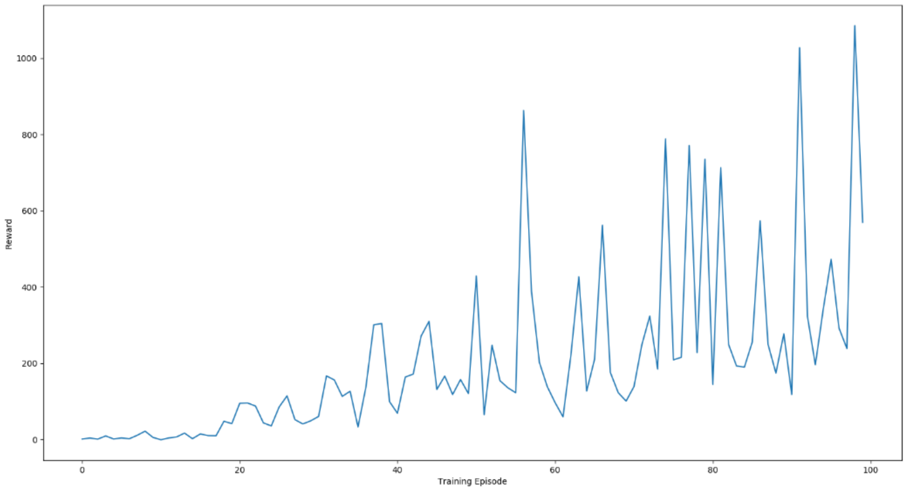
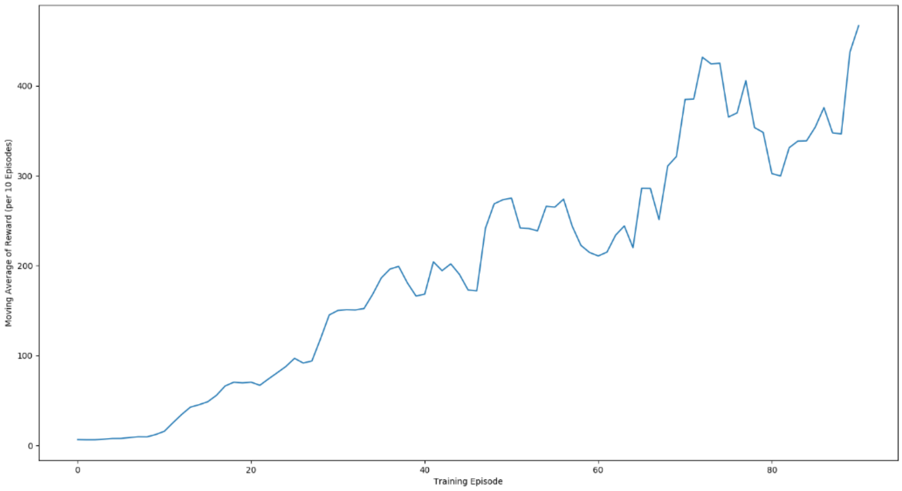


The picture above is my result about CartPole. The x-axis is training episode and the y-axis is episodic reward. I ran a total of 100 episodes. As can be seen from the figure, rewards generally show a clear upward trend.

Since I changed the calculation of reward in the code so that the agent can learn more efficiently, the reward in the figure will be higher than the default value, but I don't think this will affect the result of the agent's stable learning.



The above picture is more intuitive than the first picture. The x-axis is training episode and the y-axis is the moving average of rewards. From this picture, we can see that the reward has risen from about 10 points at the beginning to about 400 points at the 100th episode. If I continue to run the program to 500 episodes or even 1000 episodes, according to the trend, the rewards will be higher.

Based on the above, I think I solved CartPole.