Wet2 – 046203

Question1

* 1. We will define the game as a Markov decision process. We will define each state as the sum of the cards of the player (X) and the value of the first dealer card (Y) (When it’s the dealer turn there are no more actions to be taken so those states are not important. There are 10 states for the dealer card value, 17 Player sum values (that are lower to 21) and 1 terminal state. The number of states:   
     We don’t care about the different terminal states because the player can’t take any action in those states. Furthermore, we will model the reward function such that we receive a stochastic reward in each state based on the action taken. in the terminal state we don’t receive a reward. This is fine because the value function at different terminal states is trivial.

The action space (A) is to take a card (hit=1 or to not take a card (stick =0). The reward is stochastic, we will calculate the expectancy of it grammatically for the value iteration algorithm.

* 1. Like we mentioned in the previous subsection, we will not plot the terminal states, since the value function there is trivial (just the reward in each of those states).

Chart, surface chart

Description automatically generated

Figure: The value function for the different states of the game

We can see that the maximum value is achieved when the player has cards that sums up to 20. In those states (as we will see clearly in the next subsection), the player will stick and will have a very large change of winning. Another big value group of states are the ones where the player has cards that sums up to 10. In those states the player will hit and will have a large chance of getting a 10 (we have four cards with a value of 10) and then will stick.

* 1. We can see some interesting things from the following figure, such as:
* For a hand greater or equal to 17, we will always stick.
* For a hand lower than 12 we will always hit.
* For a Dealer showing of around 7, we prefer to hit more than other states. This is since the dealer has a higher chance of getting close to 21 in this case.

Chart

Description automatically generated

Question2

2.1) The states are the jobs left and the action is which job to take. Thus, the number of states is the number of subsets in the set of jobs, and the number of actions is the number of total jobs:

We will order the states as binary code, where the state “00000” corresponds to the state which no job is left and “00011” corresponds to the state which only the fourth and the fifth jobs left.

2.2) CODE: Solved using iterative solution, rewards = -costs

2.3) The policy and its value function

Chart, scatter chart

Description automatically generated  
Figure: The “cost only” policy and its value function

We can see for example that the state with the lowest value function is – the state where all jobs are not finished.