We have written code for deep Q learning and double deep Q learning. The code includes deep Q model, Atari game environment wrapper, deep Q, and double deep Q learning agent. There are two types of deep Q models, one with 3 convolutional layers and one with 4 convolutional layers. The 3 layer model is used to solve Pong, and both 3 and 4 layers deep Q models are tried on Breakout and Alien.

The Atari game environment wrapper has four primary functions. First, the fire wrapper makes sure that for the game have fire action, the fire action will be taken at the beginning of each episode, otherwise, the game will not start, and nothing will be learned during the training. The second wrapper is to preprocess the state of the game environment. The states of the Atari game environments are RGB frames, the preprocess wrapper converts the RGB frames to grayscale and rescales the frame from 210x60 to 84x84. The preprocess wrapper also stacks 4 frames together as one state. The frame stack allows the input state to deep Q model contain information about the dynamic of objectives in the game. Third, the clipping wrapper clipping the reward to -1, 0, and 1. In the Breakout game, the score is different when breaking the block into different layers, this makes us hard to find out exactly how many blocks are break in the game. Apply the clipping wrapper to make every block break count as reword 1. The fourth wrapper is the life episode wrapper. For games like Breakout, there are 5 lives in each game, it is efficient to count each life loss as the end of the episode, but only reset the game until all five lives are loss.

The score of the game is evaluated by the average score of 100 games. For the game Pong, the maximum possible score is 21. The Q learning reaches a score of 18.51 and double Q learning reach 19.64. The result is close to the result presented by Deep mind team. The score on Alien was not ideal, the max score is 593, which is a lot less than 2907 from Deepmind team. The third game Breakout is trained on double deep Q learning, and reaches an average score of 89 with the 4 layer Network and 64.59 with the 3 layer Network. Despite the Breakout doesn’t reach Deepmind’s score (375), the score is still more than double the average human score. In some trials, the Breakout can reach a score of 280 and it learned the trick to dig a tunnel and boost the score.