N_t(a_t) = N_{t-1}(a_t) + 1, \quad Q_t(a_t) = Q_{t-1}(a_t) \; + \; \frac{1}{N_t(a_t)}\bigl(r_t - Q_{t-1}(a_t)\bigr).$$

(3) Softmax Action Selection

$$P_t(i) = \frac{\exp\bigl(Q_{t-1}(i)/\tau\bigr)}{\sum_{j=1}^{K} \exp\bigl(Q_{t-1}(j)/\tau\bigr)}, \quad a_t \sim P_t(\cdot),$$

$$N_t(a_t) = N_{t-1}(a_t) + 1, \quad Q_t(a_t) = Q_{t-1}(a_t) \; + \; \frac{1}{N_t(a_t)}\bigl(r_t - Q_{t-1}(a_t)\bigr).$$

(4) Thompson Sampling

$$\theta_i \sim \mathrm{Beta}\bigl(\alpha_i, \beta_i\bigr), \quad a_t = \arg\max_i \theta_i,$$

$$\alpha_{a_t} \leftarrow \alpha_{a_t} + r_t, \quad \beta_{a_t} \leftarrow \beta_{a_t} + (1 - r_t).$$

($r_t \in \{0,1\}$ Bernoulli)