

Homework 4

Monte Carlo exploration starts:

We initially initialize the various necessary values like the Rewards, Number of episodes, etc. Next, we use MATLAB's `Randi` function to help us distribute the cards to both the dealer and the player. We also create a policy matrix for both the player and dealer. This is when they would stick or hit. I then looped forever and hit or stick according to our policy that is specified. If while for example, the player is playing, and he goes bust (sum is greater than 22) we automatically terminate the loop and here the dealer would win. A draw could be achieved when both end up with a similar sum. I have created 'break' points whenever the player or dealer goes bust. Also according to their policy, they are given cards to 'hit', now in case they get an ace and their sum is less than ten then we can use this ace and if not I have made sure that their cards being used is a '1' instead of an '11'.

After we play for the player and he is happy with his sum and hasn't gone bust we go to the dealer's game. In the event that the dealer is satisfied with his sum, it would go into the part of the code that checks who wins or loses. Else, we also deal with the dealer with cards and proceed according to the policy.

Now, once we are totally satisfied with both their policies, we proceed to evaluate them and check who won the game based on who would be closer to the sum 21 and yes, if they both have a sum of 21, then it's a draw. These are coded by the help of an 'if' and 'else if' statements. The next part is we assign the rewards. All this time we have been appending a state matrix that holds player sum, dealer card, actions.

Now we proceed to find the G values, and for every pair of S_t , A_t not appearing in this state matrix we append G to the return's an average this out find its max and then update our policy

Possible problems :

- I haven't used a right policy evaluation method, and hence the policy is not getting evaluated
- I haven't been able to update the Q matrix properly
- I am doing the average wrong or the max

My code is able to play the game and determine who would win based off the initial policy and is able to also display the required values as per the policy evaluation, but I am stuck or doing something wrong in my evaluation and updating