

WiDS – How Machines Learn: A Journey through Reinforcement Learning

This project introduced me to the most interesting type of machine learning, the Reinforcement Learning. It happened over 5 weeks, during which we went from python and its libraries to Dynamic Programming in RL.

Week 0:

It was for the basics of machine learning in python by learning the libraries NumPy, Pandas and Matplotlib. It not only helped me during the duration of the whole project, but would also help me in my future coding endeavors.

It also was for learning about Git, which I was learning about for the first time but I found it the most effective in this whole project duration.

Week 1:

In this week, we took off with our first RL problem, the multi-bandit problem. We started with the book “Reinforcement Learning: An introduction” by Sutton and Barto. We learnt the greedy and ϵ -greedy algorithms and also the Upper Confidence Bound and Optimal initial values algorithms. In this week, we also learnt the essence of RL by learning about states, actions, values and rewards.

Week 2:

In this week, we learnt about Markov Decision Process and modelled some simple environments, like Frozen Lake, Slippery Walk etc., using it. We’ve also learnt about goals and rewards, policies and value functions and also optimal policy and optimal value function. We’ve also learnt that tasks are divided into 2 types based on termination, Episodic and Continuing tasks. Also, for this week we’ve specifically used “Grokking Deep Learning” book too, which provided clear insights into the design of the model environments.

Week 3:

In this week, we’ve continued the next chapter in Sutton and Barto, Dynamic Programming. This chapter gave deeper insight into policies and value functions. We’ve learnt about Policy Evaluation, Policy Improvement and Policy Iteration and also Value Iteration. And this week’s assignments were on Dynamic Programming and Markov Decision Process. Given two assignments were Jack’s Car Rental and Gambler’s Problem(from Sutton and

Barto), of which I was only able to solve Gambler's Problem using both MDP and Value Iteration.

Final Week:

This was for the final project, which was an option between Rubik's Cube Solver using RL and 15 Puzzle Solver using RL. I chose the latter as my final project. I've used MDP and Value Iteration to complete my project. I've first defined the states of the problem and then used value iteration for finding the optimal policy.