

EEC-289A Reinforcement Learning

Homework #2

Jonathan Dorsey

<https://github.com/JonnyD1117/EEC-289A-RL/tree/main/HW>

April 13, 2021

Dynamic Programming:

The fundamental mechanism in evaluating policies for this homework is dynamic programming, where there is a one step look ahead bootstrapping happening at each timestep for each feasible action and future state given a previous state. This is predicated on knowledge of the dynamics of the system. For Homework #2, the dynamics of the coin flip are given and so we can apply Dynamic Programming to evaluate the policies being employed.

Since we know the dynamics of the system, we can compute the value function of the problem by iterating over every state and every action possible from that state and averaging the values and rewards associated with transitioning to each state and previous value associated with that state. This process is repeated until a user-specified accuracy is reached.

For the MDP given in the problem statement, the states of the system are defined to be...

$$s \in \{(0, 10 - 18), 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

And where the state $(0, 10 - 18)$ is the terminal state of the MDP, such that if the agent has \$0 the agent lost, and if the agent has any value between \$10-18 the agent has won since it has accumulated greater than \$10. In either case, the game is over and the probability of transitioning out the terminal

state becomes zero.

Additionally, the actions of the MDP is the amount bet by the agent, and the reward for the MDP is the amount of money won or lost on each coin flip.

Code Running Time

The cumulative time to run all of these different policy evaluation problems took approximately **35ms**.

Problem #1: Agressive Betting

For this problem, the goal is to evaluate the policy where the bet placed by the agent is always the maximum possible bet possible given the agents initial capital.

By applying DP (given the coin flip probabilities), the resulting value function (rounded to 2 decimals of precision) for $\theta = .0001$ is ...

$$V(s) = \{0, \quad 9.49, \quad 9.66, \quad 6.72, \quad 8.96, \quad 4.0, 4.80, \quad 5.60, \quad 6.40, \quad 7.20\}$$

Problem #1: Conservative Betting

For this problem, the goal is to evaluate the policy where the bet placed by the agent is always \$1 possible bet possible given the agents initial capital.

By applying DP (given the coin flip probabilities), the resulting value function (rounded to 2 decimals of precision) for $\theta = .0001$ is ...

$$V(s) = \{0, \quad 7.89, \quad 7.87, \quad 6.98, \quad 5.99, \quad 4.99, 4.0, \quad 3.0, \quad 2.0, \quad 1.0\}$$

Problem #1: Uniform Random Betting

For this problem, the goal is to evaluate the policy where the bet placed by the agent is uniformly random from the range of bets possible given the agents initial capital.

By applying DP (given the coin flip probabilities), the resulting value function (rounded to 2 decimals of precision) for $\theta = .0001$ is ...

$$V(s) = \{0, \quad 8.65, \quad 8.72, \quad 8.15, \quad 7.61, \quad 6.67, 5.62, \quad 5.05, \quad 4.75, \quad 4.61\}$$

Problem #2: Aggressive Betting

Same policy as in **Problem #1 Aggressive Betting**, with the exception that the dynamics of the coin flip are reversed, such that the probability of head is .1 and the probability of tail is .9. Under this policy and these dynamics the value function of the problem is...

$$V(s) = \{0, \quad -0.9984, \quad -1.984, \quad -2.88, \quad -3.84, \quad -4.0, \quad -4.80, \quad -5.6, \quad -6.4, \quad -7.2\}$$

Problem #2: Conservative Betting

Same policy as in **Problem #1 Conservative Betting**, with the exception that the dynamics of the coin flip are reversed, such that the probability of head is .1 and the probability of tail is .9. Under this policy and these dynamics the value function of the problem is...

$$V(s) = \{0, \quad -1, \quad -2, \quad -3, \quad -4, \quad -4.999, \quad -5.998, \quad -6.986, \quad -6.9863, \quad -7.876\}$$

Problem #2: Uniform Random Betting

Same policy as in **Problem #1 Uniform Random Betting**, with the exception that the dynamics of the coin flip are reversed, such that the probability of head is .1 and the probability of tail is .9. Under this policy and these dynamics the value function of the problem is...

$$V(s) = \{0, \quad -.999, \quad -1.9928, \quad -2.96, \quad -3.90, \quad -4.67, \quad -5.31, \quad -5.95, \quad -6.59, \quad -7.24\}$$