

Q-Learning

CSE-5364 – HOMEWORK 3 - FALL 2017

ROHITSHANKAR MISHRA 1001522815

Learning to balance an inverted pendulum

Q-learning is a model-free reinforcement technique. Specifically, Q-learning can be used to find an optimal action-selection policy for any given (finite) Markov Decision Process. It works by learning an action value function that ultimately gives the expected utility of taking a given action in each state and following the optimal policy thereafter. -[wikipedia](#)

The following equation has been used for solving the problem -

$$Q_{t+1}^*(s, a) = (1-\alpha) \times Q_t^*(s, a) + \alpha \times (r(s) + \gamma \times \max_b Q_t^*(s', b))$$

where

Q_{t+1}^* : optimal Q value of current state

s : current state

a : current action

α : learning rate

Q_t : optimal Q value of previous state

$r(s)$: Reward associated with current state

γ : Discount factor

s' : next state

b : all possible actions

The code is attached alongside.

Below are the results of the learning curve obtained after implementing the above equation

