# Reinforcement Learning HW4

Hadi Askari

6th May 2023

1 Implementation Details

I tried to implement the Black Jack game and find the optimal policy and value functions. The code takes approximately 1 minute to run and plots 2 scatter plots and 2 contour plots. One of them is for the unusable ace and the other one is for the usable ace. In order to implement the hit or stick on the soft 17 we need to comment out (or keep) lines 295 to 299 and re-run the code. We initialize states, actions, q values and returns at the start. The function run_iteration implements the game and returns the policy, q table and returns which are then used for plotting. The function is run for 500000 iterations as indicated in the book. Lines 324 onwards implement the MC ES algorithm. There was some error in my implementation that I failed to debug which is leading to sub-optimal convergence.
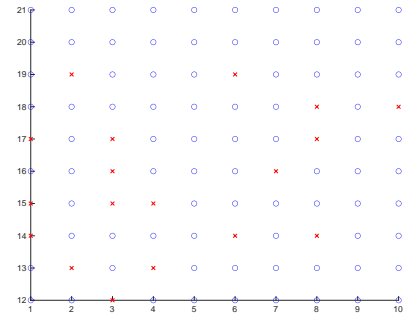
2 Figures

2.1 Hit on Soft 17

The following are the figures generated when we Hit on Soft 17 for both the case of a usable ace or an unusable ace.
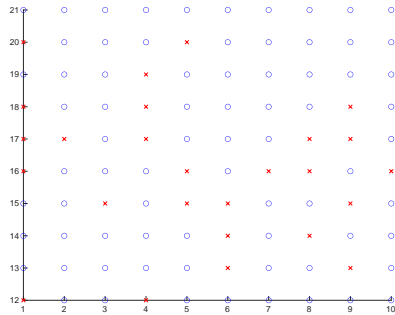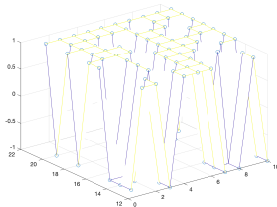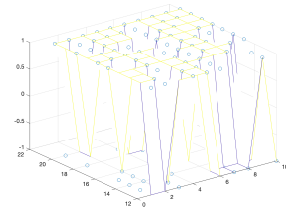

(a) Scatter Plot for Usable Aces

(b) Scatter Plot for Unsable Aces
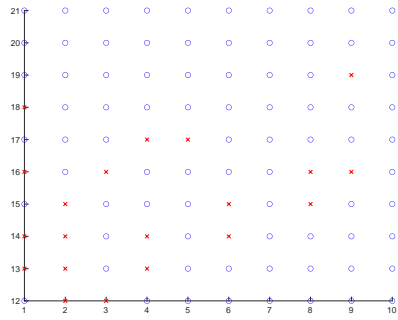
Figure 1: Scatter Plots for Hitting on Soft 17

(a) Contour Plot for Usable Aces

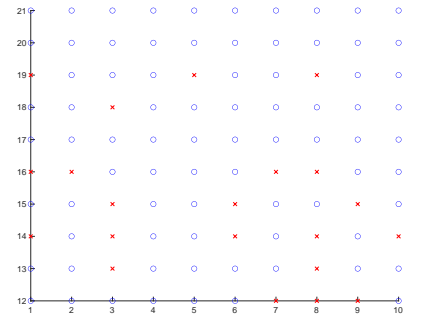(b) Contour Plot for Unsable Aces

Figure 2: Contour Plots for Hitting on Soft 17
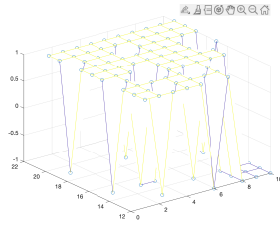
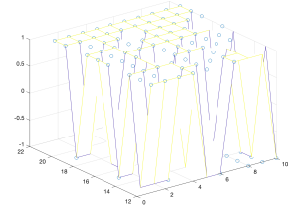## 2.2 Stick on Soft 17

(a) Scatter Plot for Usable Aces

(b) Scatter Plot for Unsable Aces

Figure 3: Scatter Plots for Hitting on Soft 17

(a) Contour Plot for Usable Aces

(b) Contour Plot for Unsable Aces

Figure 4: Contour Plots for Sticking on Soft 17