10: Reinforcement Learning

★ Slide - 48

- For infinite horizon, just like $V_\pi(s)$, we get a system of equations for $Q^*(s,a)$. The equations are not exactly linear because of the max() function term.

★ Slide - 55

- Now, we could calculate the states / measure the state of soil fairly accurately maybe by using moisture sensors or other parameters. We could also determine the set of actions we can take. But, in real ~~life~~ life, the transition model & ~~cal~~ forming the reward function isn't so easy or you could say it isn't so vanilla.

★ Slide - ~~106~~ 111

- **Exploration:-** means we are trying to understand the system or process by trying to do things & random & analyse the results of these actions.

- **Exploitation:-** Trying to do the best thing. (Not necessarily to plant ~~the~~ in the soil) In general to do an action which yeilds the highest reward

- One option to choose the tradeoff is use an ε-greedy strategy. This ε-greedy is basically a Bernoulli random variable.

$$X = \begin{cases} 1-\varepsilon, & \text{exploit}, \; P_X(\;) = 1-\varepsilon \\ \varepsilon, & \text{explore}, \; P_X(\;) = \varepsilon \end{cases}$$

Slide - 133

- One way in which we can learn Q is by estimating the transition model T & R (reward function)
- For initial estimate of $T(s, a, s')$ we assume all states in S to be equally likely. Hence, probability of getting to a random S is $\frac{1}{|S|}$.

Hence, $\hat{T}(s, a, s') = \frac{1}{|S|}$

The '^' indicates initial value.

- Since, initially anything hasn't happened yet ie. no action has been taken, $\hat{R}(s, a) = 0$

- The select_action() functions chooses an action based on what we want to do. If we choose an ε-greedy strategy, we would choose to exploit $100 \cdot (1-\varepsilon)$ times out of 100 times & explore $100 \cdot \varepsilon$ of out 100 times. If we choose to exploit, our function would depend on Q ie. we would choose the best action based on Q. If we choose to explore, our function won't depend on Q & would choose any action randomly uniformly.

- The execute() function executes the action & the nature returns the output state & reward.