$$UCB1(s_i) = \frac{w_i}{n_i} + c\sqrt{\frac{\ln t_i}{n_i}}$$

w = no. of wins after the ith move

n = no. of simulation after ith move

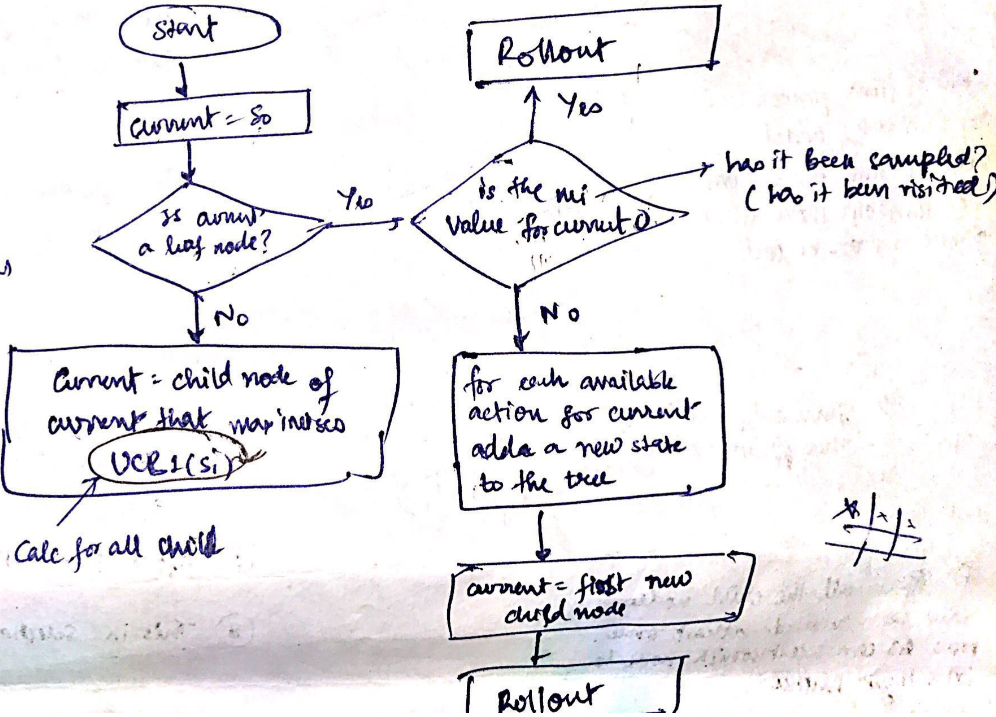
c = exploration parameter

t = total no. of simulations for parent-node

$$v_i = \frac{t_i}{n_i}$$

$$UCBT(s_i) = v_i + c\sqrt{\frac{\ln t_i}{n_i}}$$

Algorithm for tree traversal and node expansion:



(4) 2 stages

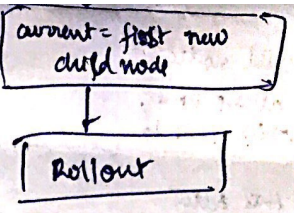
- Selection
- Expansion
- Simulation
- Backprop / update

MCTS

MCSI

MCTS

Rollout (S_i) (Simulation)



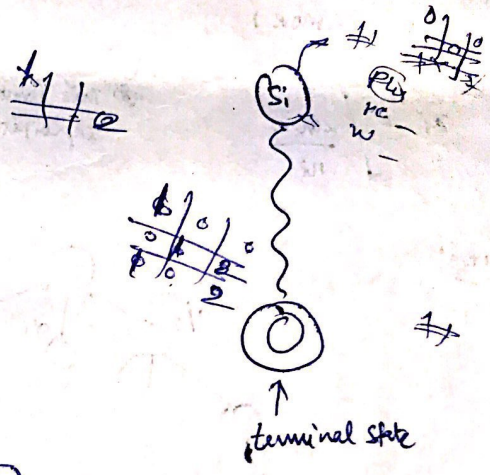
Loop forever

if S_i is terminal state:

return value (S_i)

A_i = random (available_actions (S_i))

S_i = simulate (A_i, S_i)



Worked example:

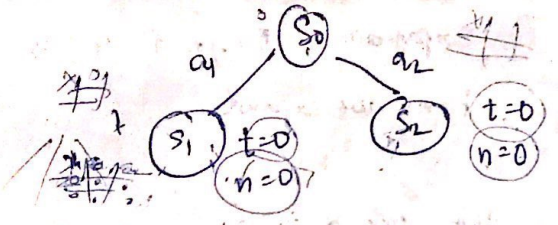
$$UCB1(S_i) = \bar{V}_i + C\sqrt{\frac{\ln N}{n_i}}$$

exploitation

exploration

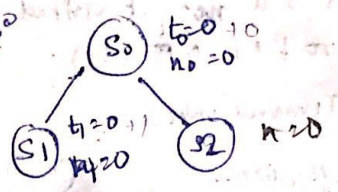
Let $c = 2$

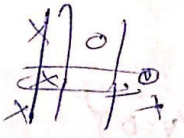
Significance of $UCB1(S_i)$: It depicts the tradeoff b/w exploitation and exploration, c is used to adjust b/w the two



UCB1(S_i) = infinite for both S_1 and S_2 as $n=0$ for both

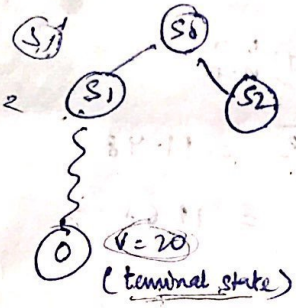
hence, we can select any.





As it is a leaf node, we check, has it been visited or not yet?

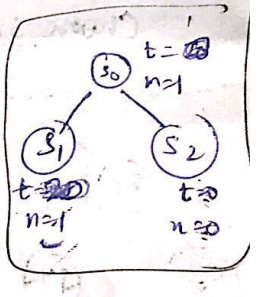
S_1 here hasn't been visited, therefore ROLL OUT(S_1)



Now, we go, to back propogation phase

So, $t_1 = 20$, $n_1 = 1$ } for S_1

$t_0 = 20$, $n = 1$ } for S_0.



After 1st Itr

So, you can run it for any no. of iterations.

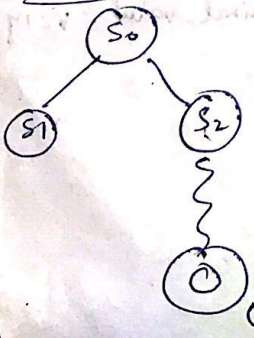
Itr 2

UCB1(S_1) = $20 + 2\times\sqrt{\frac{\ln 1}{1}}$ = 20

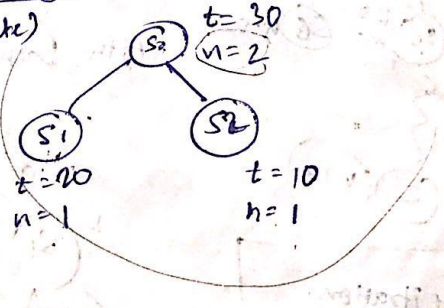
UCB1(S_2) = ∞

Hence, we visit S_2, (in general if you have a sibling with $n=0$, then it is chosen first)

Rollout(S_2)



After Backprop stage:
(2nd Itr Complete)



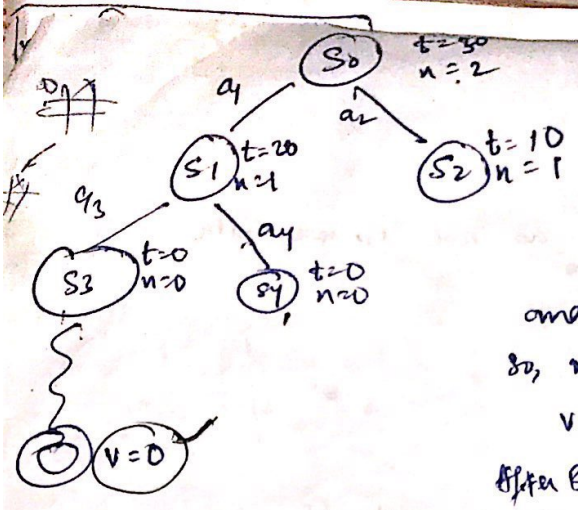
Itr 3

UCB1(S_1) = $20 + 2\sqrt{\frac{\ln 2}{1}}$

= 21.67

UCB(S_2) = $10 + 2\sqrt{\frac{\ln 2}{1}}$

= 11.67



Expansion: ts for S1 n ≠ 0

Hence we expand

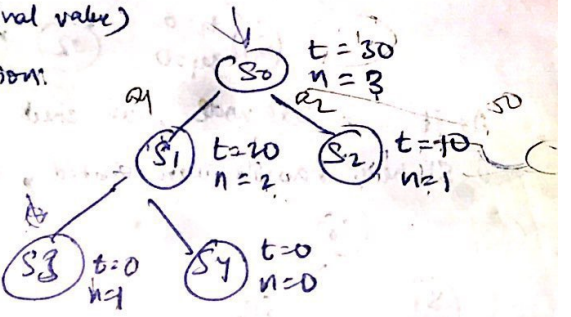
Chose one of S3 & S4, but they have same value, hence chose S3

and as it is a not yet visited leaf node,

So, we do ROLLOUT (S3):

V=0 (Terminal value)

After Backpropogation:



3rd iter

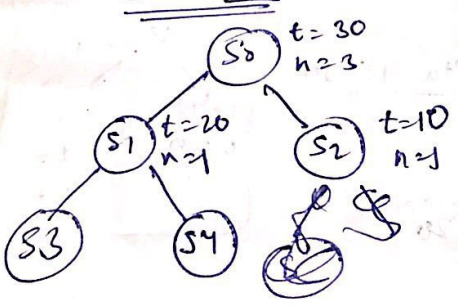
Chose b/w S1, S2:

$UCB(S1) = 20 + 2\sqrt{\frac{\ln 2}{2}} = 11.48$

$UCB(S2) = 10 + 2\sqrt{\frac{\ln 1}{1}} = 11.67$

4 → 4th

Choose S2:

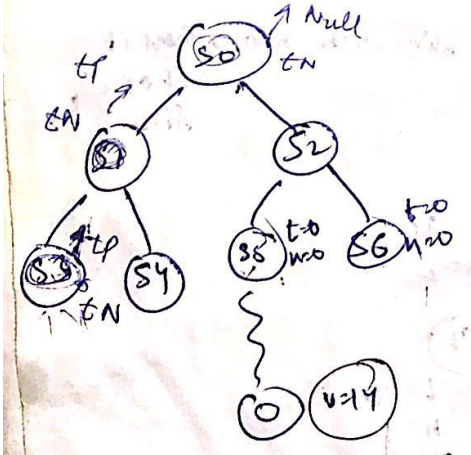


S2 is a leaf node but it is already explored, hence we look for all actions that can take place, let actions be a5, a6 then

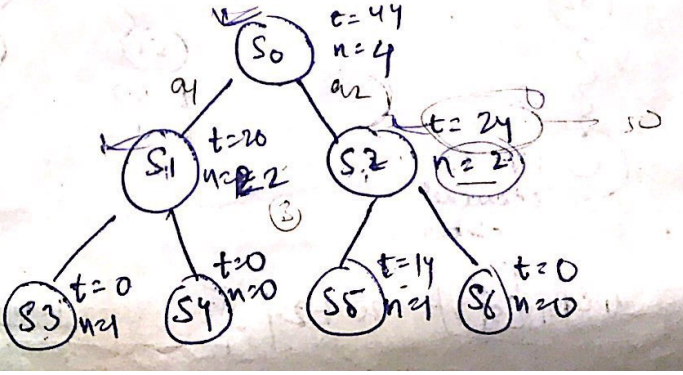
Chose any of S5 and S6.

Let us chose S5, it is unvisited leaf node, hence ROLLOUT (S5):

Let terminal value = V=14



After Backpropogation:



So, after 4 iterations we conclude that a2 is the best action as it has highest value of