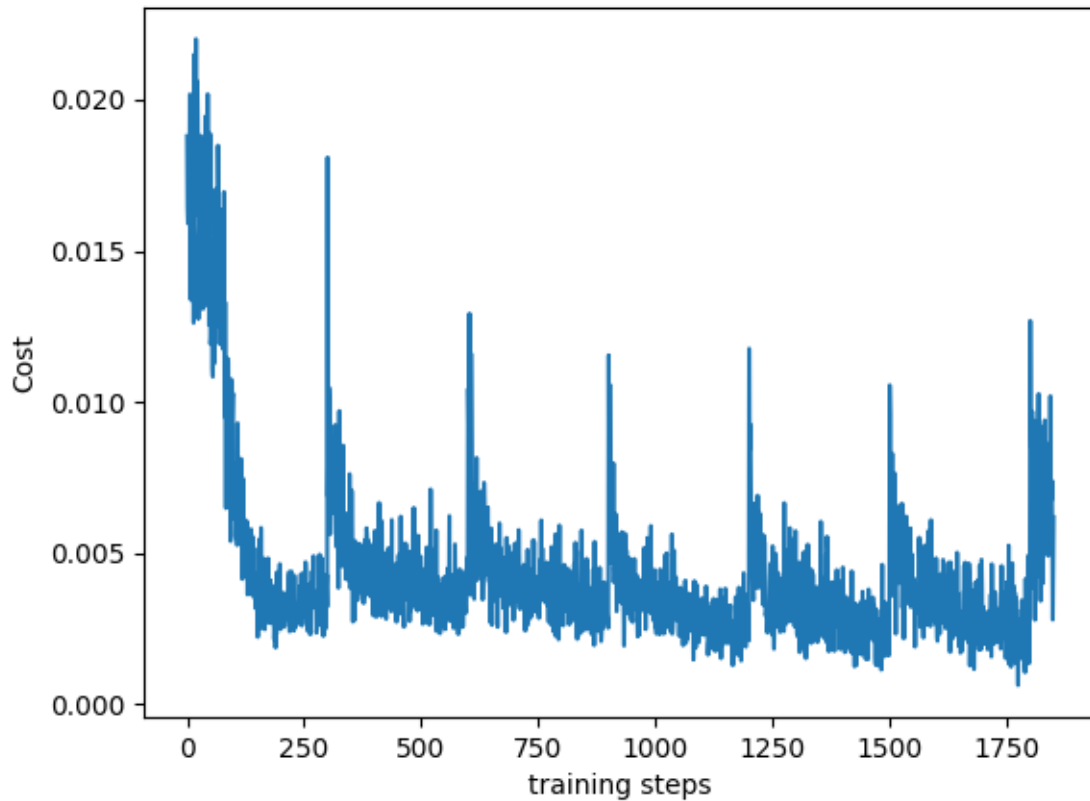


Report Analysis

$\text{reward} = \text{abs}(\text{position} + 0.5)$ ，車開的越高，reward 就越大
下圖是 step 和 cost 的關係圖



Questions

1. What kind of RL algorithms did you use? value-based, policy-based, model-based? Why? (10%)

ANS: 我使用 Deep Q Learning。是 value-based 演算法，因為車每走一步都能通過函式更新 reward，從而計算出下一步的 action。

2. This algorithms is off-policy or on-policy? Why? (10%)

ANS: 這個演算法是 off-policy 的，因為有每一個 state 都會更新 reward，並且通過保存每一次的 action 和 reward 等資訊去學習更新下一步的 action。所估計的 value function 和生成樣本時所採用的並不完全一樣，所以是 off-policy。

3. How does your algorithm solve the correlation problem in the same MDP? (10%)

ANS: 利用保存每一步 step 的資訊的方法，對於相同的類似的問題，我認為都能夠套用相同的演算法解決。只是在 reward 和 action 的參數上需要稍作調整即可。

References

<https://morvanzhou.github.io/tutorials/machine-learning/reinforcement-learning/4-4-gym/>