Journal

Ankur Mishra

09/4/2017 - 9/15/2017

Contents

1 RL Notes

1.1 Markov Processes

- Where the environmnent fully observable
- Almost all RL problems can be characterized as MDPs

1.1.1 Markov Property

- $\mathrm{P}[\mathrm{S}_{(t+1)} \mid \mathrm{S}_{t]} = \mathrm{P}[\mathrm{S}_{(t+1)} | \mathrm{S}_1, \ldots, \mathrm{S}_{t]}$
- Future is irrelavent of past, only related to present
- Given $\mathrm{S}_{(t)}$, you don't need anything else to find to find next state s'
- Transition Matrix P defines probabilites for all successive states S'

1.1.2 Markov Chains

M = {S, T}

- Episodes are random sequences that are sampled.
- S = State Space
- T = Transition Probability or the probabililty of entering the next state
-

1.2 Markov Reward Process

M = {S, T, R}

- MRP is a tuple of (S is a finite set of states, P is a state of the transitionprobability matrix, Reward Function R, dicount factor γ)
- $R = E[R_{(t+1)} \mid S_t = s]$

$R_{(t+1)}$ is the amount of reward we get from state s

- We care about the cumulative reward

1.2.1 Return (goal)

Definition: total discounted reward from time-step t

- $G_t = R_{(t+1)} + \gamma * (R_{(t+1)}) + \ldots$
- Made finite by the γ
- γ is going to have to be [0,1]; 0 discounted factor means you only care about present Reward, 1 factor means you care about all of them
- Discount factor is used because we don't have a perfect model, avoids infinite returns, and animals show a preference for immediate reward

1.2.2 Bellman Equation

The Bellman Equation determines value of a state. It is comprised of immediate reward ($R_{(t+1)}$) and value of next state (γ*v($S_{(t+1)}$))

- Equation: $v(s) = E[G_t \mid S_t = s] = E[R_{(t+1)} + \gamma * v(S_{(t+1)}) \mid S_t = s]$

It is a linear quation and can be solved.

1.3 Markov Decison Process

M = {S, A, T, R}

- MDP is the same as MRP except with the addition of A (the action space)

1.3.1 Policy

$\pi(\mathrm{a}|\mathrm{s}) = \mathrm{P}[\mathrm{A_t} = \mathrm{a} \mid \mathrm{S_t} = \mathrm{s}]$

- A policy defines the behavior of an agent. It picks the actions that get the most reward.
-