HOMEWORK 1

You could also do a group assignment with no more than 3 people in a group.
Due date: Oct. 18.

Q1. Assume that in the Blackjack game, there are m decks of cards, and n players (a dealer vs $n-1$ players). The rules of the game are explained in Example 5.1 of Sutton and Barto.

(1) Find the optimal policy for the Blackjack, when $m=\infty$ and $n=2$. You can use any of the methods learned so far in class, e.g. Monte Carlo, TD, or Q-Learning. If you use more than one method, do they reach the same optimal policy?

(2) Visualise the value functions and policy as done in Figures 5.1 and 5.2.

(3) Redo (1) for different combinations of (m,n), e.g. m=6, 3, 1, and n=3,4,6. What are differences?

Q2. Find the optimal policy for the Windy Gridworld example as in Example 6.5 of Sutton and Barto, using any of the methods learned so far.