

Multi-armed bandit algorithms
on-line learning in stochastic bandits

1. ϵ -greedy

2. Upper Confidence Bound (UCB)

3. Thompson Sampling

4. Linear bandits

5. Contextual bandits

6. Deep bandits

UCB algorithm

Algorithm 1: UCB

1. Initialize $n_i = 0$ for all arms i .

2. For $t = 1$ to T :

 a. Compute $U_i(t) = \bar{\mu}_i + \sqrt{\frac{2 \ln t}{n_i}}$ for all arms i .

 b. Select arm $i_t = \arg \max_i U_i(t)$.

 c. Pull arm i_t and observe reward $r_{i_t}(t)$.

 d. Update $n_{i_t} = n_{i_t} + 1$ and $\bar{\mu}_{i_t} = \frac{1}{n_{i_t}} \sum_{s=1}^t r_{i_t}(s)$.

3. Return $\sum_{t=1}^T r_{i_t}(t)$.

