DS 5004: Applied Reinforcement Learning

Reinforcement Learning Journal

Bruce McGregor (BM3PK)

Journaling Instructions

The purpose of this exercise is to track your learning and growth through this course.
You will add an entry for each Module (ideally on a weekly basis).
You will submit your journal over the course of the term and it will grow with your new entries.

Each entry should answer these questions:

- What were some things you learned in the module?
- What do you think were the most important concepts?
- What was challenging for you? How can you learn it better?
- Which parts did you enjoy?

Constructive feedback is also fine (e.g., Exercise X, Part Y wasn't clear) but negative criticism should be avoided.

Each entry should:

- Use full sentences to show good writing and clear thought process
- Be limited to one or two paragraphs (it should be fairly brief)

File format: Use a Word doc or similar. Recycle the same file throughout the course.

Module 1

Reinforcement Learning Fundamentals

What were some things you learned in the module?

One of things I learned is that RL ability to achieve the effects of planning and lookahead without using a model of the opponent in the tic tac toe example. In addition, it does not need to do an explicit search over possible sequences of future states and actions. I think that is incredible, and hard to grasp. So, I am looking forward to exploring further how RL achieves such things.

What do you think were the most important concepts?

I think understanding the four main elements of TL system, that is the policy, reward signal, a value function, and model (optional). Understanding the difference between reward and value is a ky concept to grasp. It is also important to understand how RL allows for individual states to be evaluated and adjust policy. This contrasts with evolutionary methods, where the individual states are not evaluated, only the final result.

What was challenging for you? How can you learn it better?

Retention is a challenge. I agree that learning is recursive. I find that during my first reading through the text (even when concentrating hard), I struggle to recall it later. I find that re-reading after the lecture helps and highlighting and underlining key concepts helps. I iteratively must review the previous material to ensure retention. Also, I still use hard copy textbooks that allow me to flip pages back and forth to really absorb the material. I find it hard to replicate this on Kindle.

Which parts did you enjoy?

I enjoyed the cart pole lab. It was fun and challenging at the same time. I also enjoyed reading about this history of RL and what other domain it pulls from, such as control theory and statistics.

Module 2

K Armed Bandits and MDP

What were some things you learned in the module?

I learned about the K-armed bandit problem, and how it assumes a single state, which simplifies the problem. I learned how the incremental implementation is applied in RL to save computation space, by only requiring us to store the value of Q_n and n, and only the small computation for each new reward. I also learned about the Markov process, and how it provides the mathematical foundation for RL. I learned that the Bellman equation used recursion and is based around dynamic programming.

What do you think were the most important concepts?

Of the important concepts is balancing exploitation with exploration, and how this translatesinto application of greedy and e-greedy approaches. Another important concept is the difference between stationary and non-stationary problems and using different approaches to setting the step-size parameters. The two biggest take away for me was state value function vs the action value function, where the latter is used to adjust the agent's policy, and the former is a weighted summation of all the possible action in a state.

Some other important concepts are how the gains function works and the discounting factor. I found it challenging to understand how an infinite sum can provide a finite value, so I had to dig into the math behind the geometric series to really validate how it works. I learned this better by exploring beyond the text.

What was challenging for you? How can you learn it better?

I struggled a bit to understand the concepts of stochastic approximation theory and the two conditions that must be satisfied to assure convergence. In particular, understanding why both conditions are met for the sample average case, but not for constant step size parameter. When I struggle with concepts in the text, Chat GPT has been a big help.

Which parts did you enjoy?

I enjoyed the hands-on examples in the notebook, which put the concepts into action, and helped reinforce the readings. I enjoyed the lecture and explanation on the Bellman equation by breaking it down and using the backup diagram to understand how it works.

Module 3

Solving MDPs

I began learning about ways to solve MDPs. The first was dynamic programming, which is a good foundation, but not used in practice. It requires a perfect model of the environment, which is not practical in many real-world problems. As a side note, I had just completed the Algorithm class, where I first studied dynamic programming, so it was nice to see how it relates in another class. I also learned the key difference in Monte Carlo is that it samples (often from a simulator) and does not require a model of the environment

Some of the most important concepts I feel are the value iteration and policy iteration algorithms. I found these require a lot of study and concentration to fully comprehend. The lab offers a useful application to see how value iteration worked in practice. This was useful and helped me understand the algorithm over the pseudocode example in the textbook.

I enjoyed reading about the Blackjack Examples in the Chapter on Monte Carlo. We will be discussing this in class on Thursday, which should provide additional reinforcement.