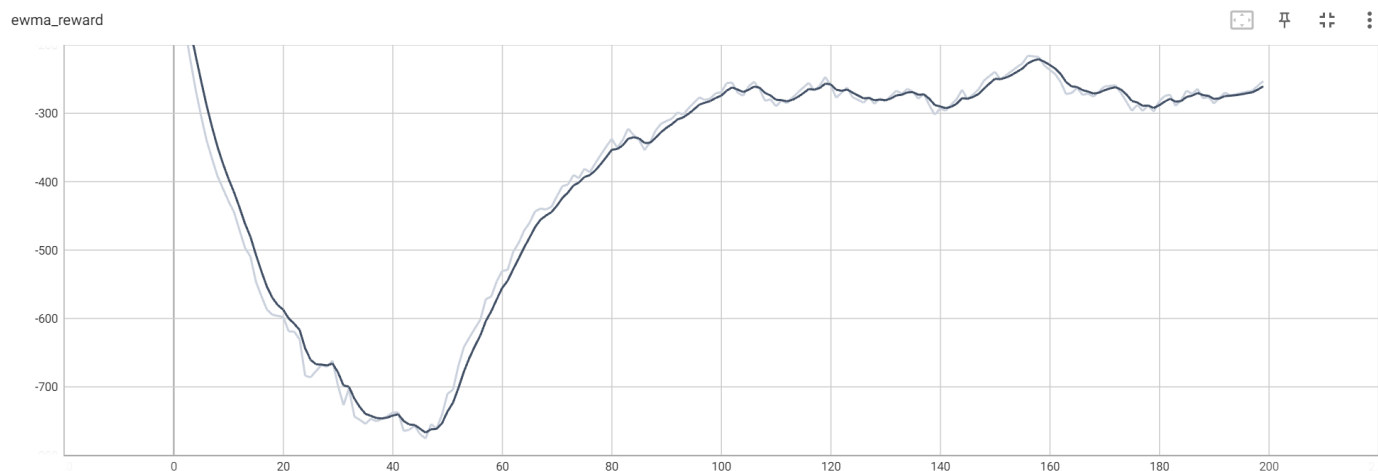
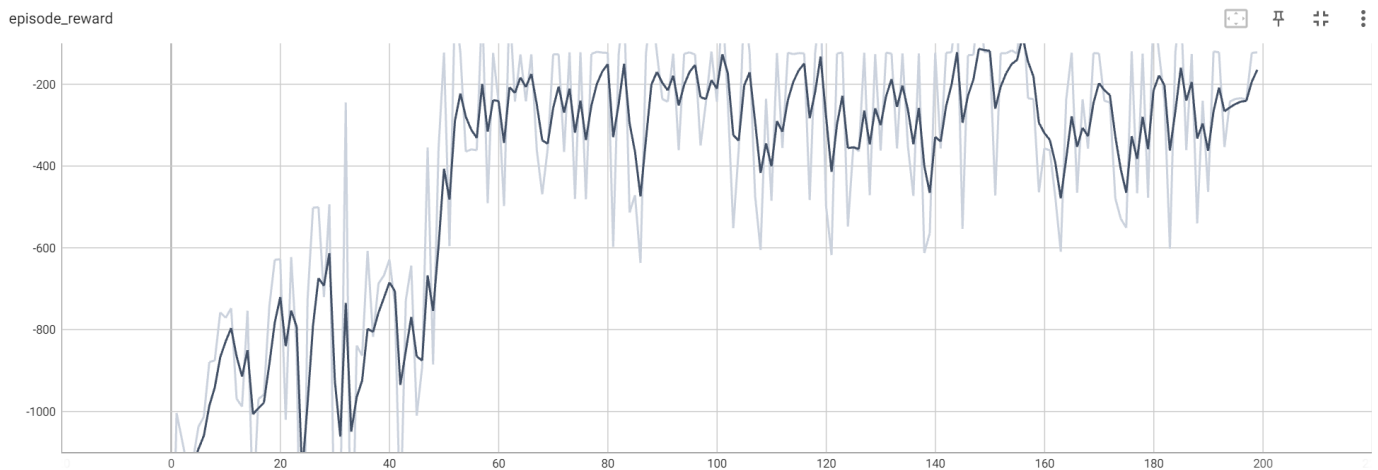


Reinforce Learning hw2

Problem3 (a.)

In this problem, I only train one model for this problem.
With the default parameter in the provided file.

- Hidden size => 128
- Noise scale => 0.1
- Tau => 0.005
- Gamma => 0.995
- Batch size => 128



As we can see, the model has significant improve after about 50 episodes whose policy can really finish our request.

Problem3 (b.)

Using the same code as Problem (3a.) but do not have a model that can really finish the request perfectly every time.

The parameter I used will be present in the following way:

Episodes_Gamma_Tau_Hidden size_Noise scale

First I tried different Hidden size for this problem as I think it was more complex problem; therefore, the more hidden size may make it easier to find the fit policy for it.



As we can see, the higher layer doesn't stand for well model on difficult problem.

I watched some of results render video, and I found that it may stuck on the air and didn't choose to land.

Thus, I thought this may be the problem that it didn't explore enough for this problem. This time, I chose to change noise scale value and hoped it will improve the exploration of the model.



✓ 1000_0.95_0.005_256_0.1



✓ 1000_0.95_0.005_256_0.2

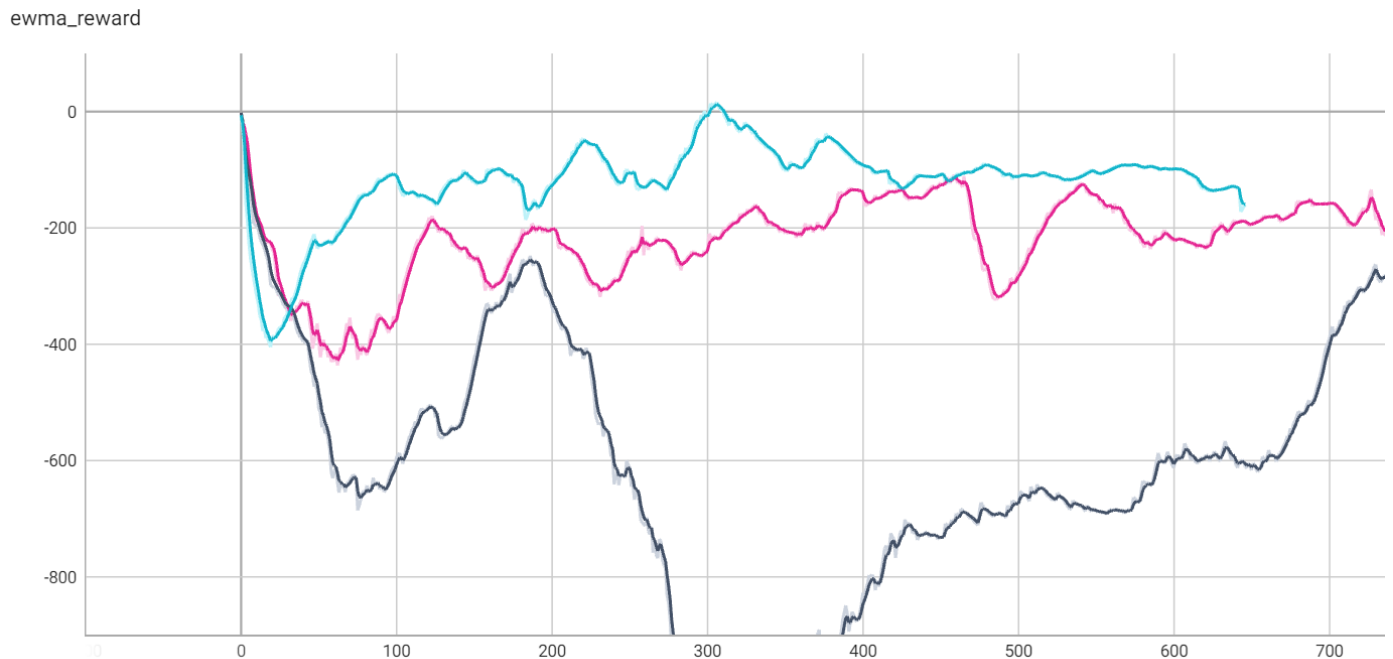


✓ 1000_0.95_0.005_256_0.25



This time, it got the better score on the middle of our train. Nevertheless, it just much earlier found the no-landing status so I think it did not really help me to find better model.

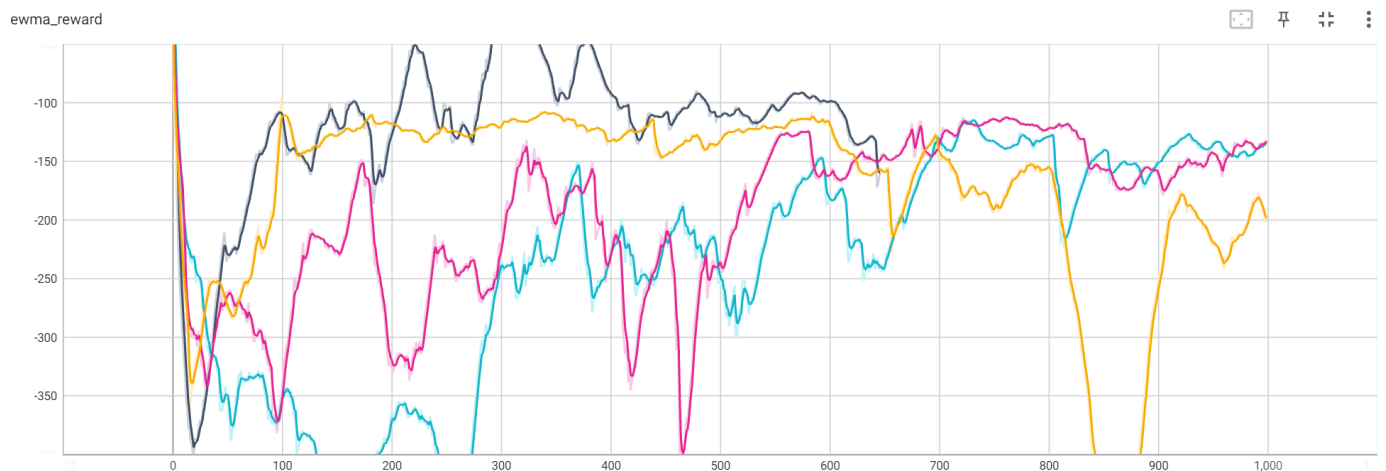
After that, I attempted to different γ to make our model be more concerned about the current action rewards.



- ✓ 1000_0.995_0.005_256_0.1 ●
- ✓ 1000_0.9_0.005_256_0.1 ●
- ✓ 1000_0.95_0.005_256_0.1 ●

The result told me that it found a better way to fly lower on the no-landing status! Maybe it was not a bad news for me.

Consequently, I made the same test as I made for the $\gamma = 0.95$ model. I tried to change the noise scale to make it has better exploration capability.



- ✓ 1000_0.9_0.005_256_0.1 ●
- ✓ 1000_0.9_0.005_256_0.2 ●
- ✓ 1000_0.9_0.005_256_0.5 ●
- ✓ 1000_0.9_0.005_256_1.0 ●

Every model just find the no-landing state and kept on it.

I thought maybe it will be helpful to change τ and made our model has more update on the correct direction when it found how to land.



✓ 1000_0.9_0.005_256_0.5

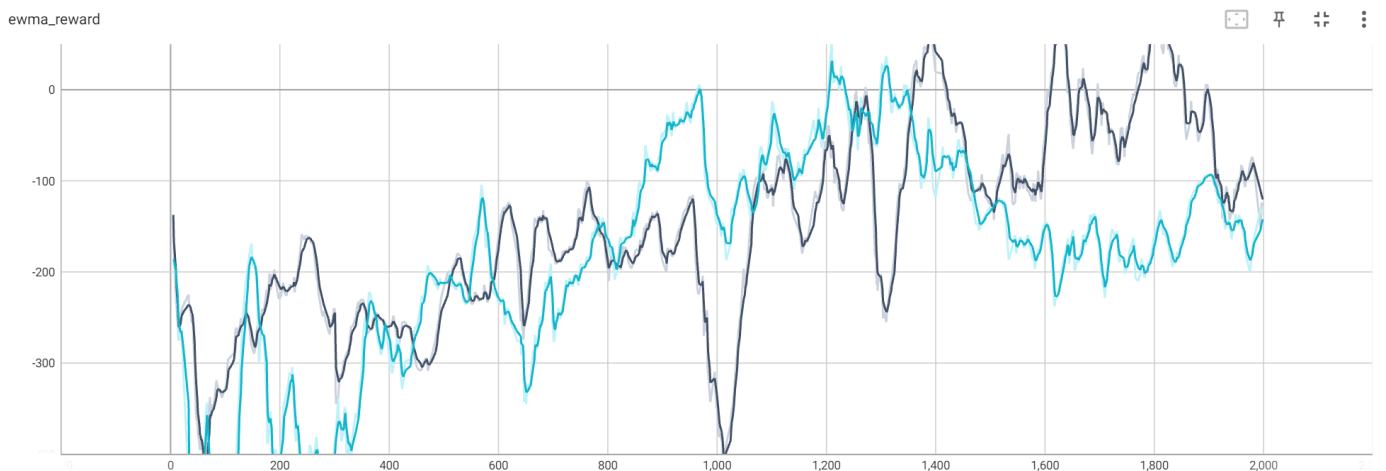


✓ 1000_0.9_0.025_256_0.5



It had better perform than the $\tau = 0.005$ case on beginning, whereas it did not have outstanding or stead improvement after that.

After finishing all this experiment, I try to enhance the episode value to make my model has more time to learn anything about landing.



✓ 2000_0.95_0.005_256_0.1



✓ 2000_0.95_0.005_256_0.15



The times it success is indeed increasing as the training time getting longer. I also change its noise scale value on this experiment. I have a conclusion is that higher noise scale value truely help the model to explore on the early time; nonetheless, it may make the model has worse performance when it has to exploitation as the better exploration makes its searching space be larger.