

Feedback from Group 20 on HW3 for Group 22

Hey Group 22,

We read through your code for HW3 and would like to share our feedback with you. We really enjoyed how well structured and easy to understand your code was. Using a deque as a Buffer is a nice idea and keeps the implementation simple. You even thought of what to do with the datapoints that have `done=true` in your loss calculation.

We do however have some suggestions on how you could improve it even further:

- The variable names in the `ReplayBuffer.sample` method could lead to some confusions
- Instead of iterating through the model weights in the `Agent.soft_update_network` method you could also directly use `model.set_weights` somewhat like this:
`model_target.set_weights(0.9*np.array(model_target.get_weights(),dtype = object) + 0.1*np.array(model.get_weights(),dtype = object))`
- Using the average reward instead of the sum of rewards as a tool to evaluate the agents performance would be better, since the trajectories are always complete, i.e. they always run until they are done. Thus the trajectories have different lengths and the sum cannot be compared