1. General Idea of Greedy Algorithm

Greedy Algorithm may be the most heuristic algorithm that targets at online decision problems. It follows the two steps for each action:

1) Estimate a model from historical data

2) Select the optimal action based on the estimation

The solution is greedy if each action is chosen solely to maximize immediate reward.

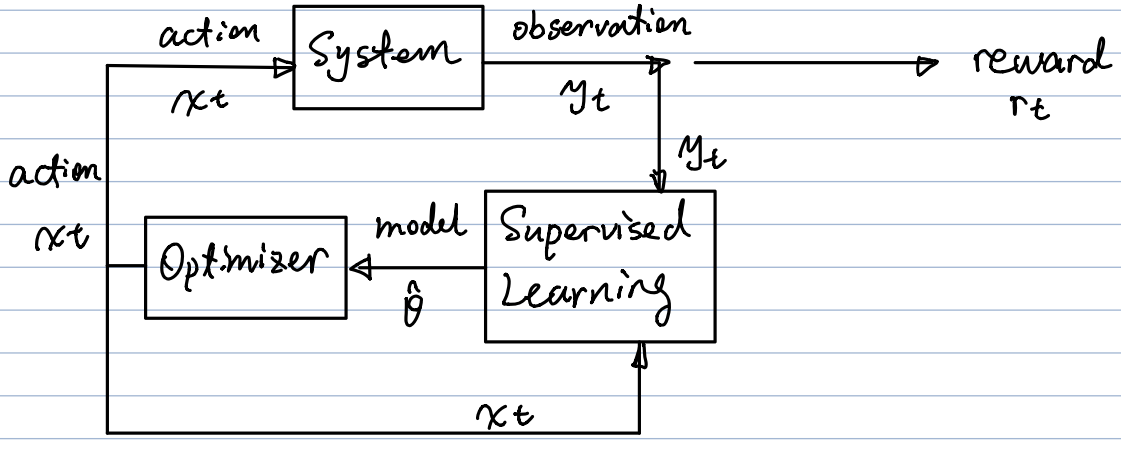
Mathematically, we use the following Language to describle a Greedy Algorithm

At time t, the historic data $\mathcal{H}_{t-1} = \{(x_i, y_i)\}_{i=1}^{t-1}$

where $\begin{cases} x_i \text{ is the action taken at time } i \\ y_i \text{ is the outcome} \end{cases}$

Reward at time t is $r_t = r(y_t)$, which is a function of y_t.

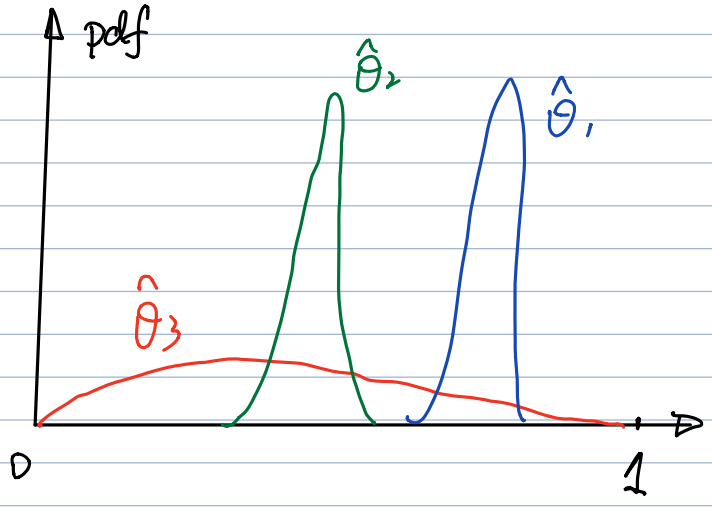
The algorithm uses $\mathcal{H}_{t-1}$ to estimate the system parameter $\hat{\theta}$ and selects the optimal action maximize r_t.



Example 1.

For a three-arm bandit machine, each time, one out of three arms can be pulled. Each arm has the probability of rewarding with probability θ_i for $i = 1, 2, 3$.

Initially, $\theta = \{\theta_i\}_{i=1}^{3}$ is not known to the player. The player played for a while and acquired historic data H_{t-1}, and the posterior estimation of $\theta_1, \theta_2, \theta_3$ are plotted as below.



It's obvious to see that

$\hat{\theta}_1 > \hat{\theta}_2$ with very small error prob.

$\hat{\theta}_1 > \theta_3$, $\hat{\theta}_2 > \hat{\theta}_3$, with relative small error prob.

So it's most likely that the first arm is going to be pulled ad infinitum. (forever). However, if the player is willing to do some exploration, it's possible that arm 3 can be the best, even it's not clear at present. But the greedy algorithm won't take the risk, and only action 1 is taken.

2. Greedy Algorithm with Dithering.

Dithering is a common approach that force to explore the system by perturbating the action.

One straightforward variation is ε-greedy exploration.

$\Rightarrow$ { apply greedy algorithm with prob $1-\varepsilon$
{ uniformly select an action with prob ε

It can be observed that ε-greedy does push the decision algorithm to do some exploration, while it's also true that, if we are almost sure one action is sub-optimal there is no reason to waste time on it.

$\Rightarrow$ Idea: We should allocate resources on actions that may be the best but are not appeared so at present due to the lack of data.

$\Rightarrow$ How to allocate resources wisely becomes the key problem to solve.