National Taiwan University
Department of Electrical Engineering
Name: TU-CHIN CHIANG
ID: R11921038

112-1 (Fall 2023)
Reinforcement Learning

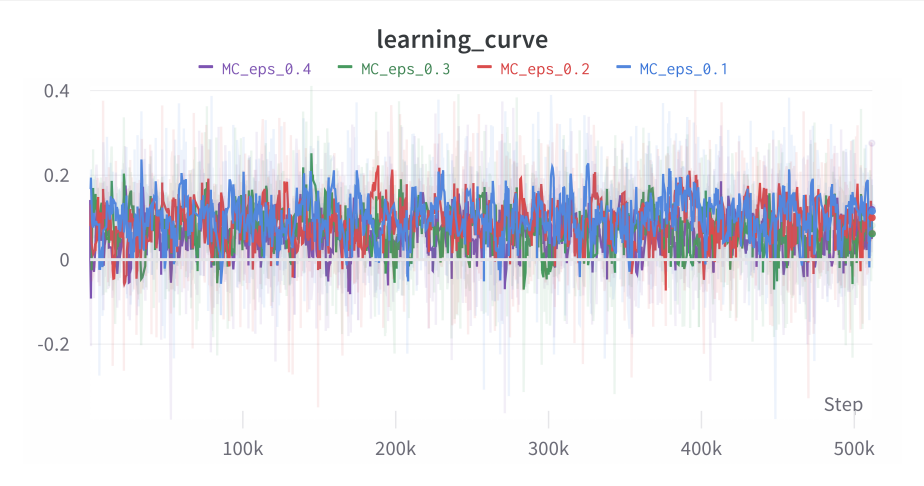
Assignment 2 Report

Q1. Discuss and plot learning curves under ϵ values of $(0.1, 0.2, 0.3, 0.4)$ on MC, SARSA, and Q-Learning

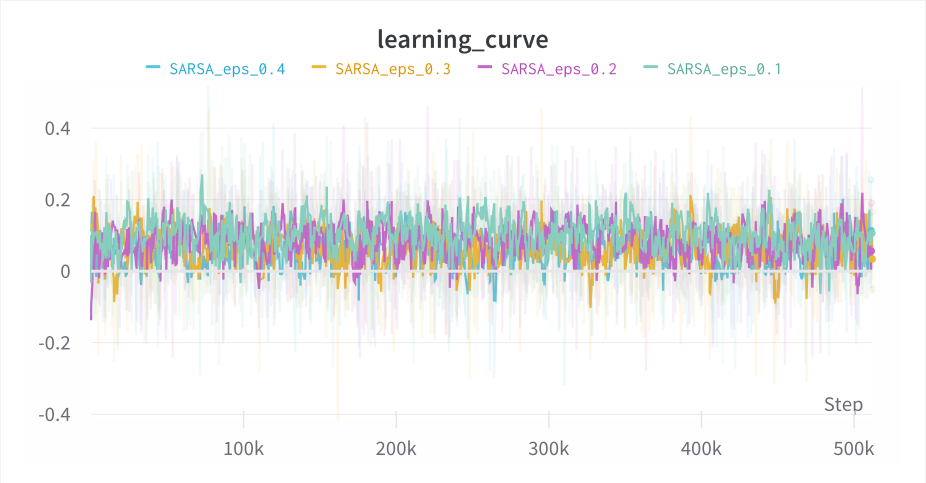
The graphs presented in Figures 1a, 1b, and 1c were generated using Weights & Biases, incorporating a smoothing rate of 0.5. Each data point represents the average non-discounted episodic reward calculated over the last 10 episodes. Besides, the experiment utilized default values in the presentation slide for the parameters of each method.

In general, the smoothed learning curves for MC, SARSA, and Q-Learning methods predominantly fall within the range of 0 to 0.2, showing no significant upward trend. Concerning the different methods employed, there is no discernible distinction in terms of episodic reward performance.

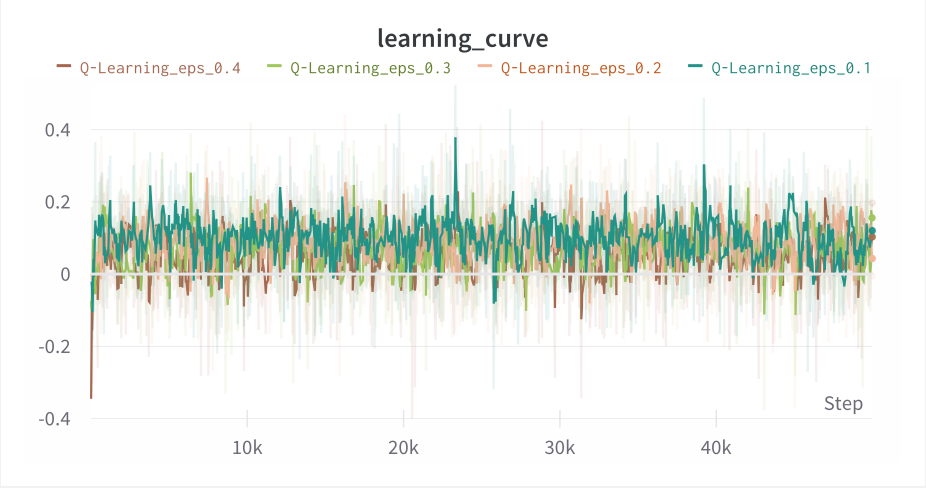
Examining the impact of varying ϵ values within the same method, a slight difference in performance is observed. Specifically, utilizing a smaller ϵ value tends to yield slightly better results compared to using a larger ϵ value. For instance, in Figure 1c, the green line ($\epsilon = 0.1$) appears to surpass the dark brown line ($\epsilon = 0.4$).



(a) MC learning curves



(b) SARSA learning curves



(c) Q-Learning learning curves

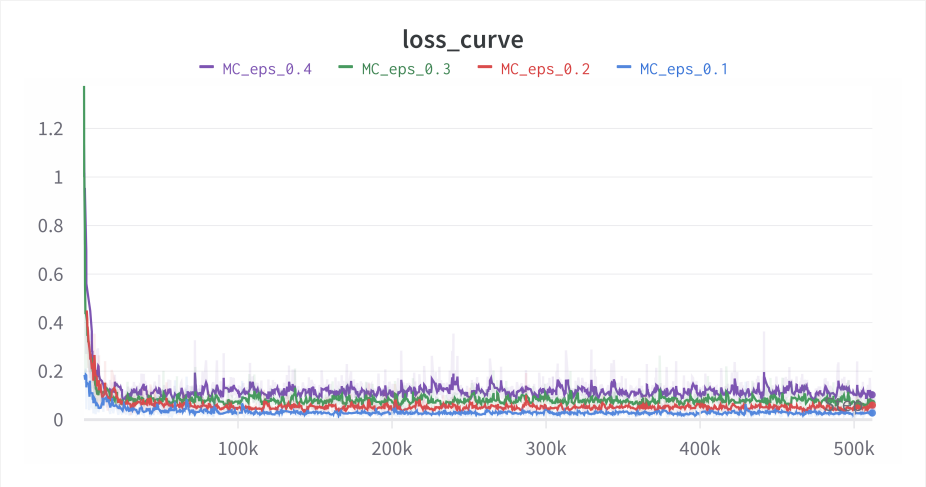
Figure 1: Learning curves with ϵ values of $(0.1, 0.2, 0.3, 0.4)$

Q2. Discuss and plot loss curves under ϵ values of $(0.1, 0.2, 0.3, 0.4)$ on MC, SARSA, and Q-Learning

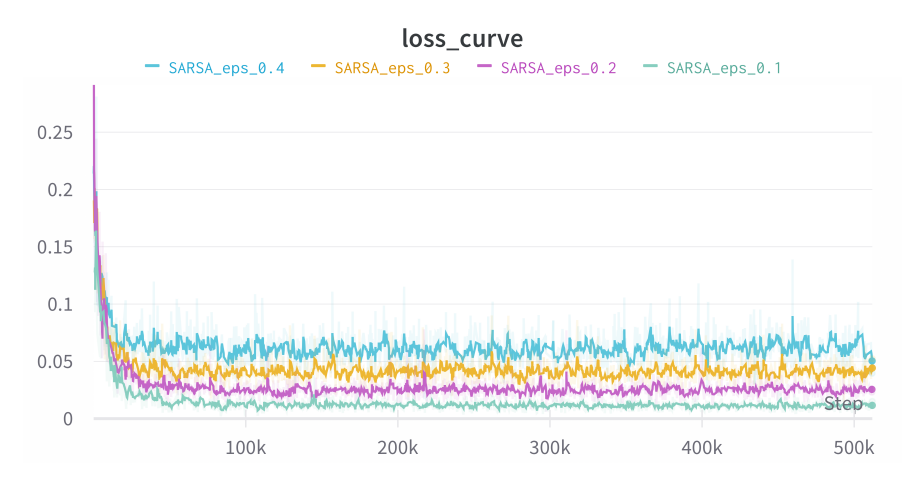
The charts depicted in Figures 2a, 2b, and 2c were generated using Weights & Biases with a smoothing rate of 0.5. Each data point represents the average absolute estimation loss calculated over each transition in the last 10 episodes. Furthermore, default parameter values, as outlined in the presentation slide, were employed in the experiment for each method.

In the graphs corresponding to the MC and SARSA methods, it is evident that utilizing a smaller ϵ value results in a noticeably reduced loss value. In the case of Q-Learning, each configuration converges to a low loss value. However, the curve associated with a larger ϵ decreases more rapidly than that of a smaller ϵ.

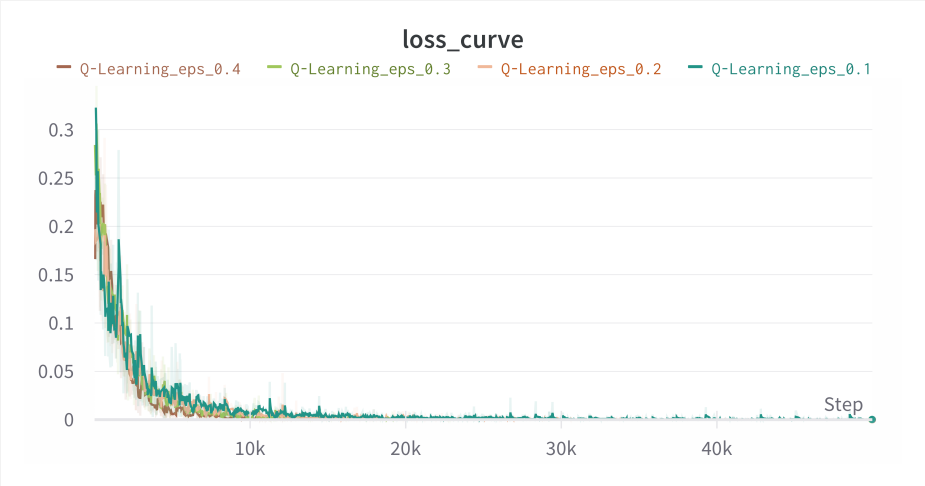
Additionally, when comparing the MC, SARSA, and Q-Learning methods, Q-Learning outperforms the other methods in terms of the estimation loss value. Therefore, Q-Learning exhibits faster convergence to a small loss, with its overall performance surpassing that of the other methods in this particular aspect.



(a) MC loss curves



(b) SARSA loss curves



(c) Q-Learning loss curves

Figure 2: Loss curves with ϵ values of $(0.1, 0.2, 0.3, 0.4)$

Q3. Discuss and plot ...

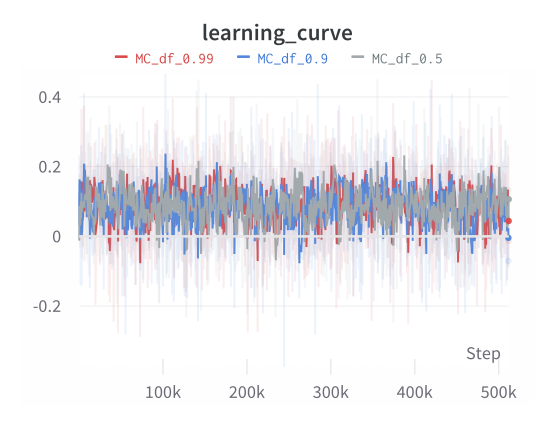
In this section, a grid search is performed to explore different discount factors, learning rates, update frequencies, and sample batch sizes. The corresponding charts for each result are depicted from Figure 3 to 6, respectively. Additionally, a detailed discussion of the learning and loss curves is provided for each figure.

Discount Factor

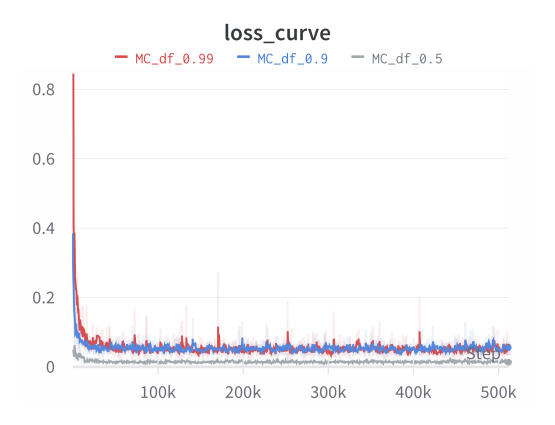
The learning curves and loss curves corresponding to different discount factor values are depicted in Figure 3. Figure 3a and 3b depict the results of MC policy iteration, and Figure 3c and 3d detail the results of SARSA. Figure 3e and 3f are the experiment records of Q-Learning. The experiments adhered to default values for the remaining parameters.

As with the previous observations, the learning curves show little variation across different discount factor values, with episodic rewards clustered within a limited range. In contrast, the loss curves reveal that the estimation loss converges rapidly to a lower value when a small discount factor is employed. This suggests that prioritizing states in the near future may enhance learning efficiency

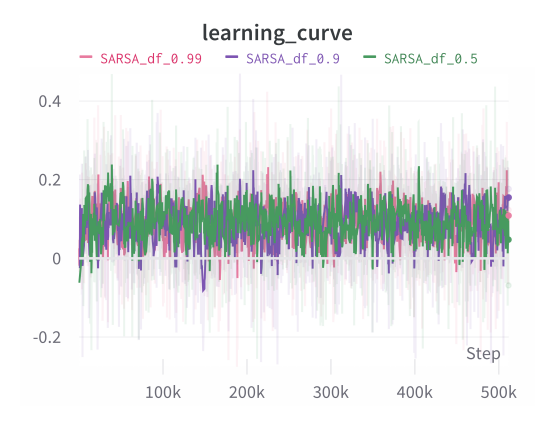
in grid world scenarios.



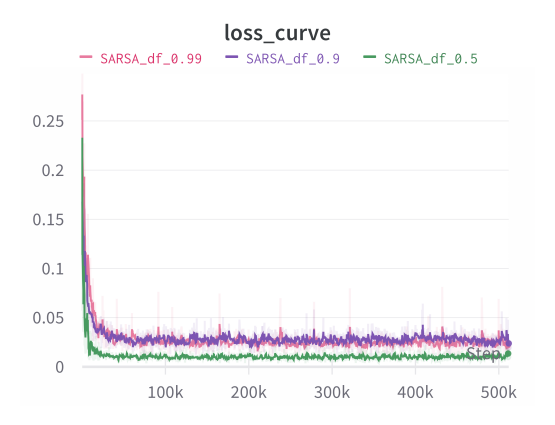
(a) MC learning curves



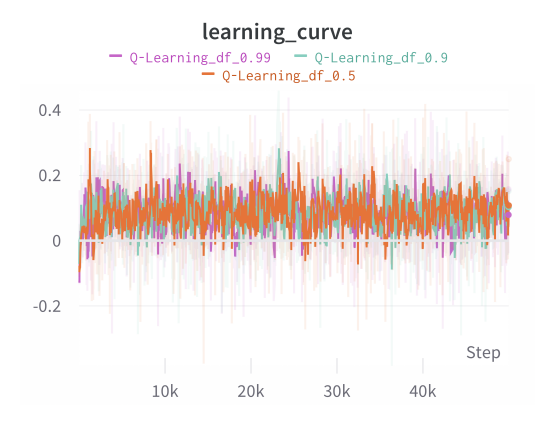
(b) MC loss curves



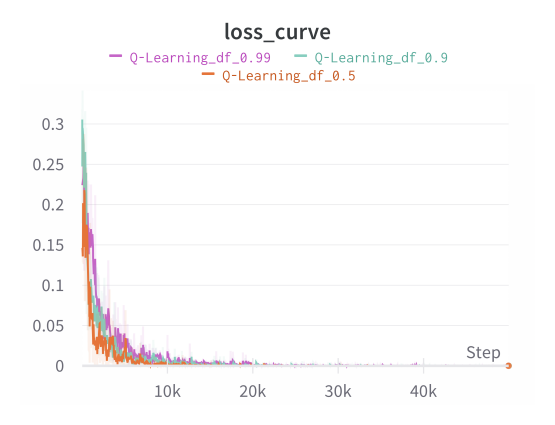
(c) SARSA learning curves



(d) SARSA loss curves



(e) Q-Learning learning curves



(f) Q-Learning loss curves

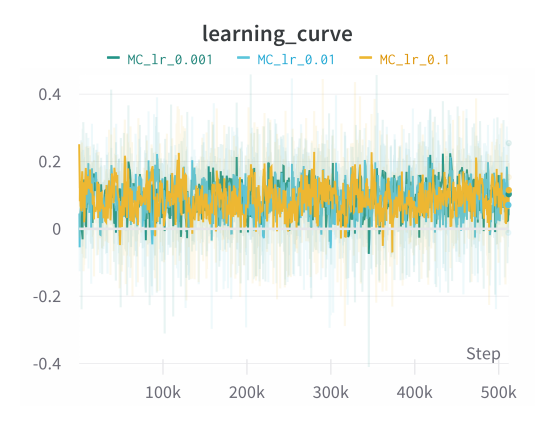
Figure 3: MC, SARSA, and Q-Learning with discount factor of $(0.5, 0.9, 0.99)$

Learning Rate

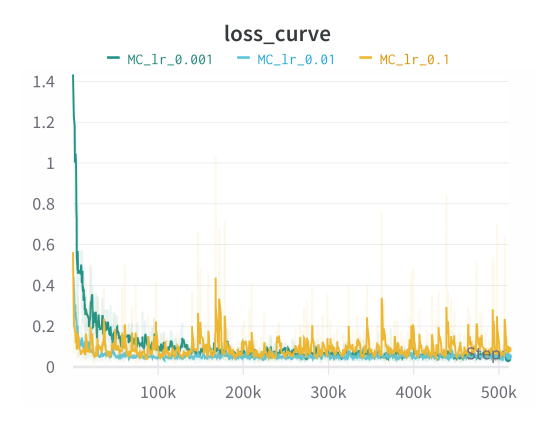
The learning curves and loss curves corresponding to different learning rates are illustrated in Figure 4. The experiments maintained default values for the remaining parameters.

Upon examination, the graphs indicate that different learning rates are suitable for various methods. In the case of MC policy iteration, a learning rate of 0.1 results in significant variance in the loss

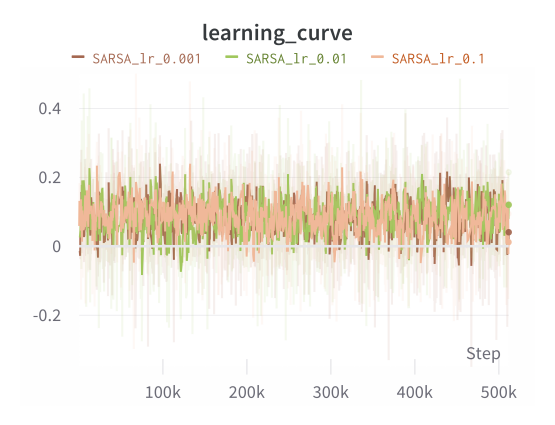
values, whereas learning rates of 0.01 and 0.001 lead to stable decreasing curves. For SARSA and Q-Learning, it appears that a learning rate of 0.001 is too low for efficient convergence. In summary, higher learning rates are more appropriate for SARSA and Q-Learning, while a lower learning rate is suitable for MC.



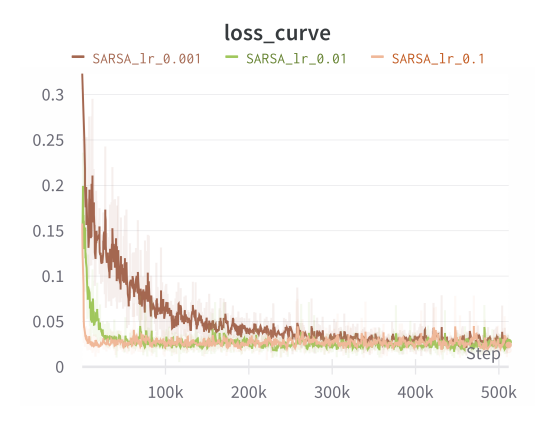
(a) MC learning curves



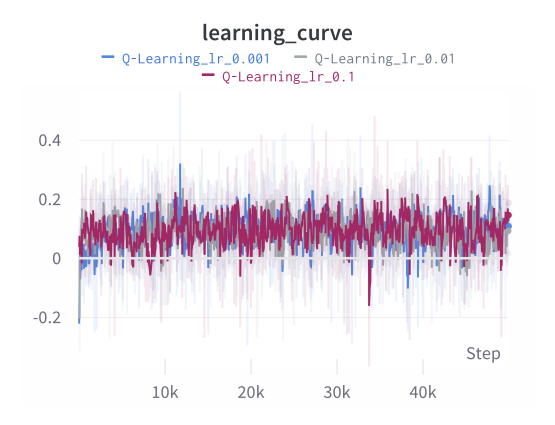
(b) MC loss curves



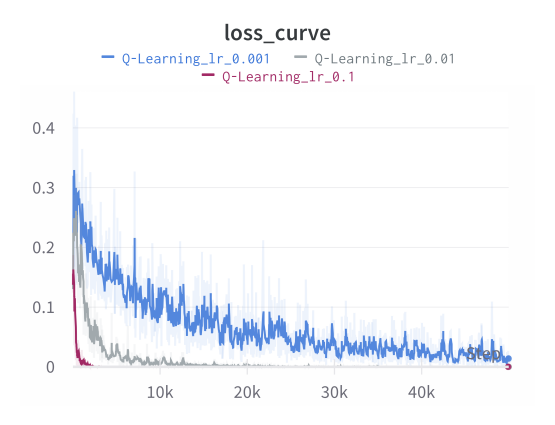
(c) SARSA learning curves



(d) SARSA loss curves



(e) Q-Learning learning curves



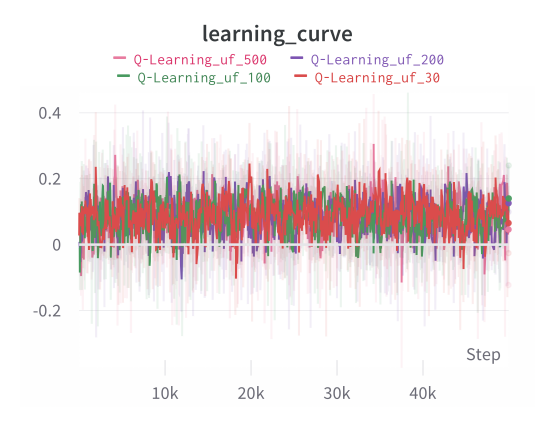
(f) Q-Learning loss curves

Figure 4: MC, SARSA, and Q-Learning with learning rate of $(0.1, 0.01, 0.001)$

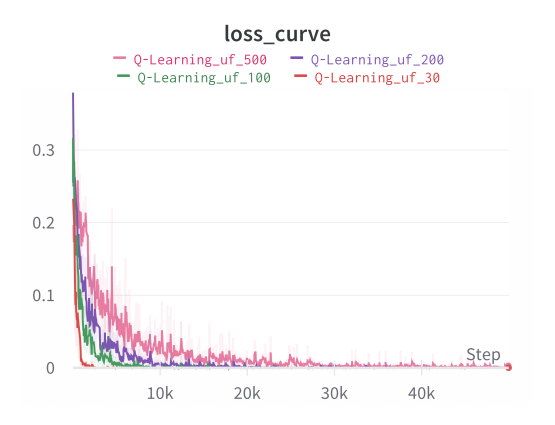
Update Frequency

The learning curves and loss curves corresponding to different learning rates are depicted in Figure

5. The experiments retained default values for the remaining parameters. Referring to Figure 5b, it is observed that a lower update frequency results in better performance in terms of estimation loss. This improvement may be attributed to the more frequent correction of Q values.



(a) Q-Learning learning curves

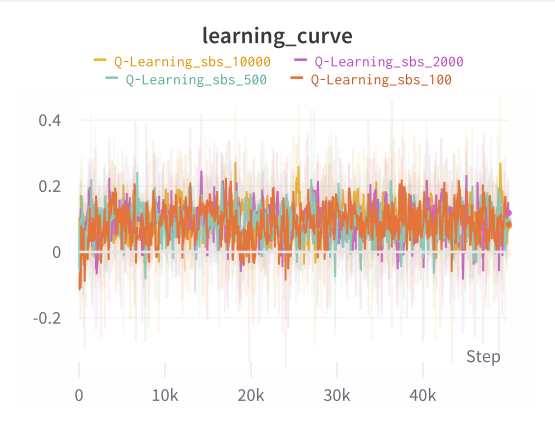


(b) Q-Learning loss curves

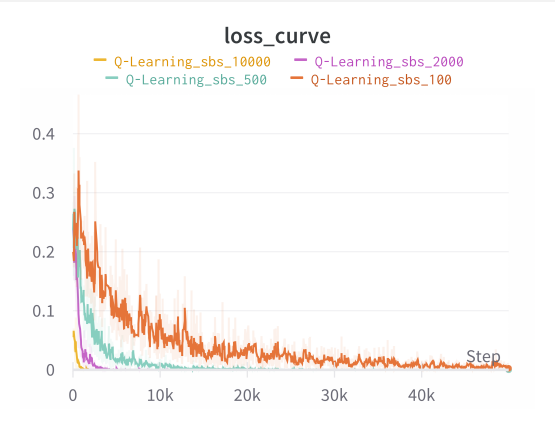
Figure 5: Q-Learning with update frequency of $(30, 100, 200, 500)$

Sample Batch Size

The learning curves and loss curves associated with different learning rates are presented in Figure 6. The experiments retained default values for the remaining parameters. Examining Figure 6b, it is evident that a larger sample batch size leads to improved performance in terms of estimation loss. This suggests that employing a larger batch size, in conjunction with buffer size, may enhance learning performance.



(a) Q-Learning learning curves



(b) Q-Learning loss curves

Figure 6: Q-Learning with sample batch size of $(100, 500, 2000, 10000)$