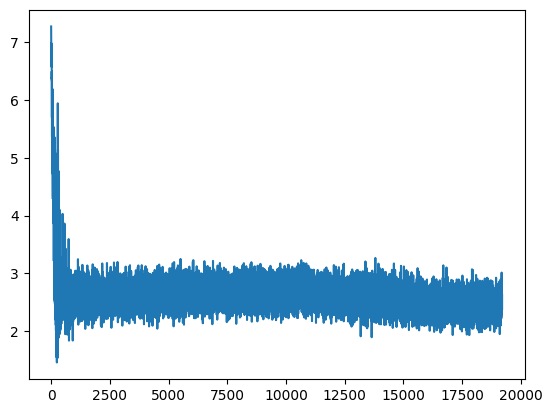
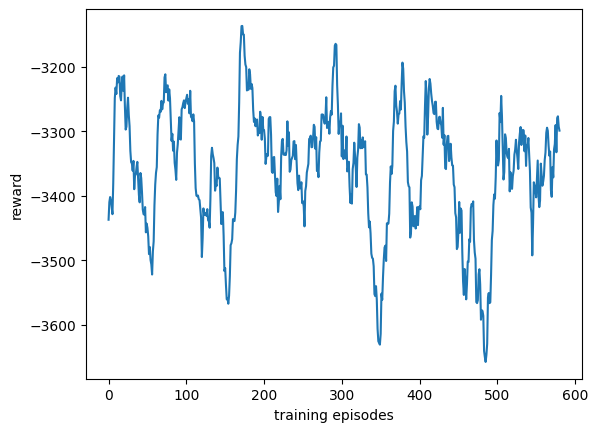
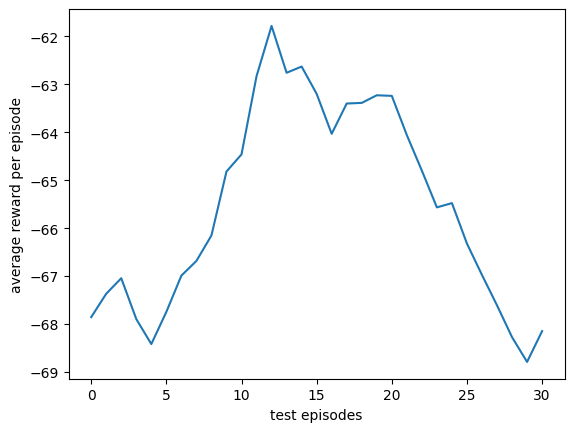
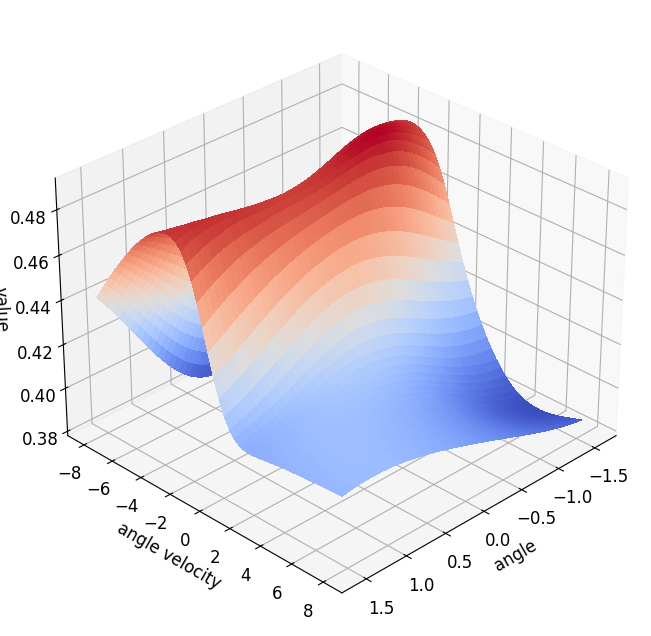
Pendulum  
  
DQN without target network:

Training loss and rewards:

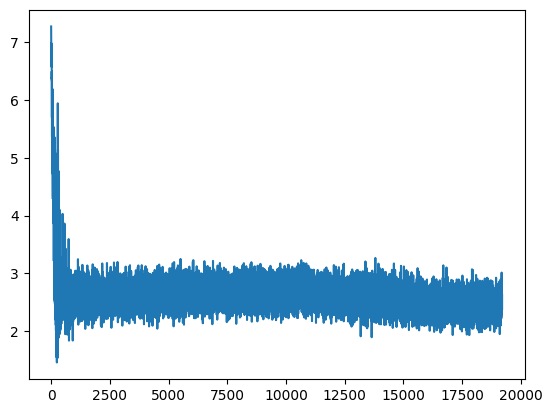
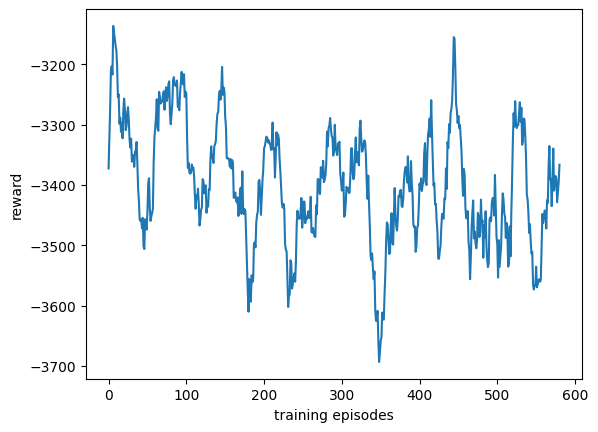


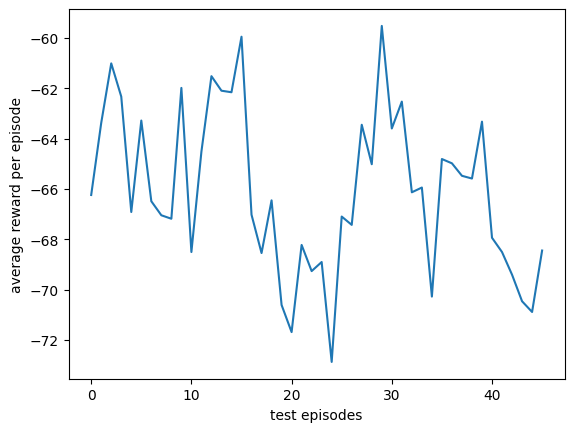
Test rewards:



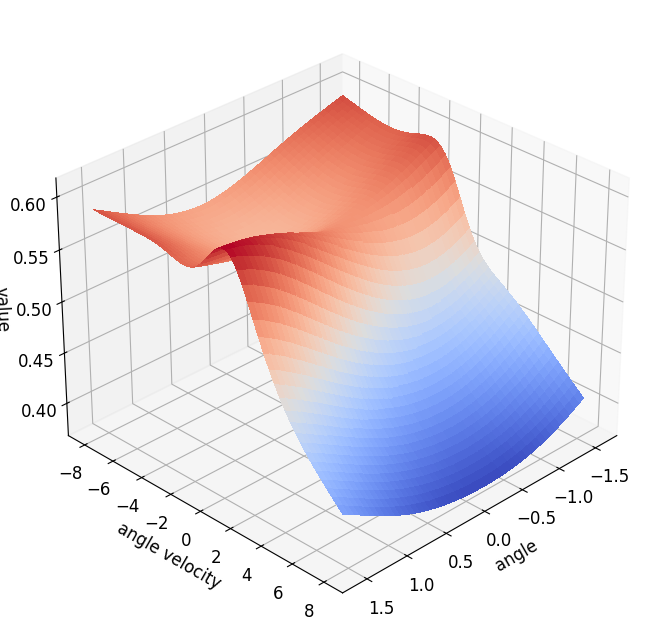


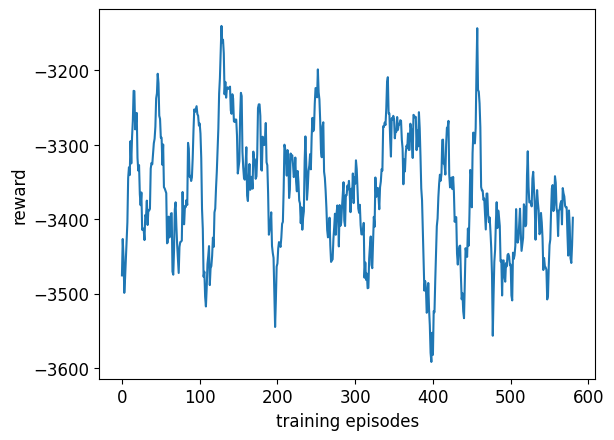
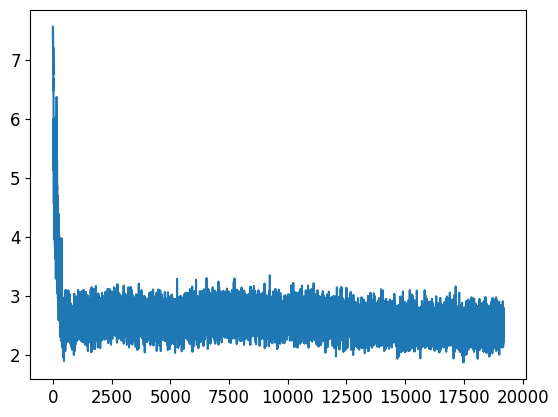
DQN with target network updated every batch training step  
  
training loss and rewards:

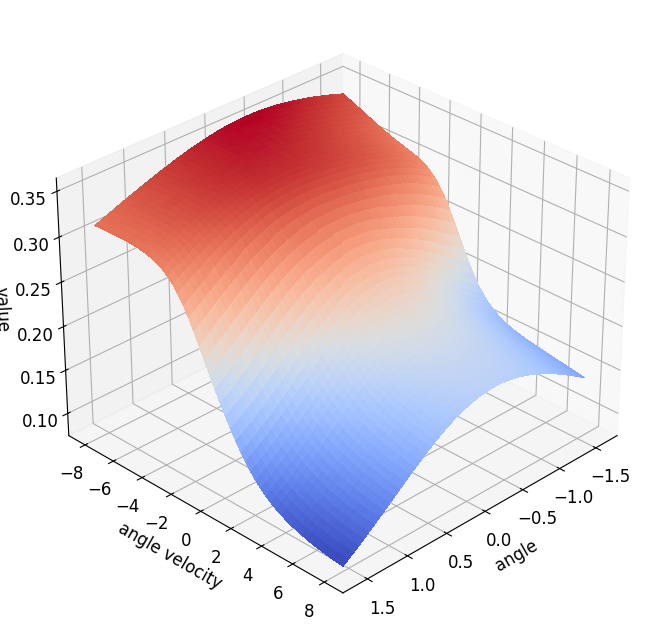
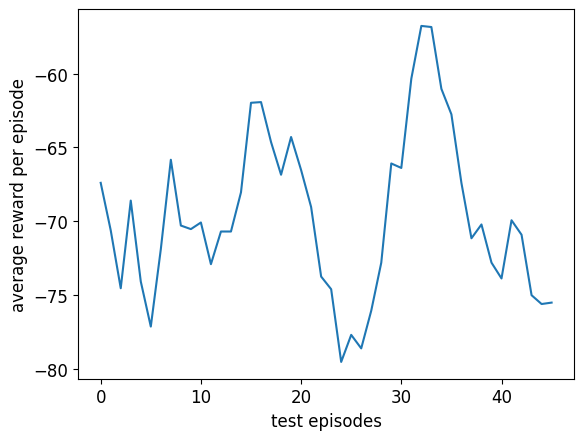
  
test rewards:



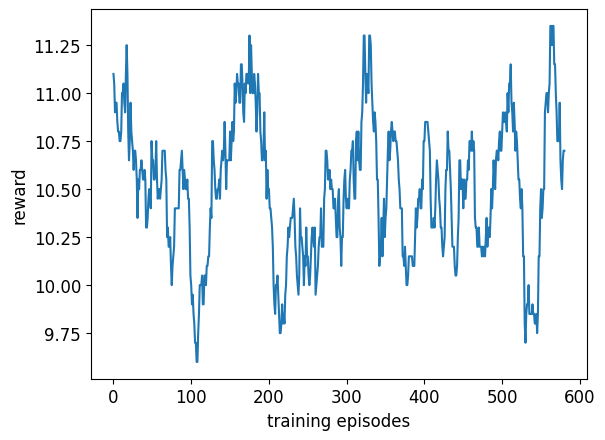
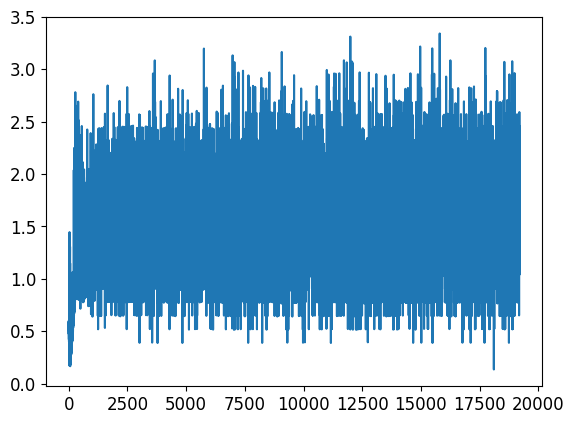
Value function (max q)

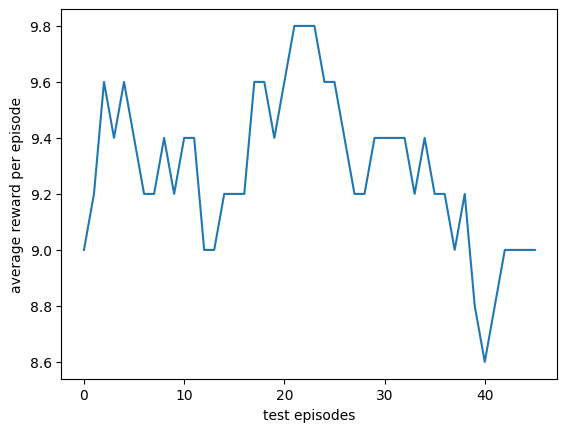
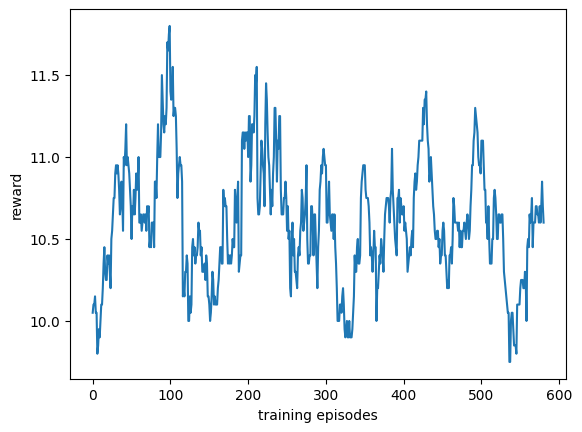
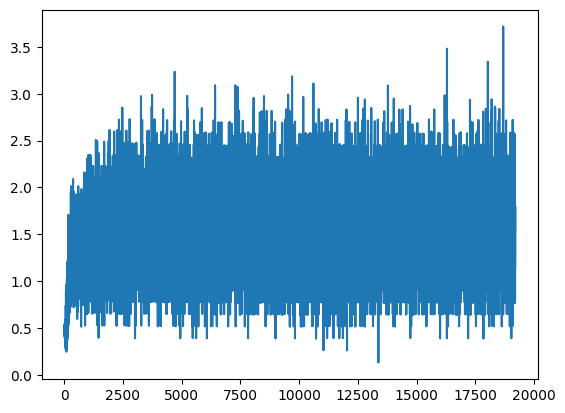


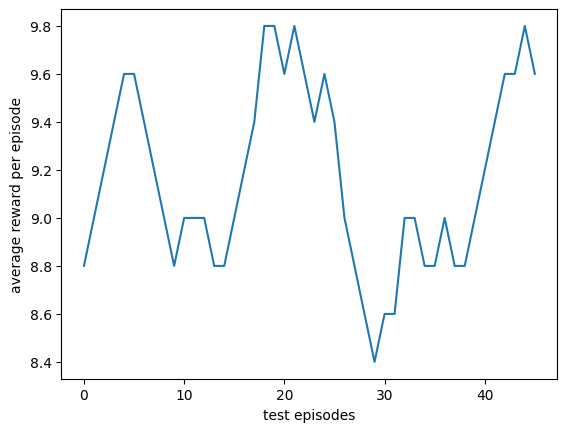
DQN with target network updated every 20 batch training steps  
 



Cartpole  
  
- without target network

  
Update every target network  
  
 



-update every 20 training calls  
  
