HLAI- ASSIGNMENT1

NAME: Roopam Verma

ID: 1001957590

**Part-1**

1. TODOs in starter code for equation of policy gradient, reward to go and reward discounting from learning\_algorithms.py file:

Text

Description automatically generated

**Part-2**

1. Util function for equation 6,7 and 8:

Text

Description automatically generated

1. Create a graph that compares the learning curve from the four trials above. Label the curves as t0, t1,t2.

Chart

Description automatically generated

1. Second version is best performing with maximum reward. And with Average trajectory reward 60.90.

On trying same trials with random seed below graph was observed. And it can be seen that in tr1 there is high variance when compared to tr0 and tr2.

Chart, scatter chart

Description automatically generated

**Part-2**

**LunarLander**

Chart, scatter chart

Description automatically generated

It has been observed that if we increased number-of-trajectory-per-rollout, there is an increase in learning curve with increase in average trajectory reward. In this case last trial had good performance.

Github repository link: <https://github.com/roopamrv/learning_algorithm.git>