Final Project Status Update: Hit or Stand?

Naixiang Gao ngao4@stanford.edu
Jack Ren zr11@stanford.edu
Jiaqi Shao jiaqis7@stanford.edu
AA228/CS238, Stanford University

1. Re-introduce the problem

In our project, we attempt to use Q-learning to train an agent to play the game of Blackjack. The algorithm will be based on the current value of our cards and the value of the agent's revealed card to decide the next step, which is to hit or stand. In our project, we will simplify the environment to consist of a simulated blackjack table with only two players (the agent plays against the dealer), and we simulate the role of an agent. For the dealer, we suppose it follows the Stand on Hard 17 rules.

2. Progress report

So far, our code has completed the function for data generation. The generated data has four columns: state, action, reward, and next state. The details of the game simulation are shown below.

In our game, we first set Aces to have a value of 1 or 11 and we set it as 11 first and reset it to 1 if the hand becomes bigger than 21. Here's how we simulate the whole environment: Since the range of our hand value's sum is from 4 to 21, there are 18 states. Also, we take into account whether we have an ace in our hand, which is 2 states (true/false), and we also put the dealer's revealed card into state space, which is another 10 states (2-11). If we combine those three categories into the state space, there would be 18*2*10 = 360 states. In the end, we add the terminal of the game as another state, and there would be 361 states in total.

There are only two actions in the action space, which are hit and stand. There would be an updated workspace if the action is hit, but if the action is stand, it would turn into the dealer's action space which we could not control.

We set 6 different kinds of rewards in the model:

1. Win: +1 (general case) / +2 (when the dealer's visible card is an Ace).
2. Loss: -1 (general case) / -2 (when the dealer's visible card is not an Ace).
3. Tie: 0 (general case) / +0.5 (when the dealer's visible card is an Ace).

3. Timeline

Date	Task
11/12	Generate Data
11/13	Finish Status Report
11/25	Construct Learning Structure
11/27	Test Algorithm and Prepare for Presentation
12/06	Record Presentation

Table 1: CS238 Final Project Timeline