October 10, 2019

Department of Computing

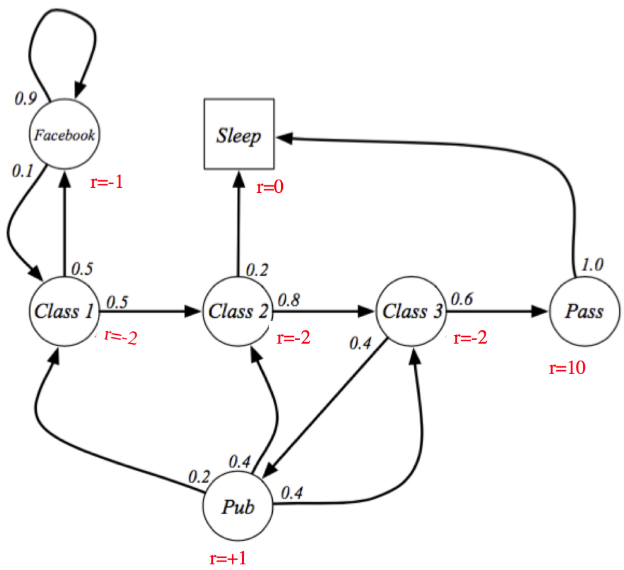
CO424H/BBB – Dr Aldo Faisal

Lab Assignment 0

Lab assignments are exercises that guide you threw your "hand's" on learning experience. They deepen your learning experience, complement the lectures and may convey content or insights not covered by the lectures, They are not assessed, you do not have to return any results. You can do the lab assignments in your own time, or better, interactively working with class mates and the GTAs to get quick feedback turn around and boost your learning experience. A solution will become available after the class had the opportunity to work on this assignment at some point before the exam.

Lab assignment: Understanding MRPs

Consider the Markov Reward Process drawn out below (as seen in the lectures)



Recall that a Markov Reward Process (MRP) is a Markov chain which emits rewards. Specifically a Markov Reward Process is a tuple $(\mathcal{S}, \mathcal{P}, \mathcal{R}, \gamma)$, with $\mathcal{S}$ is a set of states, $\mathcal{P}_{ss'}$ is a state transition probability matrix, $\mathcal{R}_s = \mathbb{E}\left[r_{t+1}|S_t = s\right]$ is an expected immediate reward that we collect upon departing state s, this reward collection occurs at time step $t+1$ $\gamma \in [0, 1]$ is a discount factor. Assume $\gamma = \frac{1}{2}$.

In a programming language of your choice you are to code up this MRP for numerical simulation:

1. Choose a simple code representation of the $\mathcal{S}$ state space and implement it.
2. Write a function that returns the state transition probability $\mathcal{P}_{ss'}$.
3. Write a function the gives you $\mathcal{R}_s$ (note that in our MRP the reward is deterministic, so the expectation is really the immediate reward).
4. Sample a trace (the sequence of state, reward, state, reward, . . . , terminal state, 0) from this MRP. Hint: You will need to write functions that e.g. implement the probabilistic state transition dynamics. Simulate a run of the MRP always using $Class1$ as the only initial state.

5. Write a function that computes the return of a specific trace (i.e. from its initial state till it reaches the terminal state.
6. Write a function that computes (by averaging over the returns of many sampled traces) the value of each state in our MRP. This is a way of computing the state value function (Why?).
7. Report the computed state values for all states. Compare your results to the solution in the lecture notes.

Bonus How many repeated samples do you need to get a good estimate of the state value. Plot for the state $Class1$ the empirical average of the returns of individual traces as a function of the number of runs.