Reinforcement Learning Tutorial 9, Week 10

—

Revision: MDPs & Semi-Gradient SARSA

Pavlos Andreadis

March 2024

Overview: The following tutorial questions relate to material taught in weeks 1 to 6 of the 2023-24 Reinforcement Learning course. They aim at encouraging engagement with the course material and facilitating a deeper understanding.

We continue this week with another look at past exam questions. The first one involves modelling a Markov Decision Process (MDP) for a variation of *the River Crossing Riddle* [Wikipedia, accessed 2021]. Note how the problem explicitly defines your state space, and does not allow for defining any further states. This constraints your choices when asked to define absorbing states in the first question. Note also that transitions here are deterministic.

In the second question, taken from the same exam, you are first asked to define a set of features for the state space. Though you are not allowed to use a one-hot encoding, there are still a few alternative formulations to explore. Some of them are preferable to others, so it is worth thinking about how your feature space affects what value function approximations can be learned.

Problem 1 - Revision: MDP Modelling

the River Crossing Riddle **[Adapted from RL exams in 2018-19]**

A farmer went to a market and purchased a wolf, a goose, and a bag of beans. On his way home, the farmer came to the bank of a river and rented a boat. But crossing the river by boat, the farmer could carry only himself and a single one of his purchases: the wolf, the goose, or the bag of beans. If left unattended together, the wolf would eat the goose, or the goose would eat the beans. The farmer could use the boat to transfer his purchases from either bank of the river to the other. The farmer could also move across the river with the boat empty.

Consider the control problem for moving all of the farmer's purchases across the river, where the current state is defined *only* by the following binary state variables:

- which side of the river the farmer is at;
- which side of the river the wolf is at;
- which side of the river the goose is at; and
- which side of the river the bag of beans is at.

1. Define a state space for the control problem. Which is the initial state? Which states need to be defined as absorbing states?
2. Given the above information, formulate a finite Markov Decision Process (MDP) with discounting for the problem of controlling the transfer of the farmer's purchases across the river. For the transition function, give only transitions from the initial state.

Problem 2 - Revision: Semi-Gradient SARSA

[Adapted from RL exams in 2018-19]

You observe an agent going through the following sequence of states, actions (i.e. `move-up`, `move-left`), and reward signals.

$$\{\text{“green”, “loud”}\}, \texttt{move-up}, -1$$
$$\{\text{“blue”, “loud”}\}, \texttt{move-left}, 0$$
$$\{\text{“green”, “silent”}\}, \texttt{move-up}$$

1. Define a set of features to represent the state (do not enumerate the states; in other words: do not define one feature per state, and do not use a state index as a feature). For simplicity, assume that states are only ever described as either “green” or “blue”, and as either “loud” or “silent”, and that the only available actions are `move-up` and `move-left`.
2. Then, using linear function approximation for the action-value function in this problem compute the first two semi-gradient one-step Sarsa updates along the episode and given the actions taken, using initial weights of 0, a discount factor of 1, and a learning rate of 0.1 (show all the steps).

References

Wikipedia. River crossing puzzle. https://en.wikipedia.org/wiki/River_crossing_puzzle, accessed 2021.