Upper Confidence Bound (UCB) Algorithm

Step #1. At each round n, we consider two numbers for each ad_i.

number_of_selections

* $N_i(n)$ - the number of times the ad_i was selected up to round n.

sum_of_rewards

* $R_i(n)$ - the sum of rewards of the ad_i up to round n

Step #2. From these two numbers we compute:

* the average reward of ad_i up to round n.

$$\bar{r}_i(n) = \frac{R_i(n)}{N_i(n)}$$

* the confidence interval $[\bar{r}_i(n) - \Delta_i(n), \bar{r}_i(n) + \Delta_i(n)]$ at round n with:

$$\Delta_i(n) = \sqrt{\frac{3}{2}\frac{\log(n)}{N_i(n)}}$$

Step #3. We select the ad_i that has the maximum

$$UCB \rightarrow \bar{r}_i(n) + \Delta_i(n)$$