

Upper Confidence Bound (UCD)

Upper Confidence Bound Intuition (UCB)

Machine Learning A-Z

© SuperDataScience

The Multi-Armed Bandit Problem



Machine Learning A-Z

© SuperDataScience

Total Exploration = Doing A/B test

We can do the A/B test for each but in that case we are doing only exploration and not exploitation and this will cost us a lot

The Multi-Armed Bandit Problem



The Multi-Armed Bandit Problem

- We have d arms. For example, arms are ads that we display to users each time they connect to a web page.
- Each time a user connects to this web page, that makes a round.
- At each round n , we choose one ad to display to the user.
- At each round n , ad i gives reward $r_i(n) \in \{0, 1\}$: $r_i(n) = 1$ if the user clicked on the ad i , 0 if the user didn't.
- Our goal is to maximize the total reward we get over many rounds.

Upper Confidence Bound Algorithm

Step 1. At each round n , we consider two numbers for each ad i :

- $N_i(n)$ - the number of times the ad i was selected up to round n ,
- $R_i(n)$ - 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

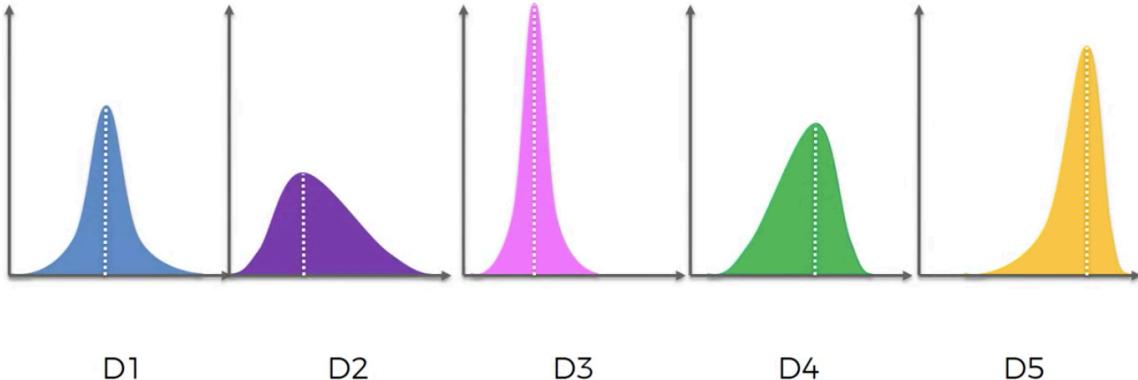
$$\Delta_i(n) = \sqrt{\frac{3 \log(n)}{2 N_i(n)}}$$

Step 3. We select the ad i that has the maximum UCB $\bar{r}_i(n) + \Delta_i(n)$.

Upper Confidence Bound Algorithm



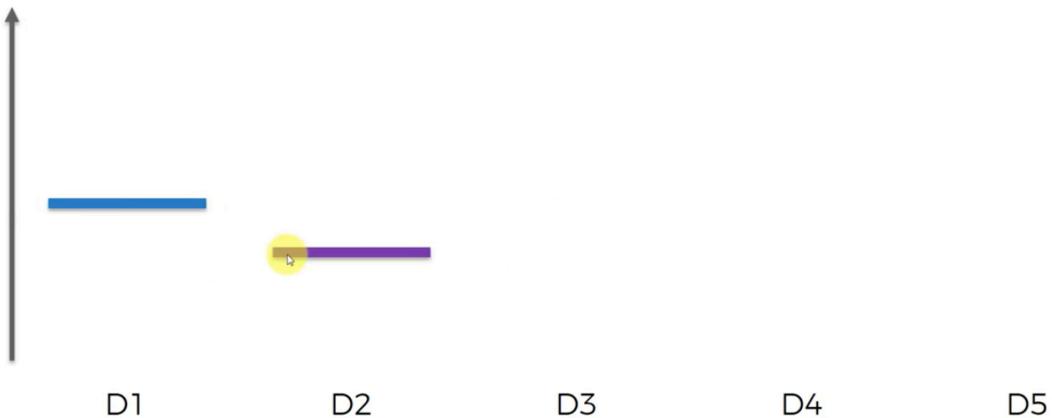
Upper Confidence Bound Algorithm



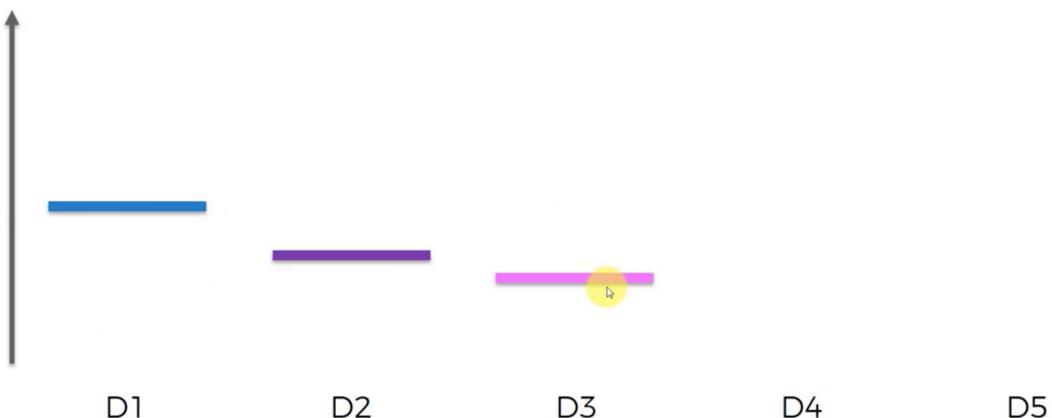
Upper Confidence Bound Algorithm



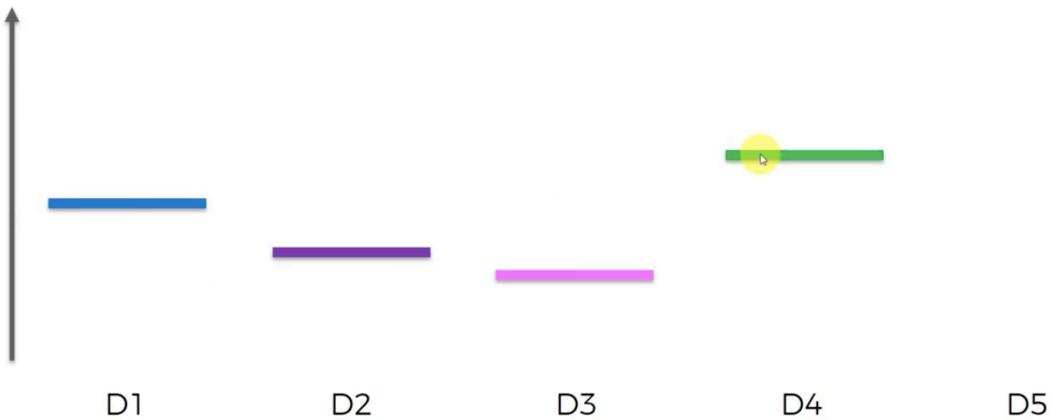
Upper Confidence Bound Algorithm



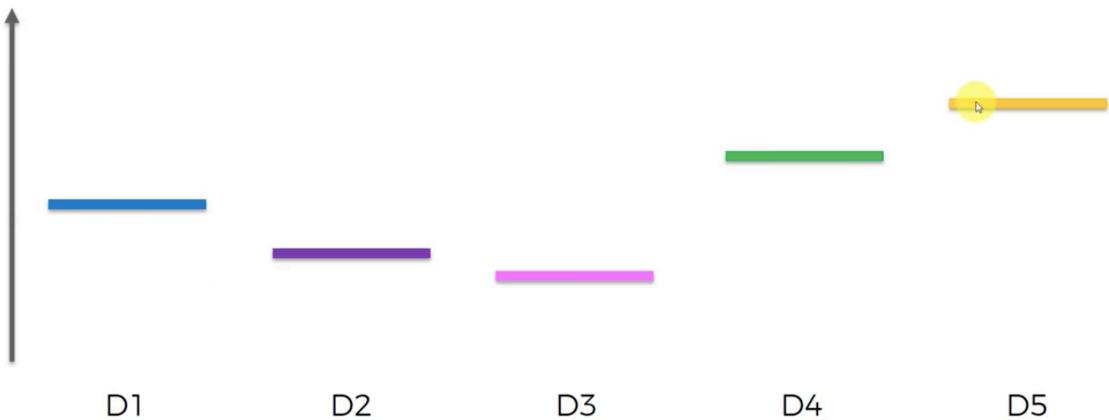
Upper Confidence Bound Algorithm



Upper Confidence Bound Algorithm

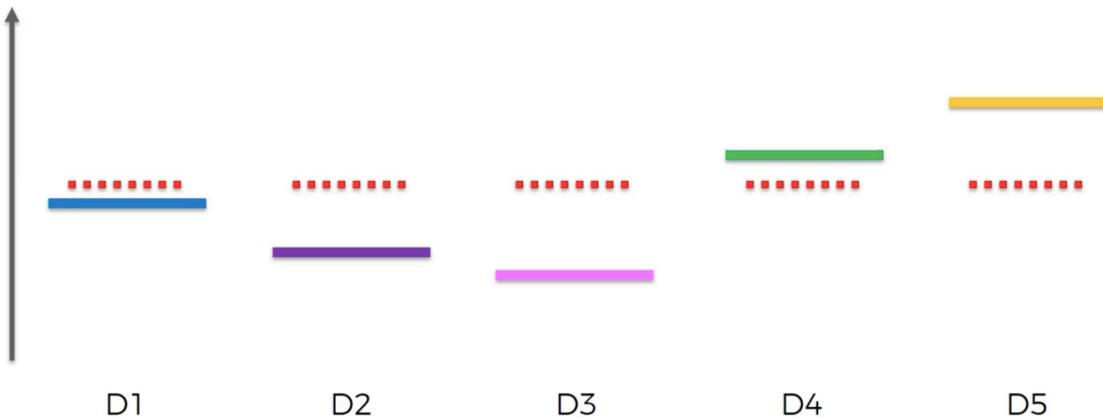


Upper Confidence Bound Algorithm



This is the expected value or return for each of those distribution in each machine.

Upper Confidence Bound Algorithm

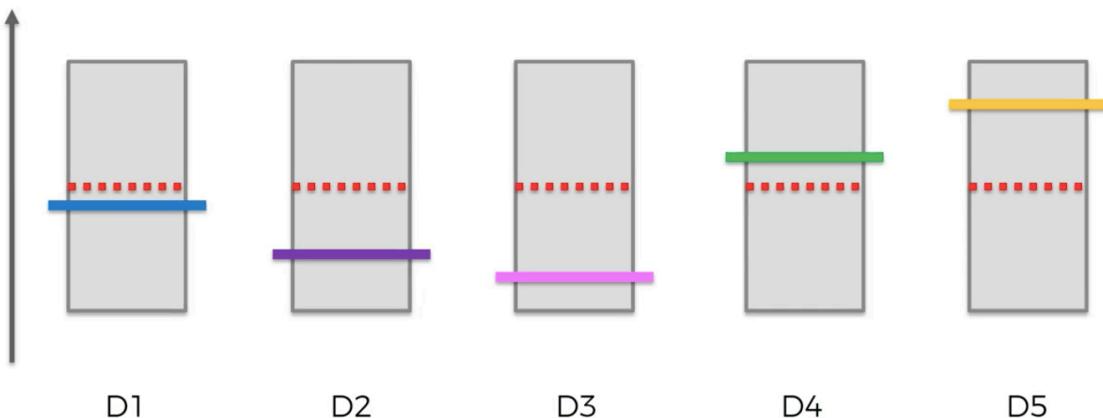


the algorithms assume that there is certain starting value (All the same) which it illustrated with dash line.

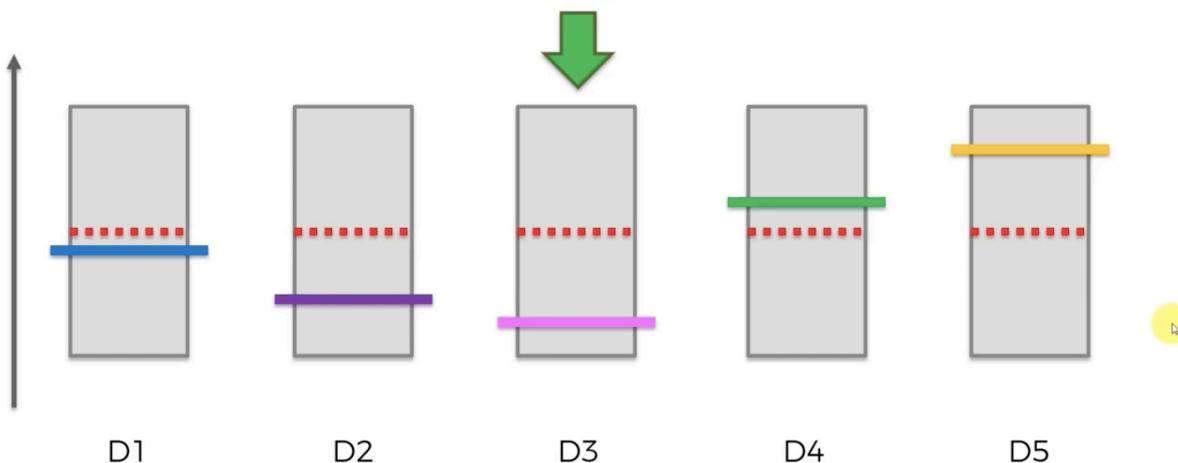
The first couple of rounds going to be trial runs.

After runs, we want our confidence bound fall into the expected value.

Upper Confidence Bound Algorithm



Upper Confidence Bound Algorithm



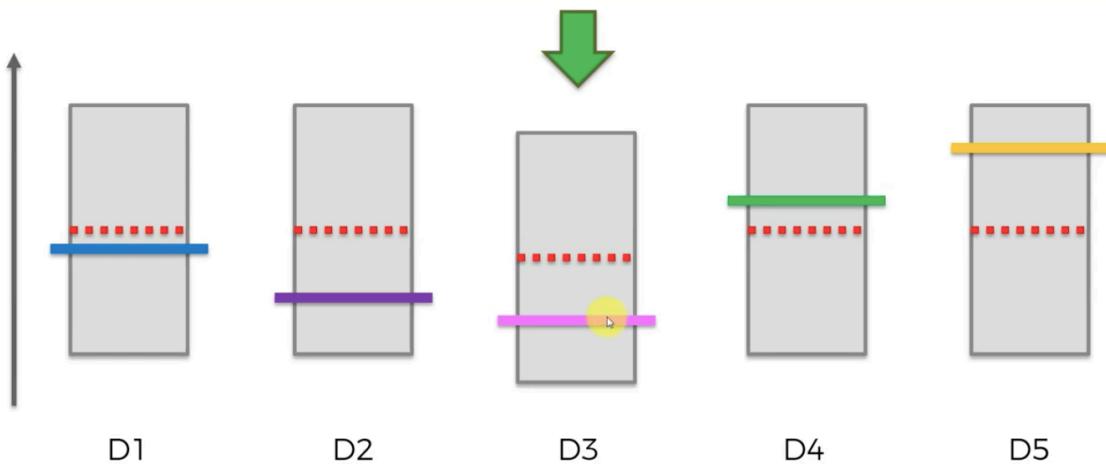
Machine Learning A-Z

© SuperDataScience

Choose the highest confidence bound (only the upper part) which in here we should choose it randomly.

Note: we do not know the expected value. At the moment, we only can see the box.

Upper Confidence Bound Algorithm



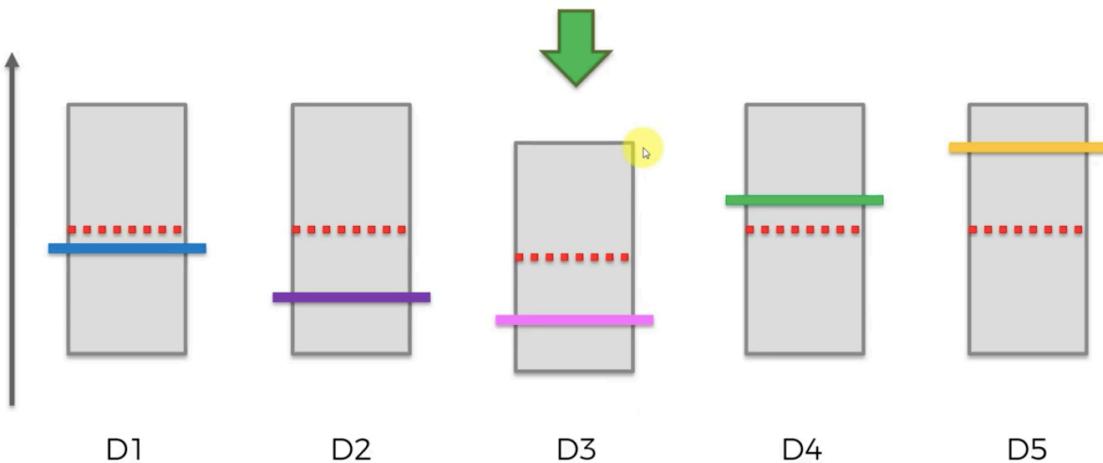
Machine Learning A-Z

© SuperDataScience

Imagine each is an ad and we want to know whether the person click on it or not. if the person didn't click on it then the box goes down, otherwise it goes up.

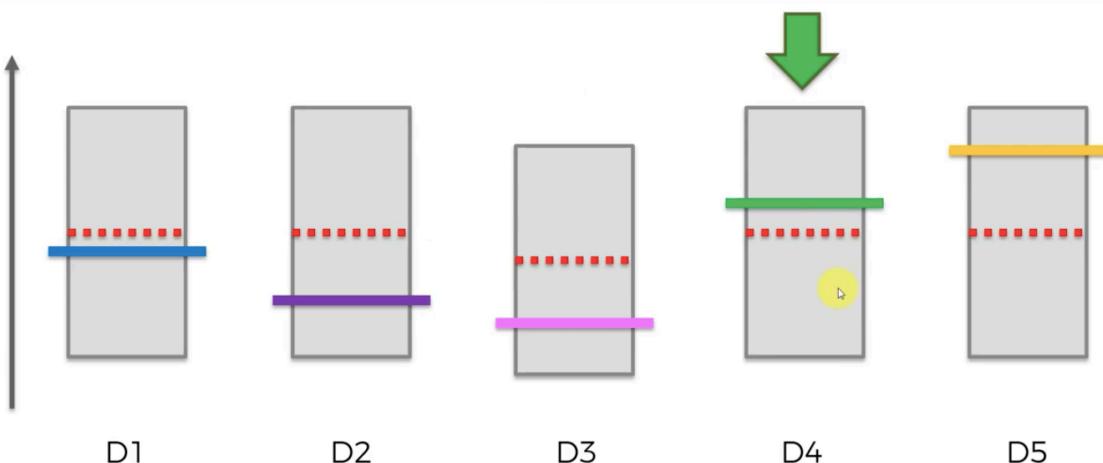
The red dash (observed average) always, in a long run, is going to converge to expected value.

Upper Confidence Bound Algorithm



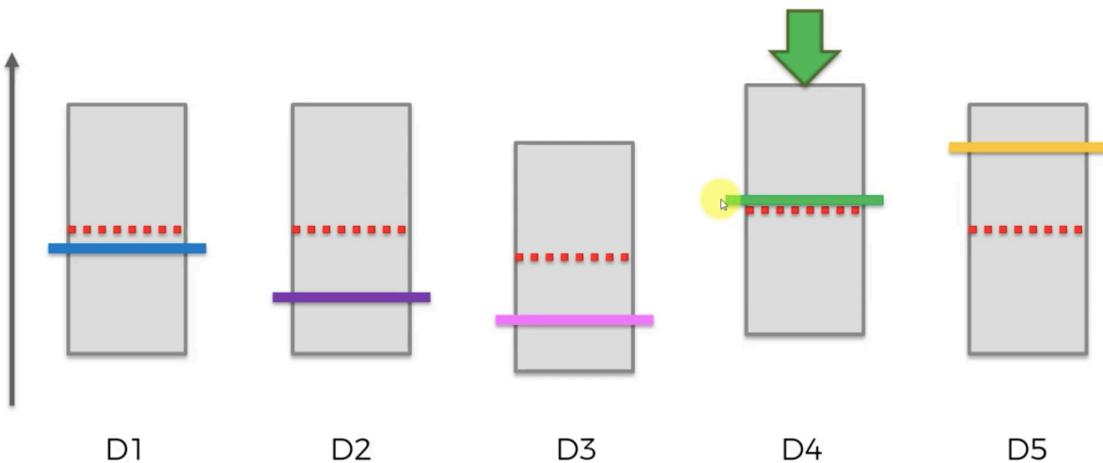
The confidence interval is getting smaller here simply because we have an additional observation.

Upper Confidence Bound Algorithm

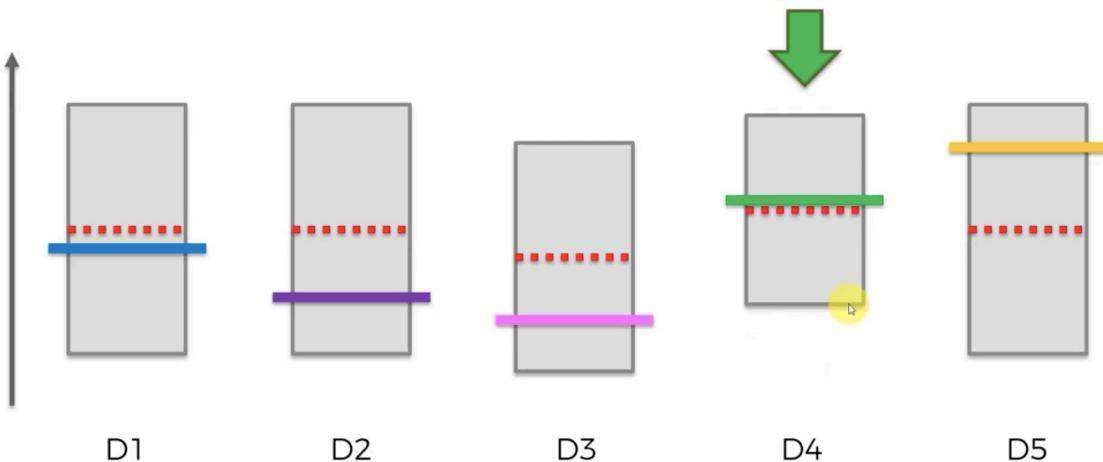


Now choose the highest confidence bound.

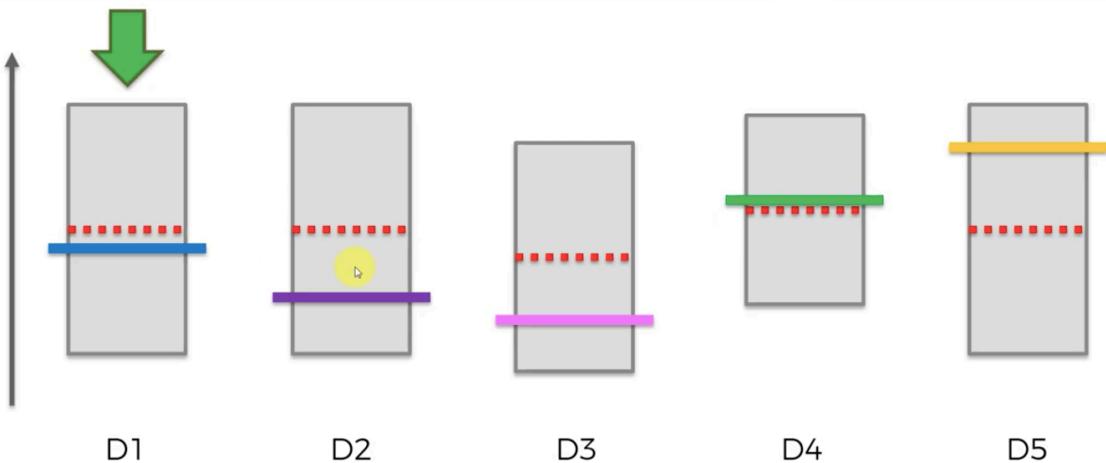
Upper Confidence Bound Algorithm



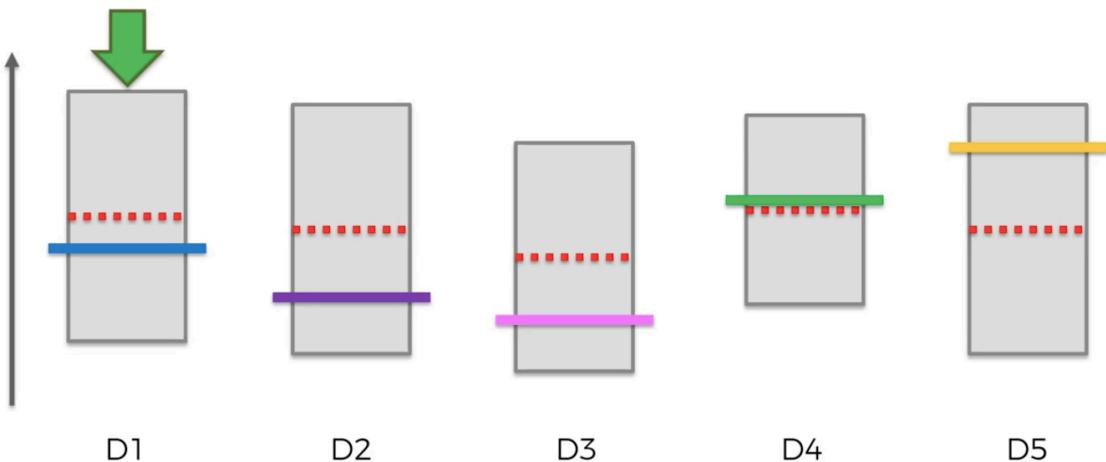
Upper Confidence Bound Algorithm



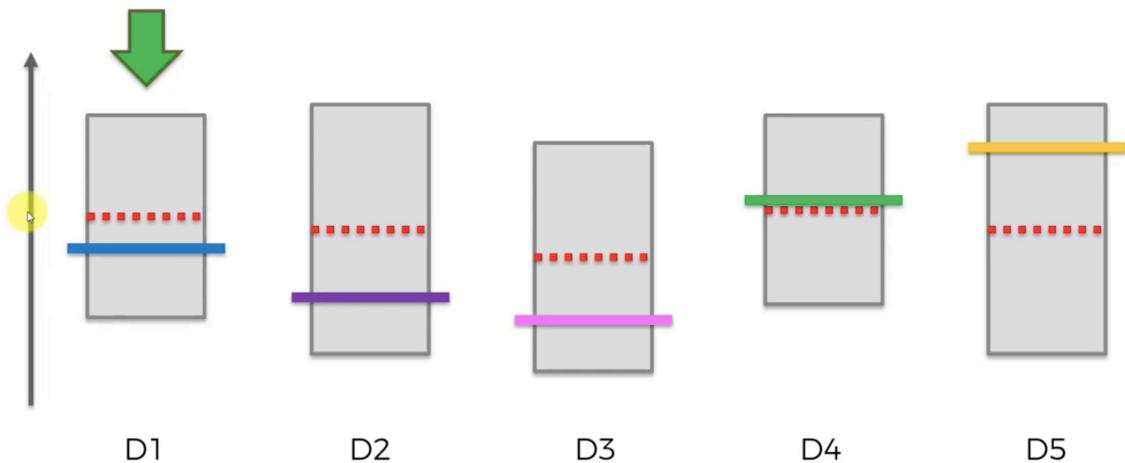
Upper Confidence Bound Algorithm



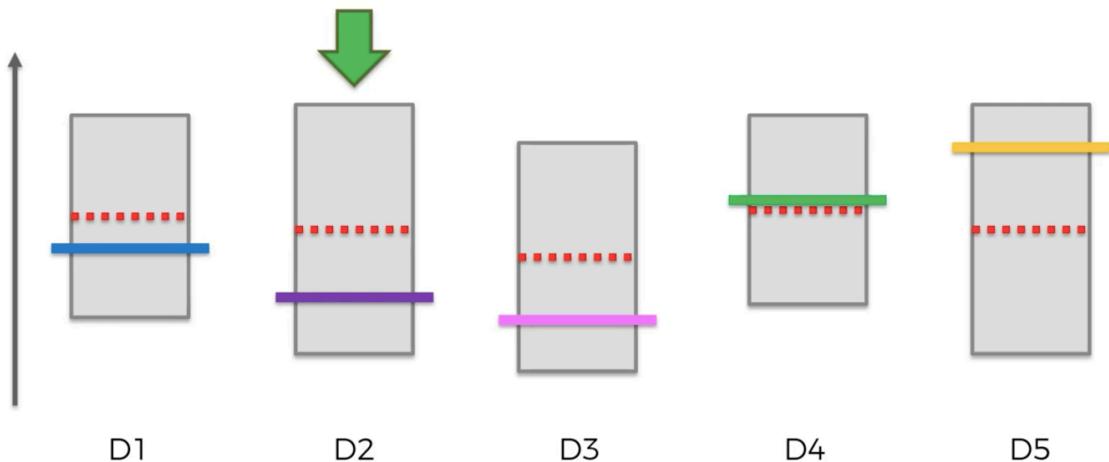
Upper Confidence Bound Algorithm



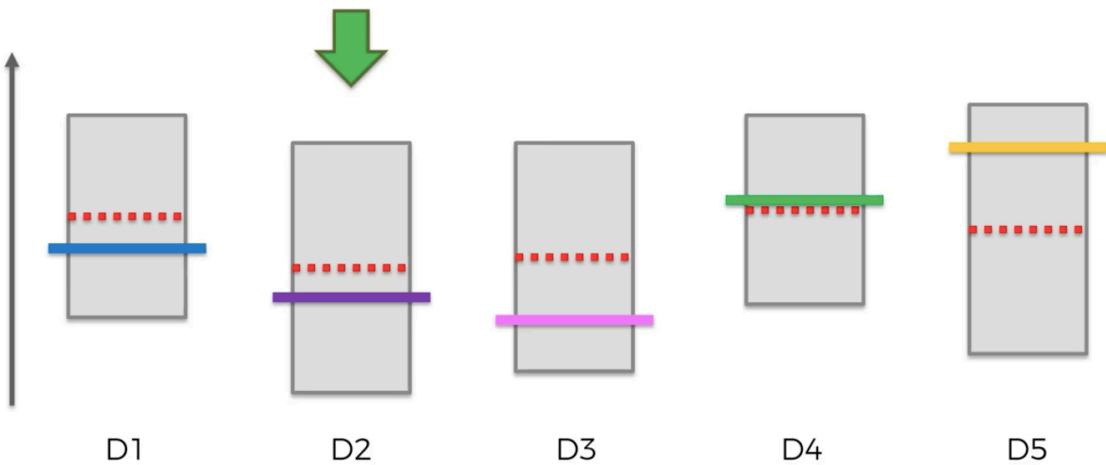
Upper Confidence Bound Algorithm



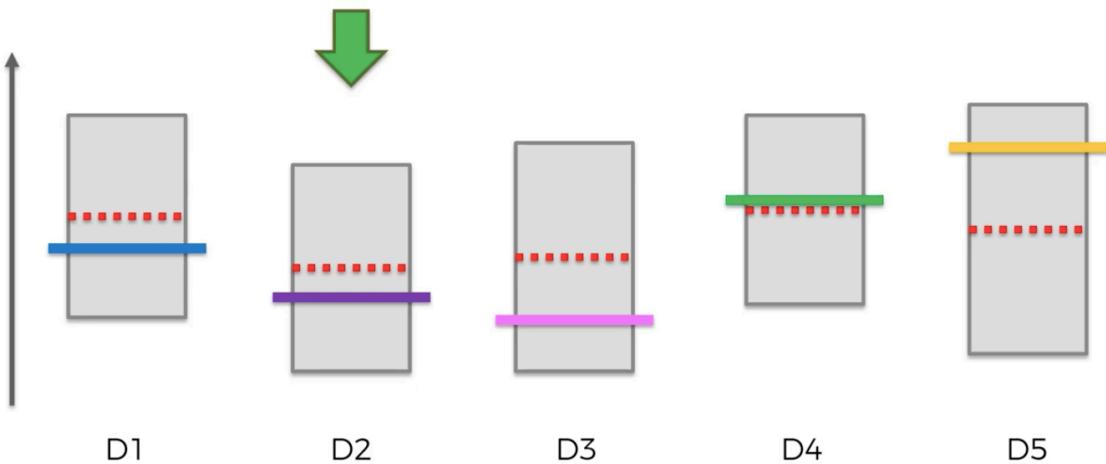
Upper Confidence Bound Algorithm



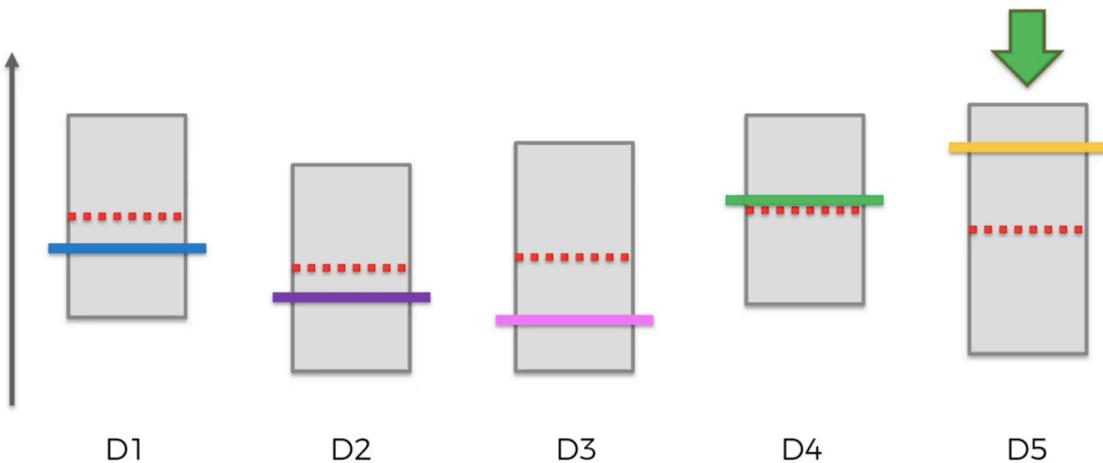
Upper Confidence Bound Algorithm



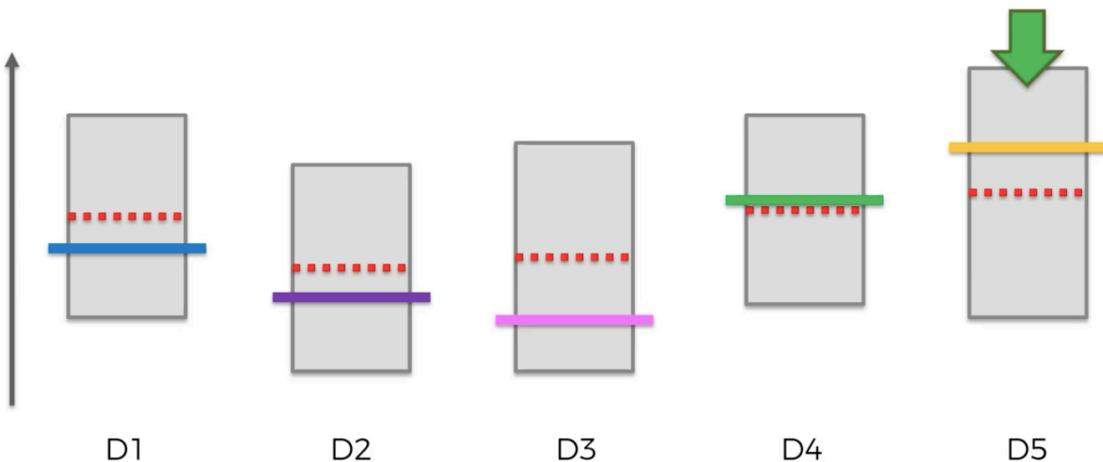
Upper Confidence Bound Algorithm



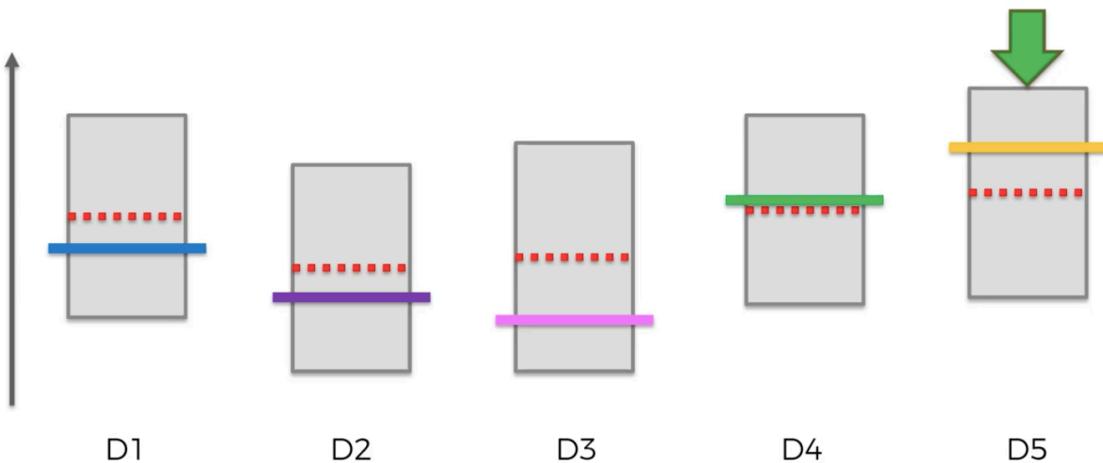
Upper Confidence Bound Algorithm



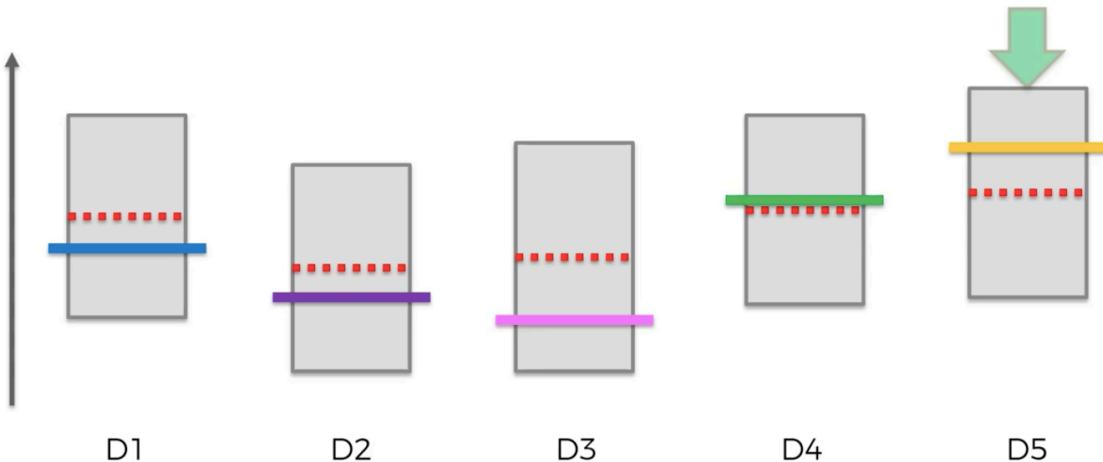
Upper Confidence Bound Algorithm



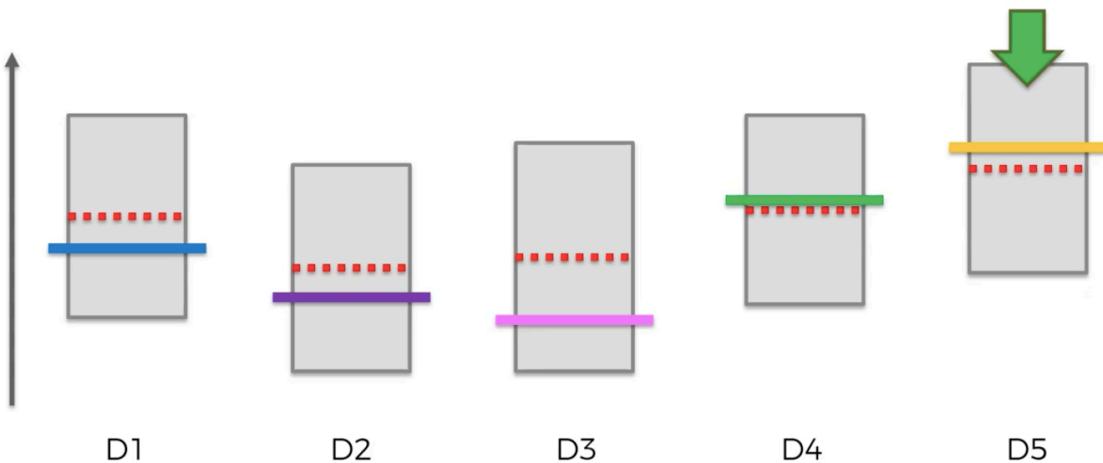
Upper Confidence Bound Algorithm



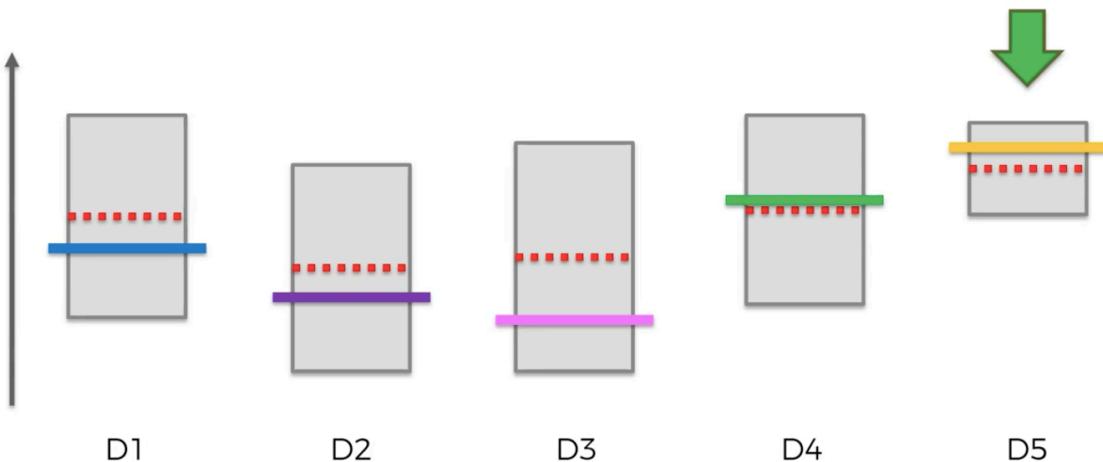
Upper Confidence Bound Algorithm



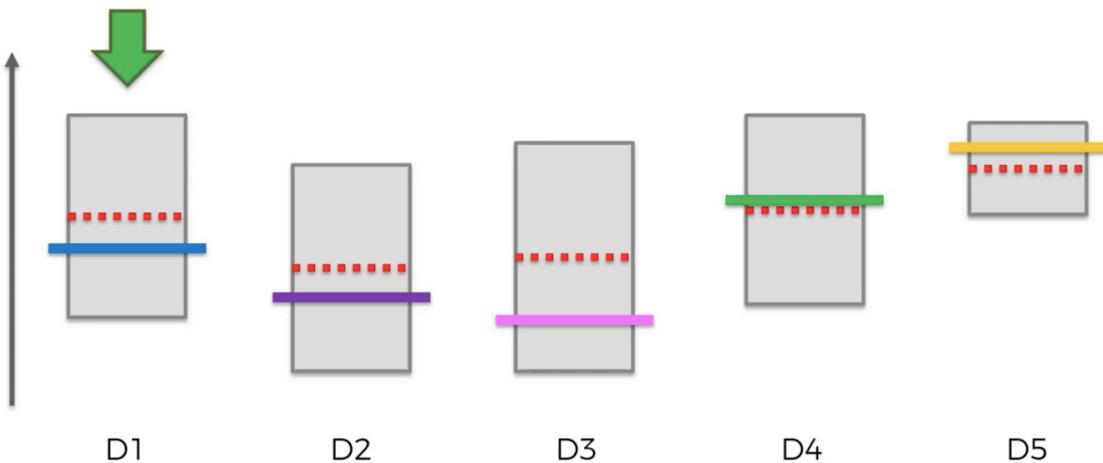
Upper Confidence Bound Algorithm



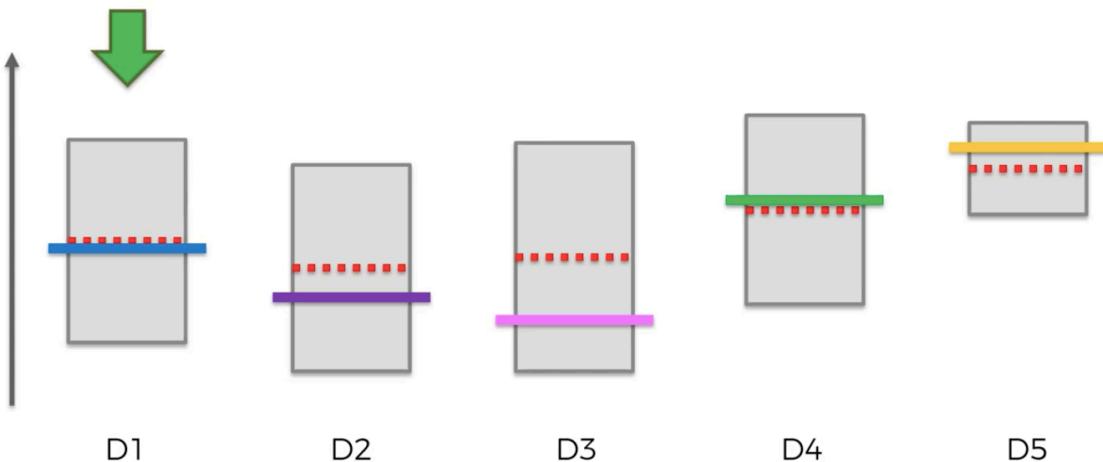
Upper Confidence Bound Algorithm



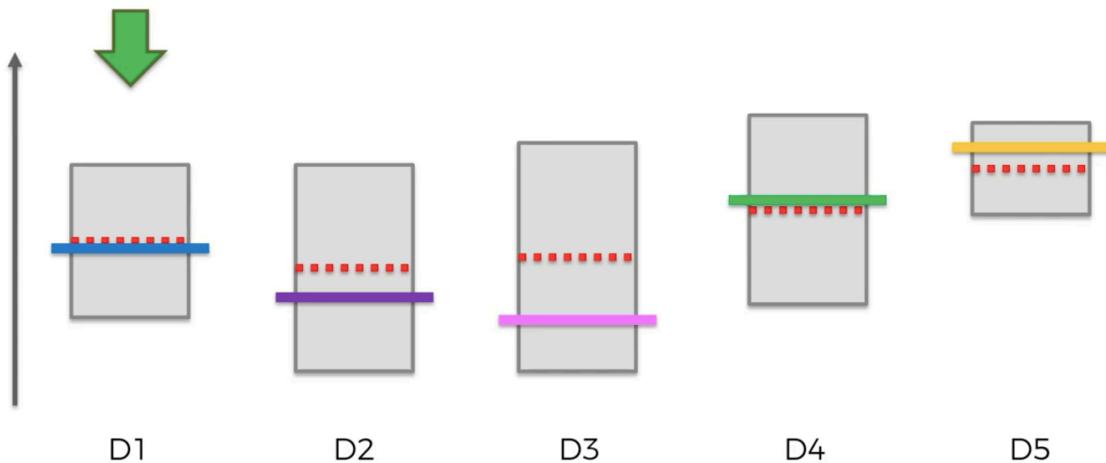
Upper Confidence Bound Algorithm



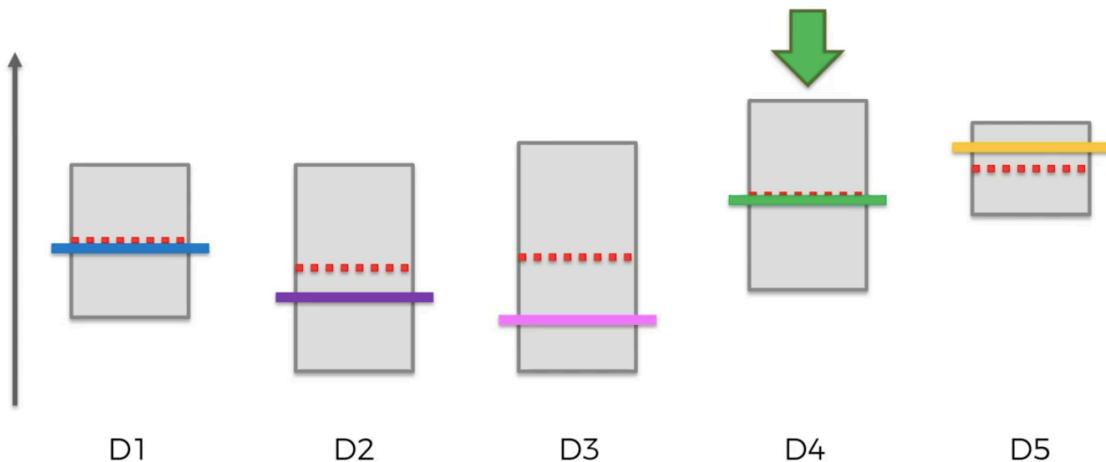
Upper Confidence Bound Algorithm



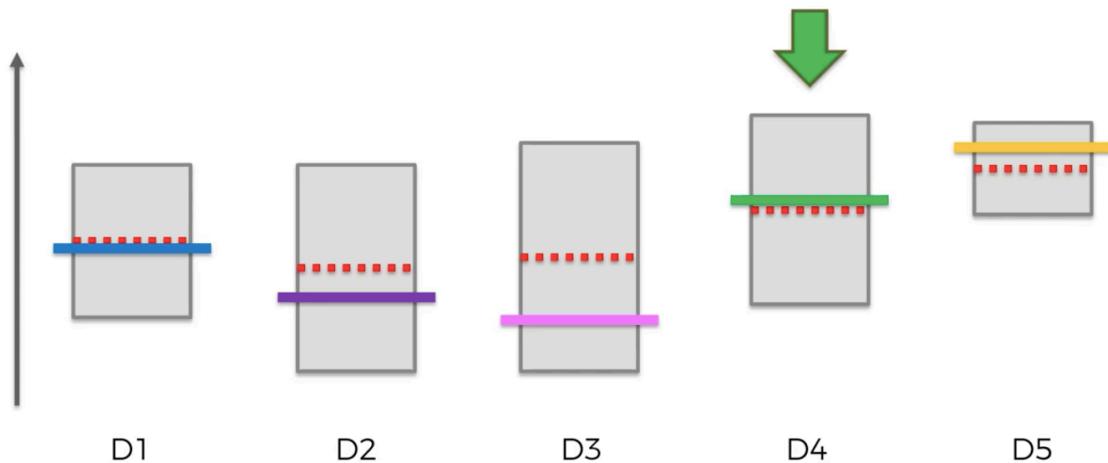
Upper Confidence Bound Algorithm



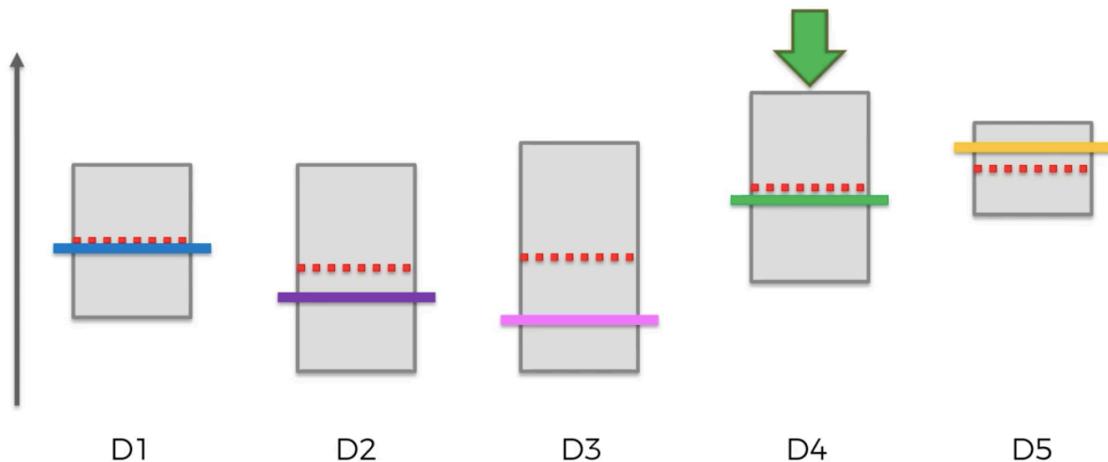
Upper Confidence Bound Algorithm



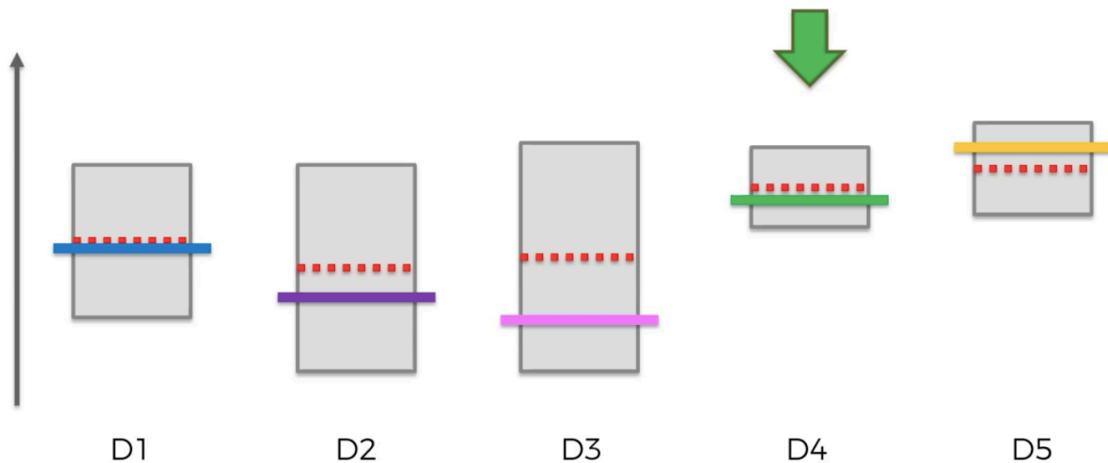
Upper Confidence Bound Algorithm



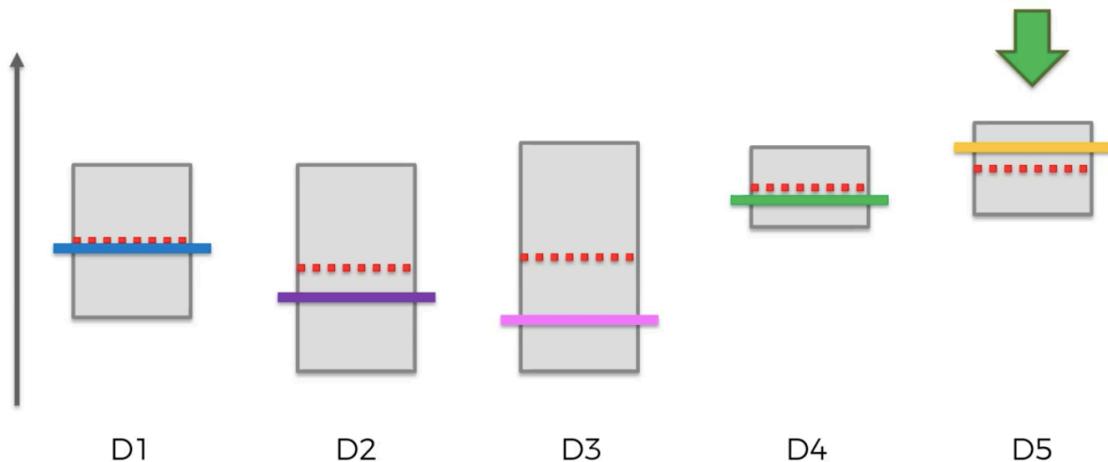
Upper Confidence Bound Algorithm



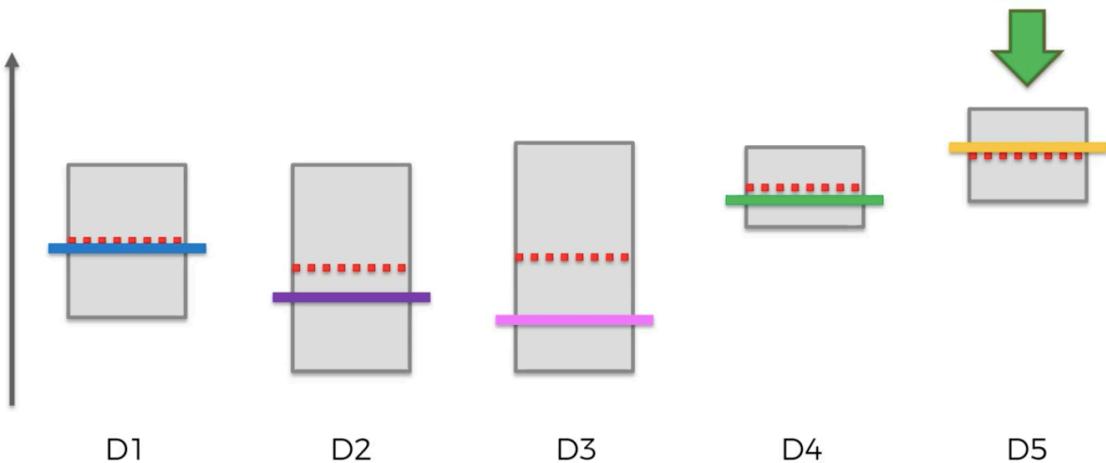
Upper Confidence Bound Algorithm



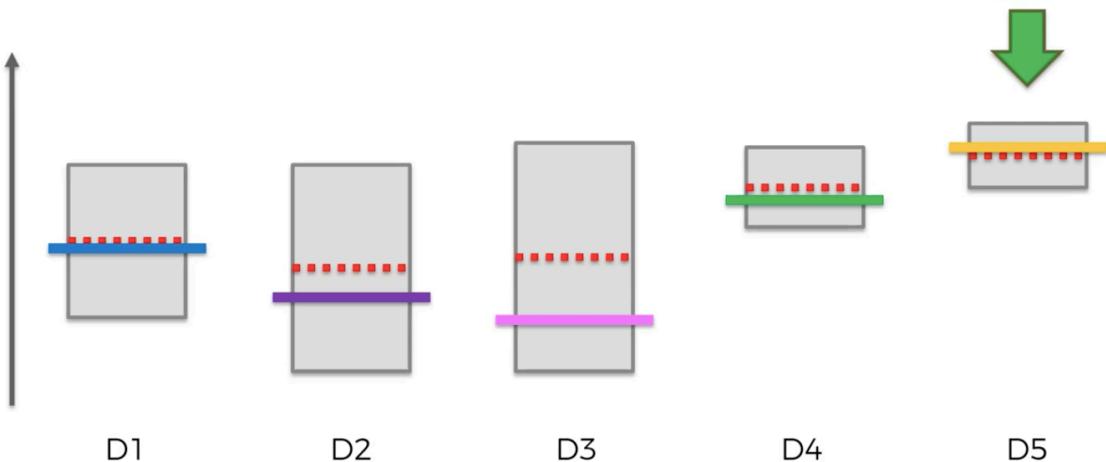
Upper Confidence Bound Algorithm



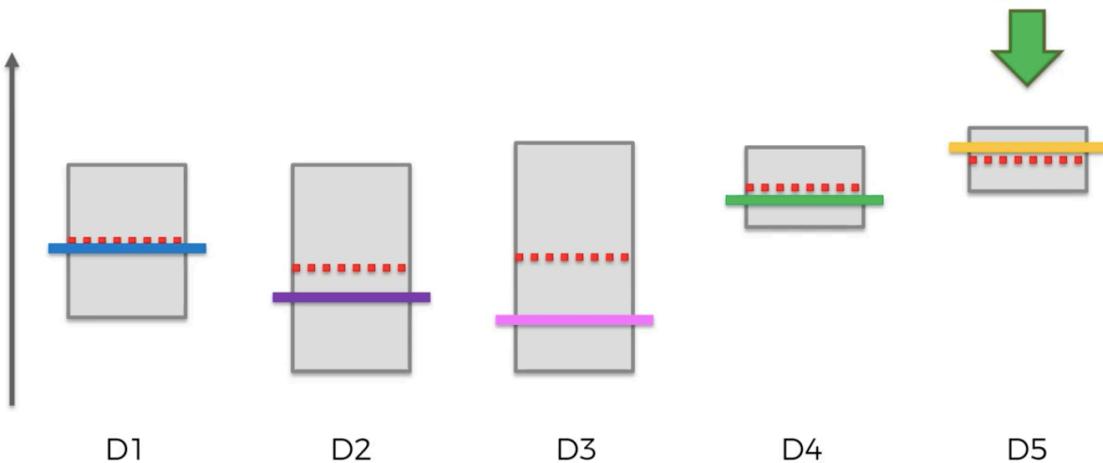
Upper Confidence Bound Algorithm



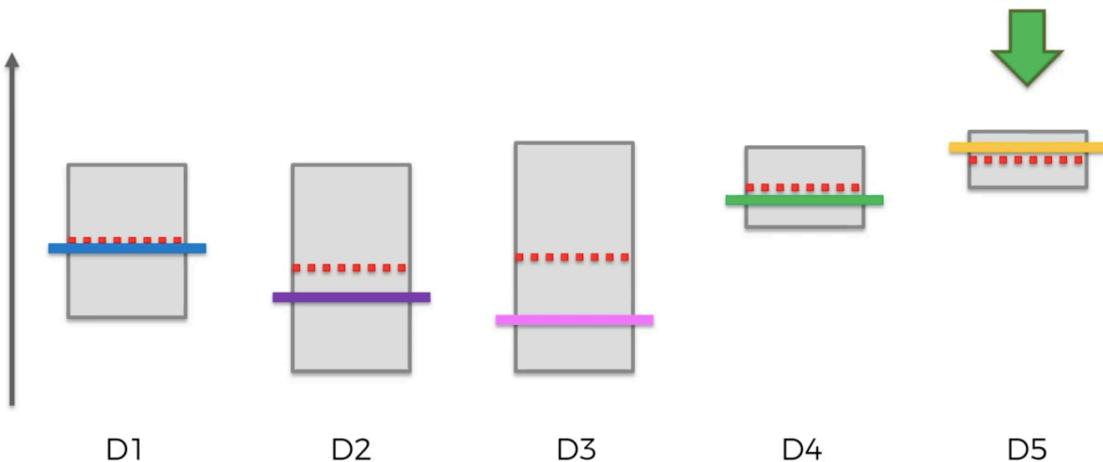
Upper Confidence Bound Algorithm



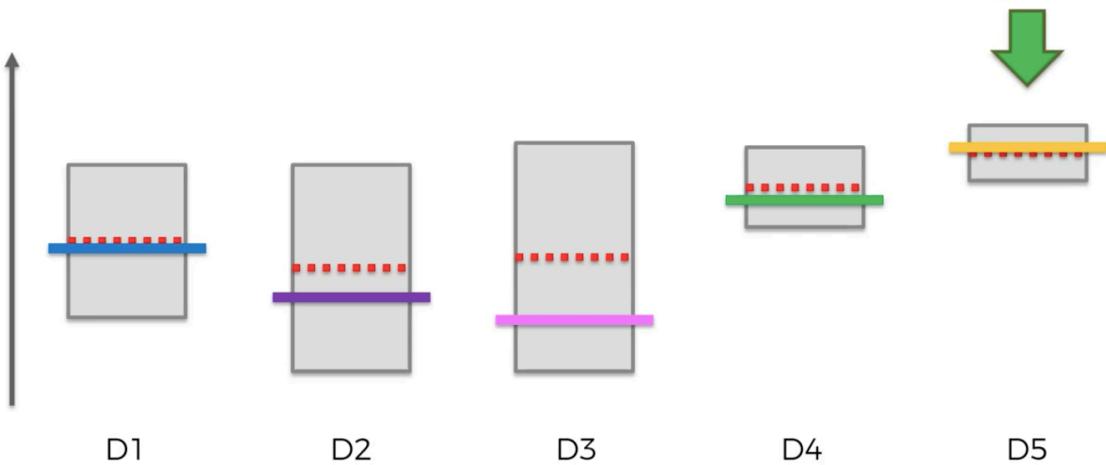
Upper Confidence Bound Algorithm



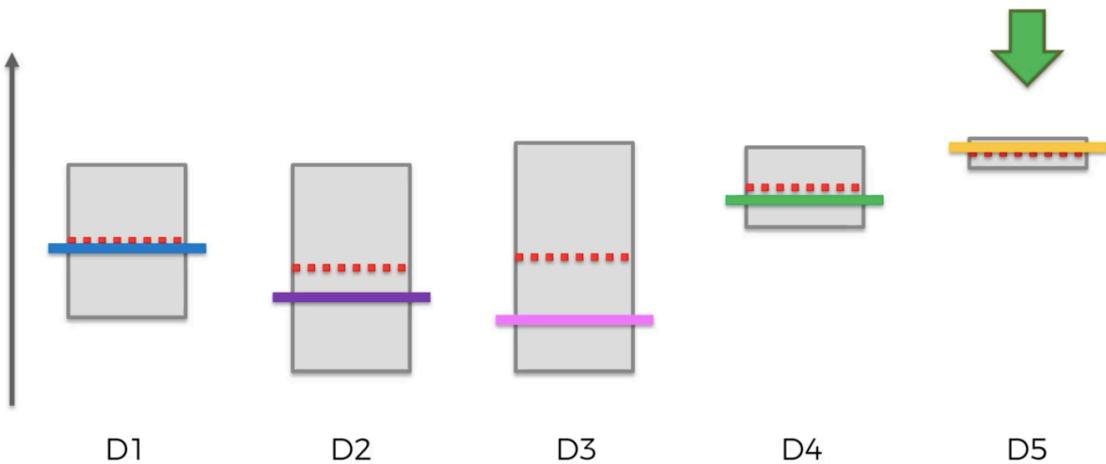
Upper Confidence Bound Algorithm



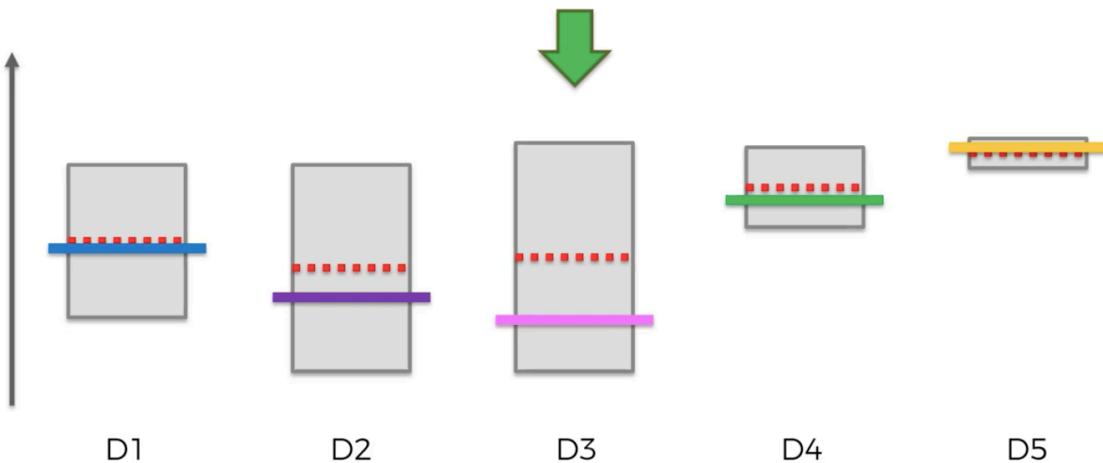
Upper Confidence Bound Algorithm



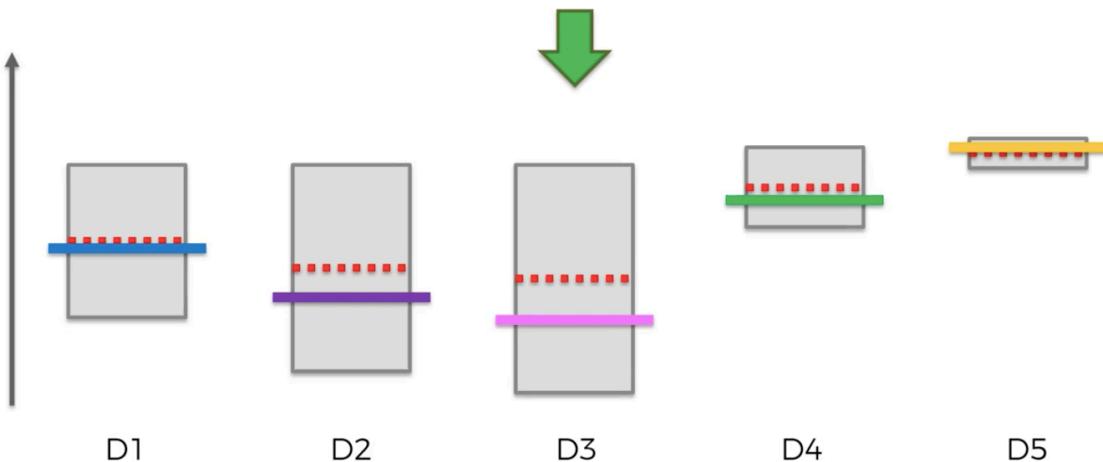
Upper Confidence Bound Algorithm



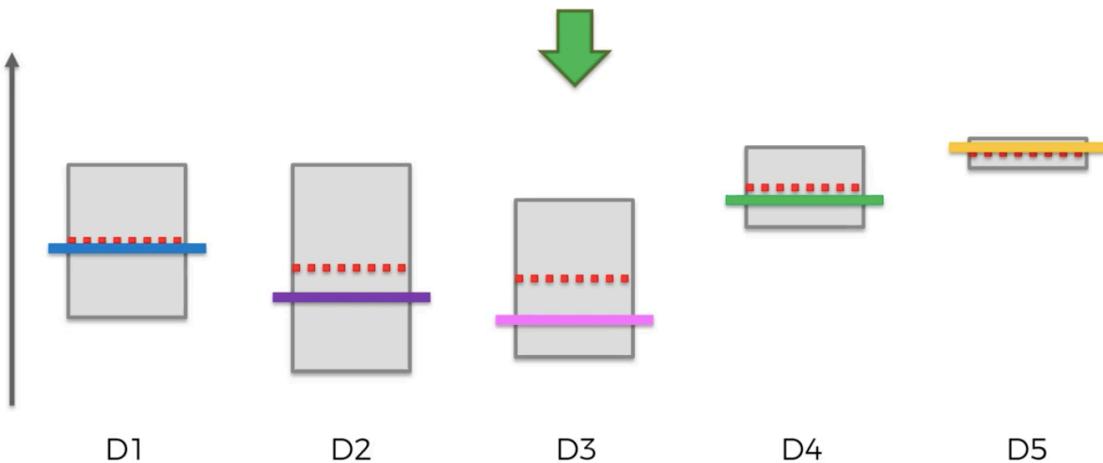
Upper Confidence Bound Algorithm



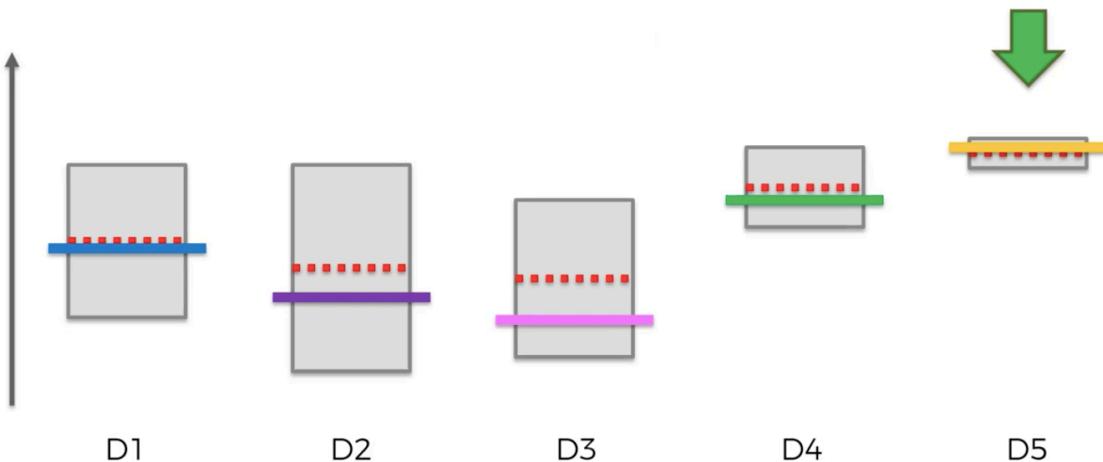
Upper Confidence Bound Algorithm



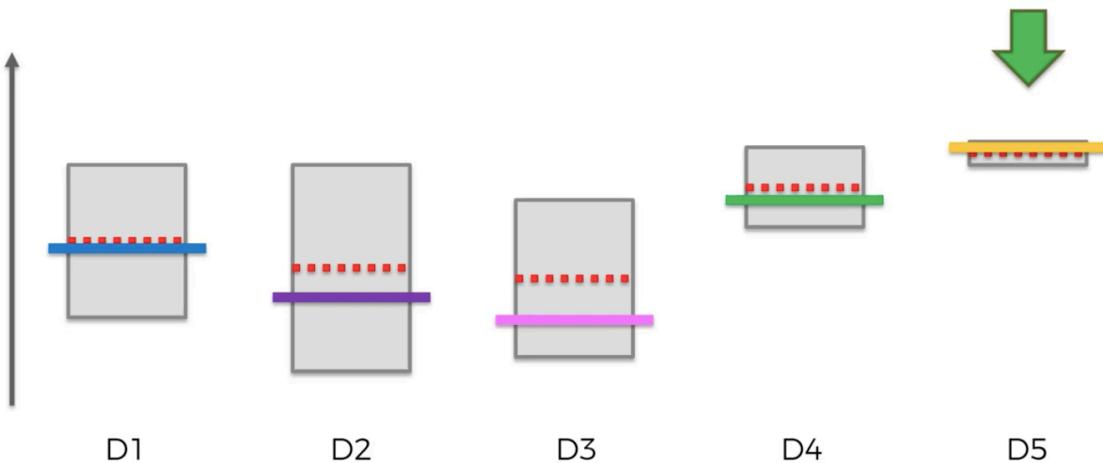
Upper Confidence Bound Algorithm



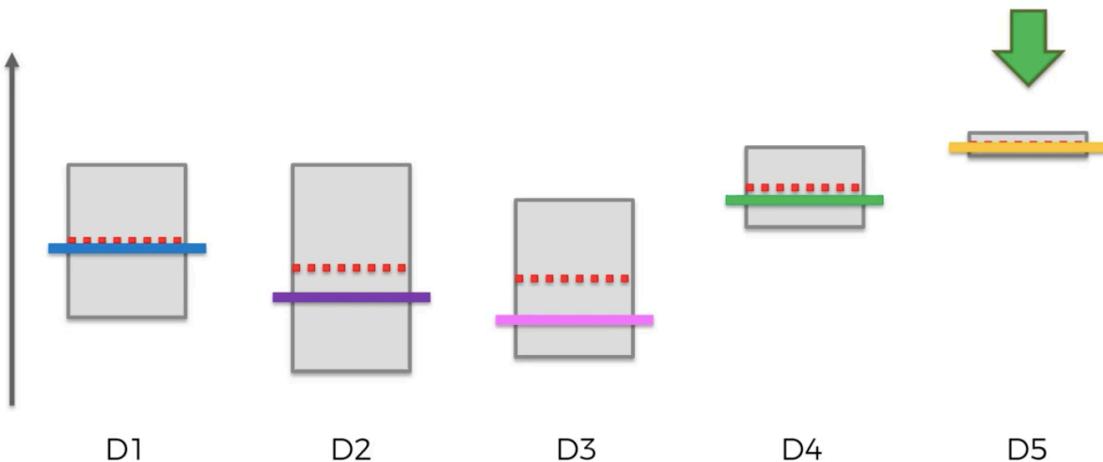
Upper Confidence Bound Algorithm



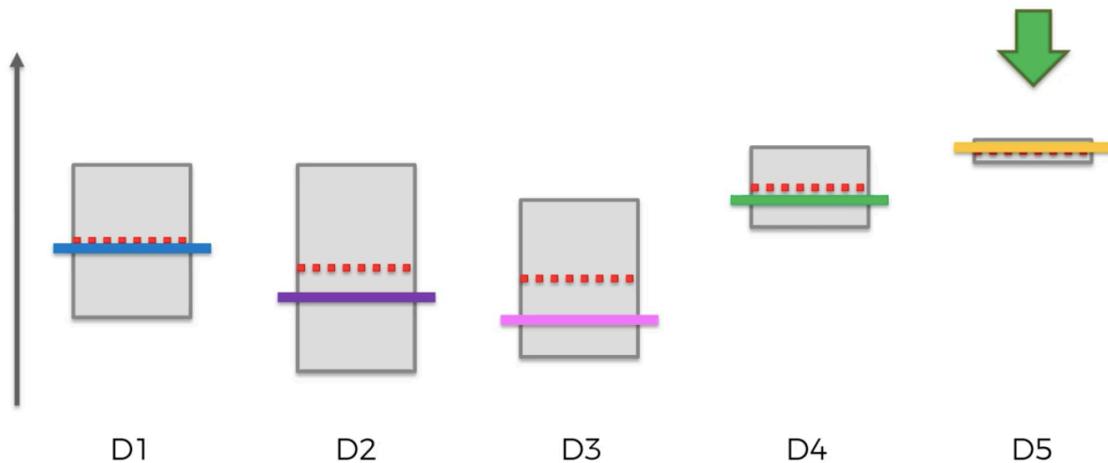
Upper Confidence Bound Algorithm



Upper Confidence Bound Algorithm



Upper Confidence Bound Algorithm



Upper Confidence Bound Algorithm

