Deep Learning Lab Course

Exercise 1 : Imitation Learning and Reinforcement Learning

Name: Roshin Rajan Panackal Matriculation No. : 4850879

1.2 Imitation Learning

Hyperparameter Settings

The network was optimized mainly for two hyperparameters, 1. Learning Rate 2. History Length. Best test time performance obtained on configuration:

History length: **5** and learning rate: **0.01** giving a result of Mean : **778.89** and Standard deviation: **229.95**

Problems Faced

- Agent was mostly stuck at the beginnings of the episodes.
- Agent slows down significantly while going over turn markers.
- Expert data contained actions capturing multiple key presses

Handling Imbalance

Each class is given same sampling probability of 1/L, where L is the number of classes. Sampling probability of a class is distributed across the samples. Then, probability of sampling a single data point x, belonging to class C is given by,

$$P(x) = 1/(L * N), \quad \text{where N is the \#samples in class C}$$

Therefore, each mini-batch has similar number of samples from each class.

Multi-key Presses

I chose to keep keys based on an order of key priorities as given below,

DOWN > LEFT, RIGHT > UP

Remarks

Most noticeable difference due to sampling by weights is that the agent's chance of failure to accelerate at the beginning of the episode is significantly reduced.

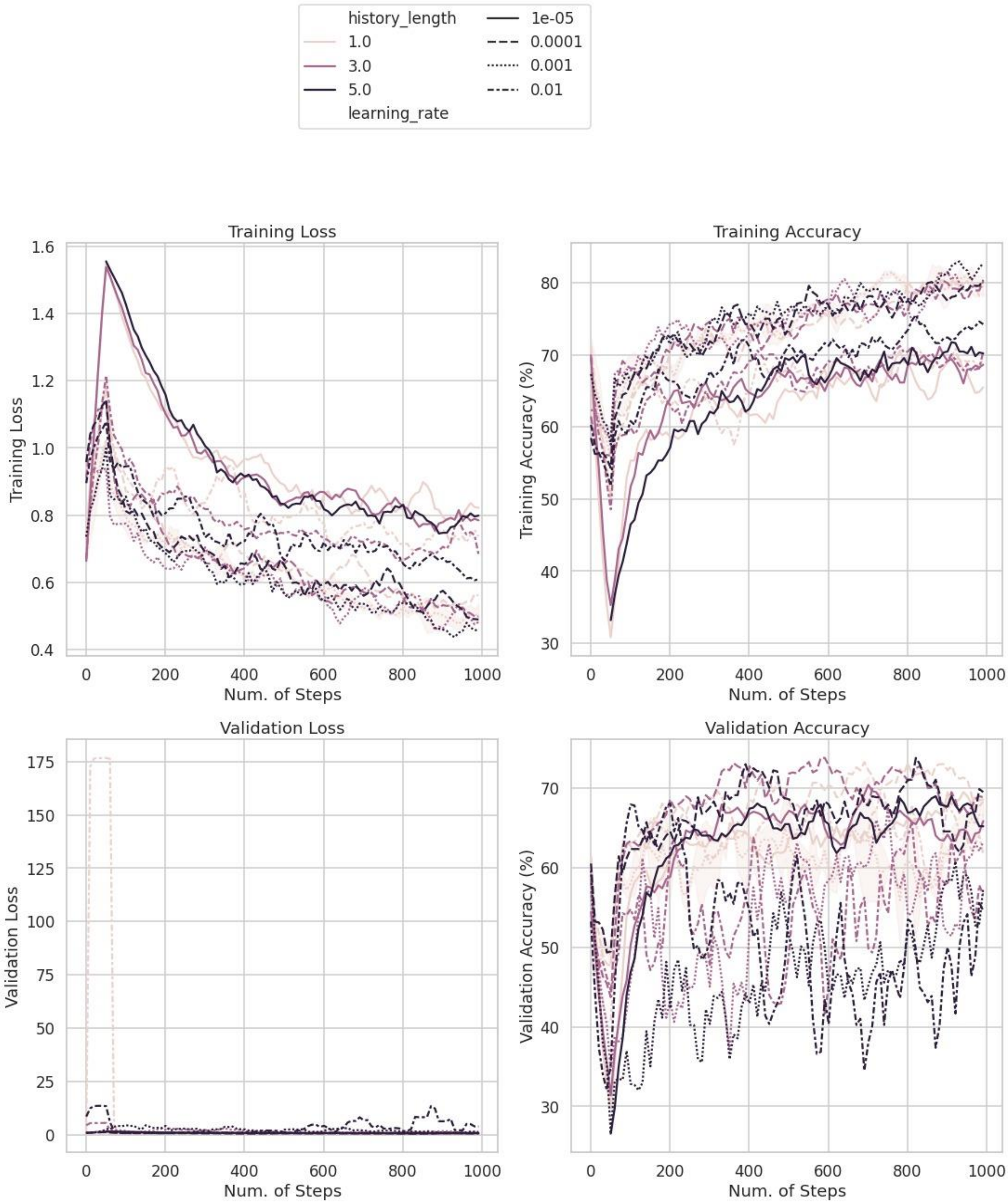
With longer history length agents tends to stick more strictly to the track, minimizing deviations.

At high learning rates, the agent was able to better return to track.

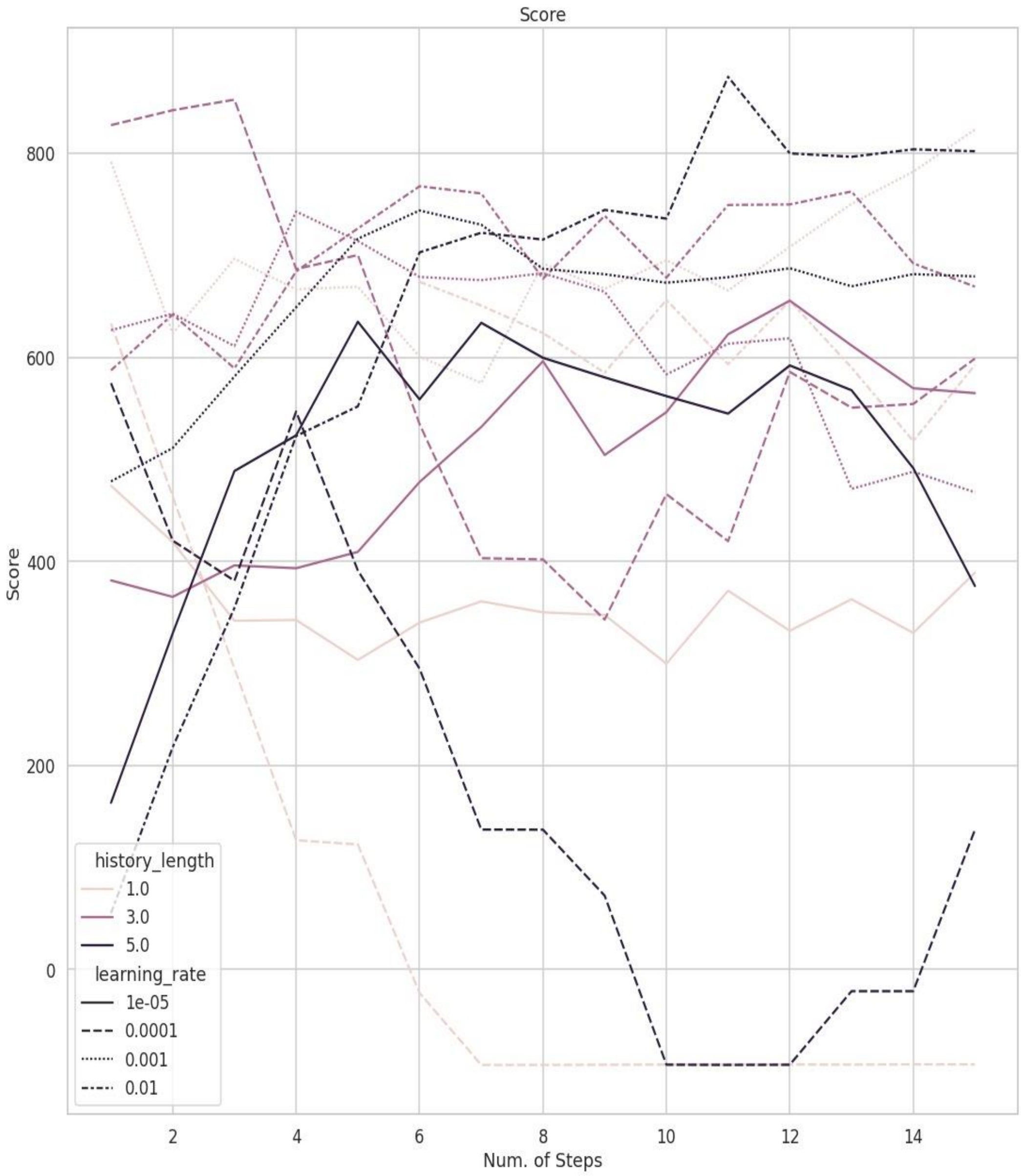
Limitations

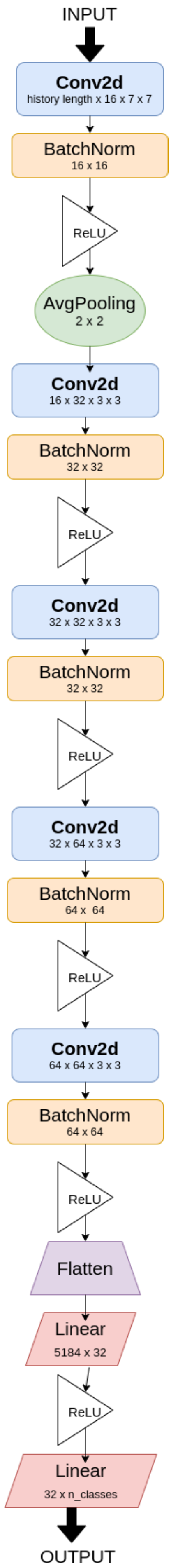
- Agent seems to rely heavily on markers (bents) on the image.
- Even though the agent trained with higher history lengths was better at sticking to the track, once gone off-track, it did poorly at its attempts to return back.

Visualization of Results



The agent performed best at higher history length and learning rates. Very low learning lead to deteriorating performance for all history lengths.





Test time performance of the agent across 15 episodes

Best top score 923 was obtained learning rate 0.01 and history length 3.

Table 1 Performance for best hyperparamter settings accross different history lengths.

History Length	Mean	Standard Deviation
1	778.509750784092	132.71139725312776
3	727.3431644627094	252.73547066962328
5	826.5686979542521	117.15375837417352

1.3 Reinforcement Learning

1.3.1 CartPole Agent

Hyperparameter Setting

Best performing setting:

Decay: 0.95 **Schedule:** Exponential Decay **Epsilon:** 0.2 **Learning Rate:** 1e4

Number of Episodes: 500

Experiments

Other than exponential decay of epsilon, linear and a custom decay schedule were tried.

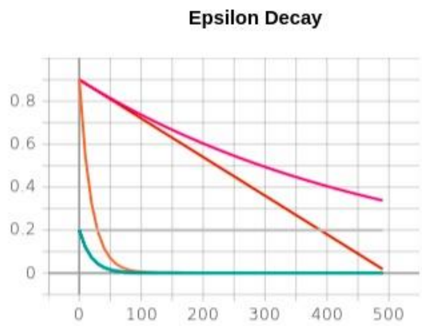
- *Observation:* Quickly cooling down the epsilon seems to give better results.

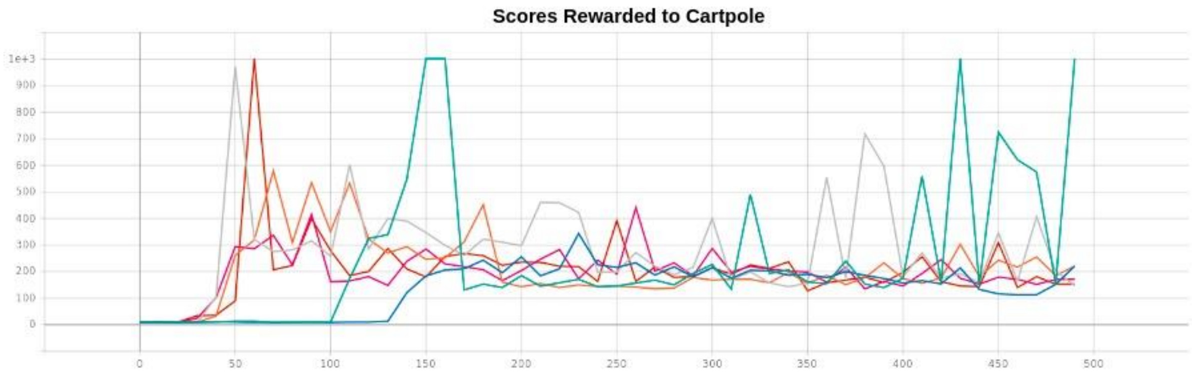
Deep Double Q-Learning agent was also tested.

- *Observation:* Even with the best observed hyperparameter settings, the agent was taking significantly longer to learn.

Visualization of Results

Color	Method and Hyperparameter
	Epsilon = 0.2, Decay=Exponential, Method = DQN
	Epsilon = 0.9, Decay=Exponential, Method = DQN
	Epsilon = 0.9, Decay=Experimental, Method = DQN
	Epsilon = 0.9, Decay=Linear, Method = DQN
	Epsilon = 0.2, Decay=None, Method = DQN
	Epsilon = 0.2, Decay=Exponential, Method = DDQN





1.3.2 Car Racing

Problems Faced

I was unable to train the DQN agent for Car racing task as my laptop kept crashing few minutes into training. Therefore, following experiments were tried for a span of approx. 8mins. I have still uploaded a fully running code of Car racing agent uploaded with this report.

Experiments

Adding frame skipping made the training process visibly faster.

Using pre-trained model from Behavioural Cloning agent, showed the agent taking more informed actions even at the early stage of learning.