**Reinforcement Learning Assignment 2 Technical report**

1. CartPole-v0 Problem: 當Running reward為195.06264631943262時會收斂，the last episode runs to 199 time steps，Train一共需要228個Episodes，使用的Learning Rate為預設的0.01。Test十次的平均Reward為200.0。
2. LunarLander-v2 Problem: 當Running reward為201.5251419139687時會收斂，the last episode runs to 303 time steps，若設定Learning Rate為預設的0.01會沒辦法收斂(跑了5000個Episodes以上仍無法達到收斂條件)，因此嘗試過多種不同的Learning Rate，當設定Learning Rate為0.001時，在經過1913個Episodes後可以收斂，經過觀察訓練過程，ewma reward會在大概第1000個Episode左右從負數變為正數，並在40-170這個range不停改變，但在第1400-1500個Episode又會從正數變為負數，在第1800-1900個Episode又會從負數變為正數，並在第1913個Episode達到收斂條件。Test十次的Reward分別是195.19862532552824, 212.47865146342127, 21.655981339513303, 269.4983432107955, 87.64718612071329, 254.02790544996992, 273.3909541449898, 248.14680720725124, 21.18916996460591, 257.8858516327336，這十次的平均Reward為184.1119475859522。