Exercise 10

Kai Schneider

July 6, 2021

Task 1   DQN on the Cart-Pole

a.)

- NNs tend to diverge (due to high correlation states and actions, NOT i.i.d)
  $\rightarrow$ Experience Replay (sampling training data from buffer of past experiences)

- oscillation due to big changes in Q-values
  $\rightarrow$ much lower learning rate than in other DL-scenarios

- bias introduced at beginning of training and instability due to similar subsequend steps
  $\rightarrow$ DDQN (2 NNs) predictions of target NN are used to backpropagate training NN. target network is only synced periodically