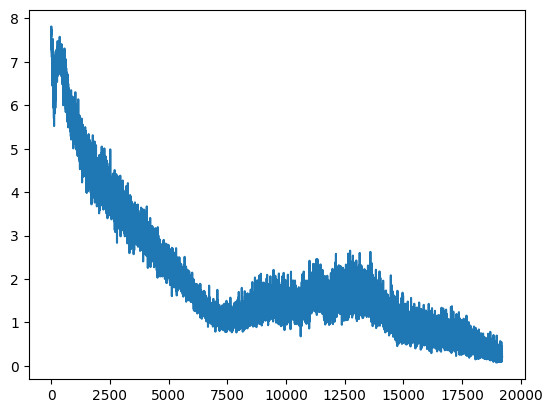
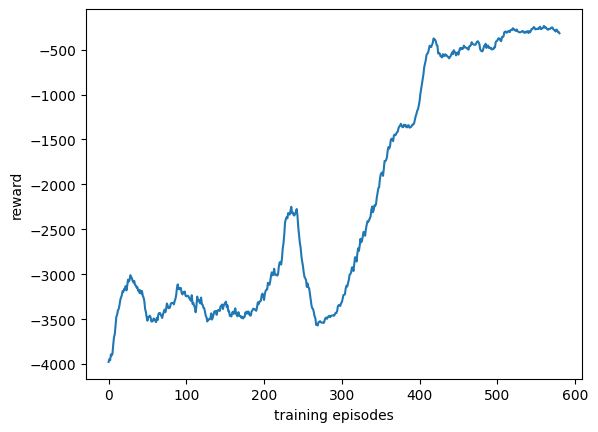
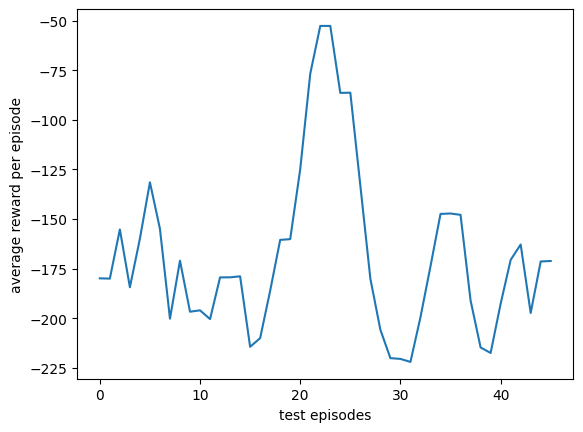
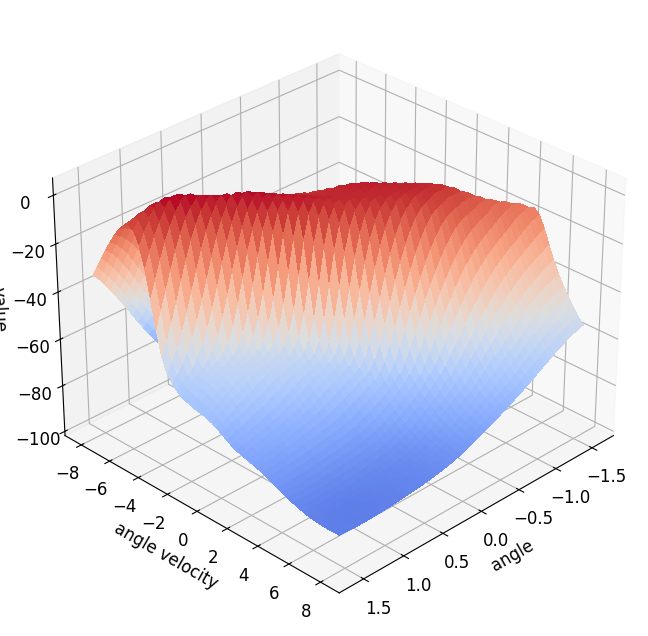
Pendulum  
  
DQN without target network:

Rewards and training loss:

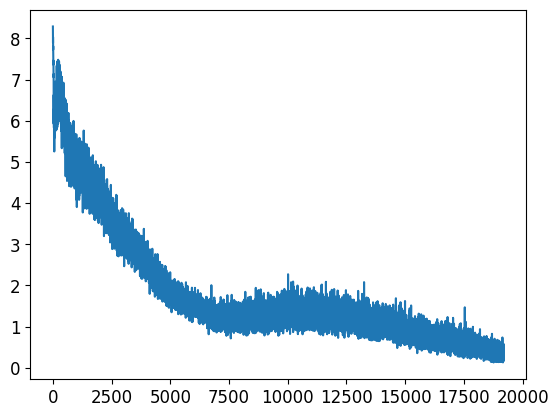
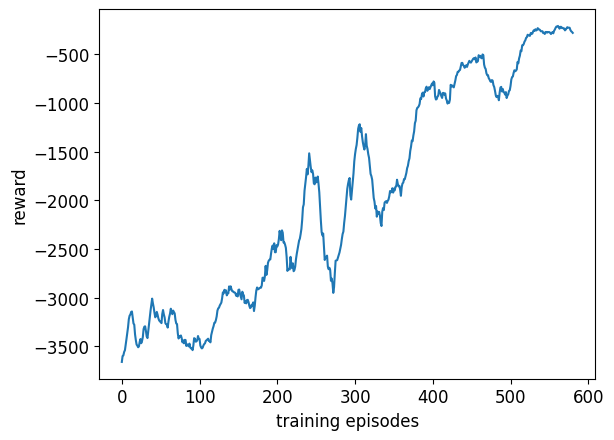


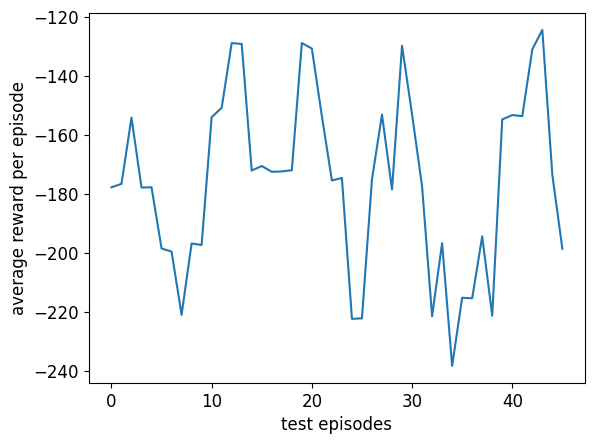
Test rewards:

avg = -166

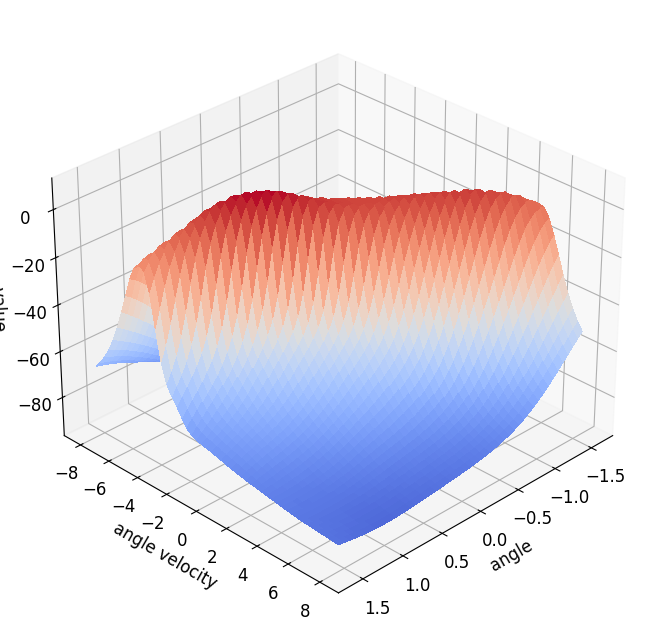


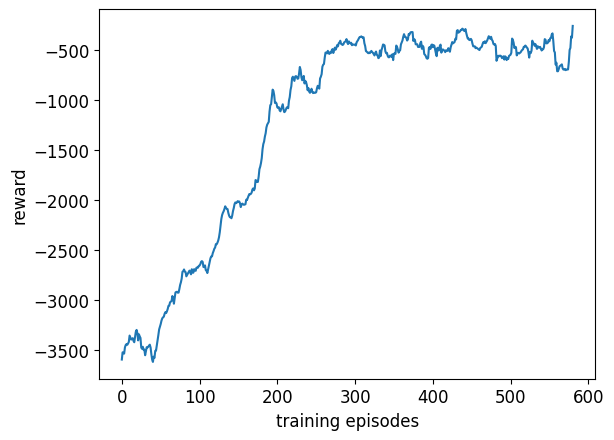
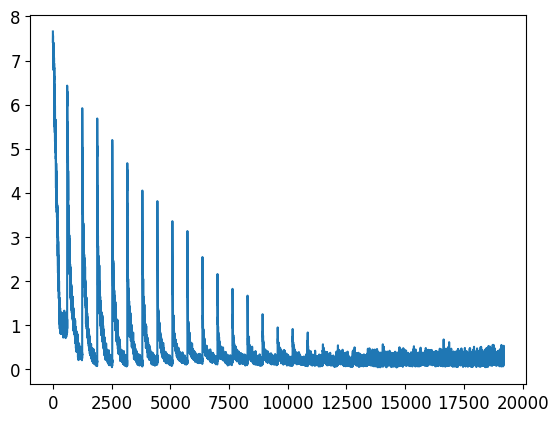
DQN with target network updated every batch training step  
  
training loss and rewards:

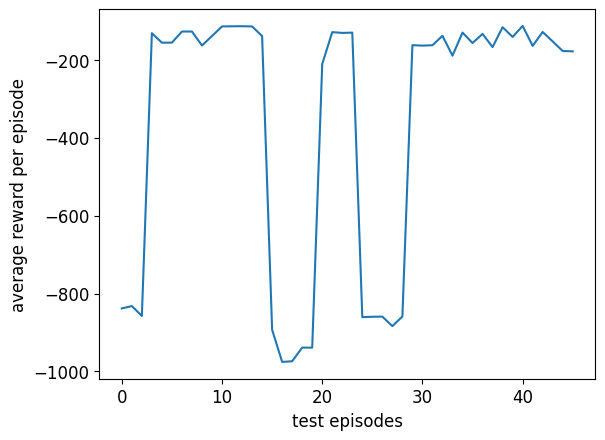
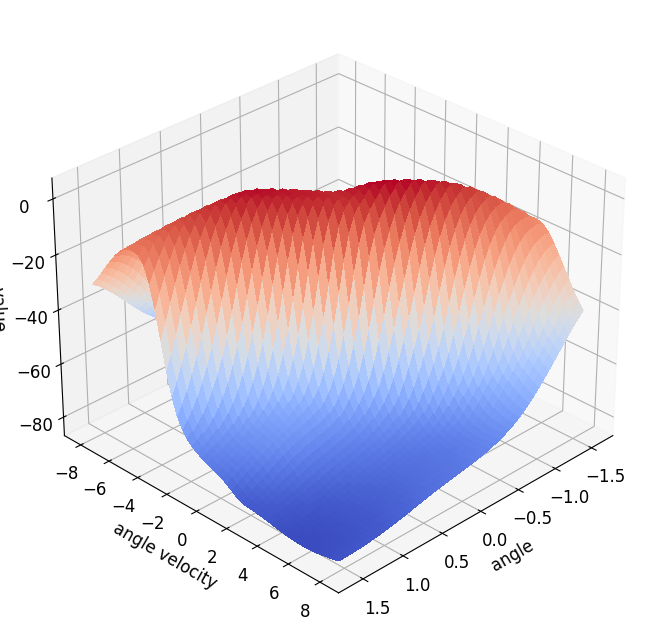
  
test rewards:

 avg = -177

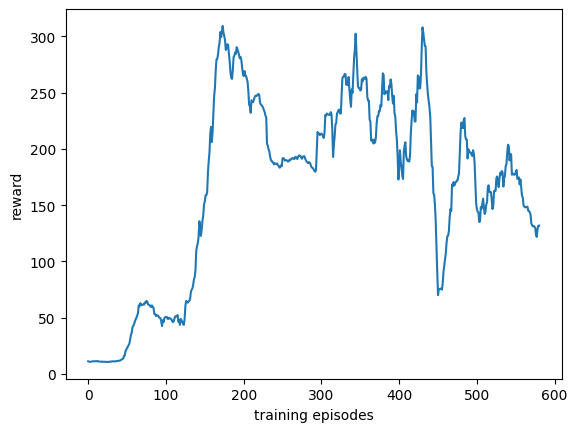
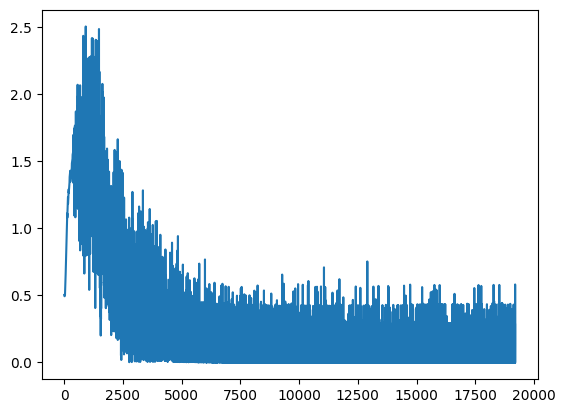
Value function (max q)

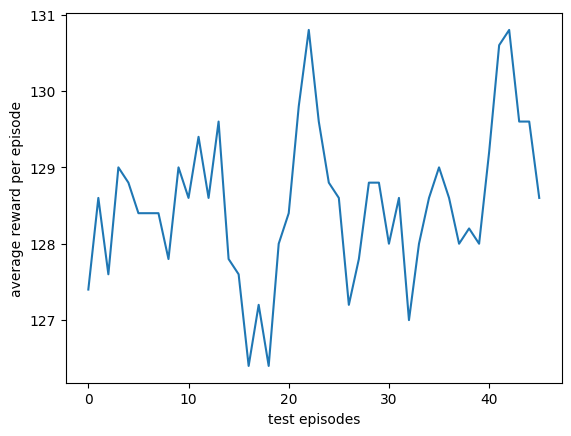
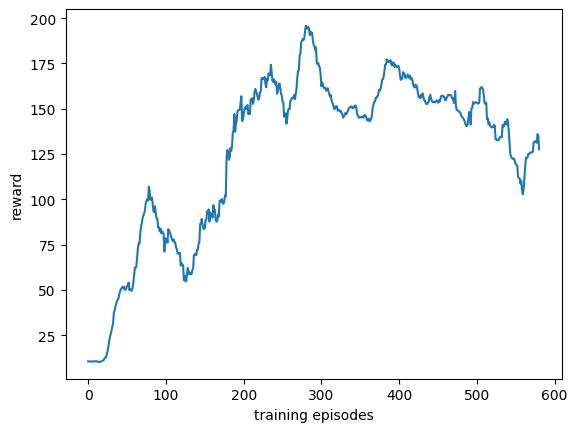
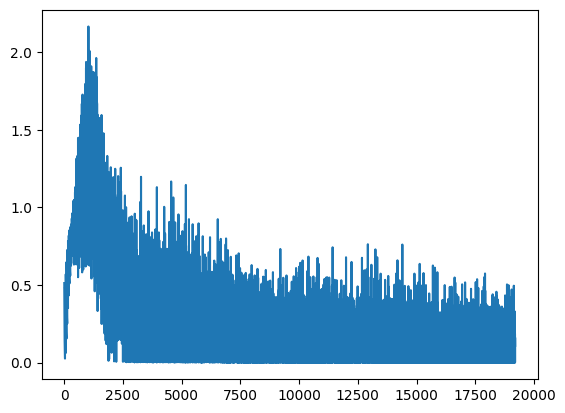


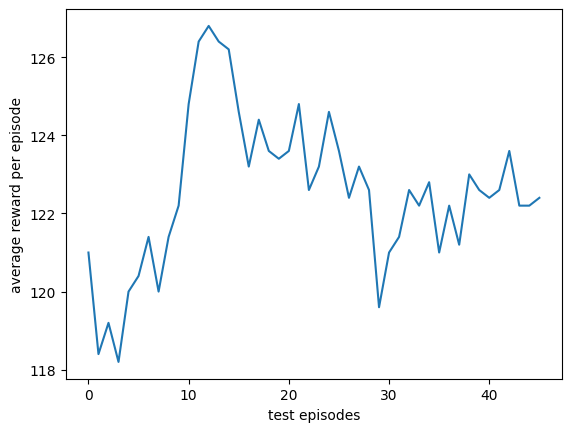
DQN with target network updated every 20 batch training steps  
 

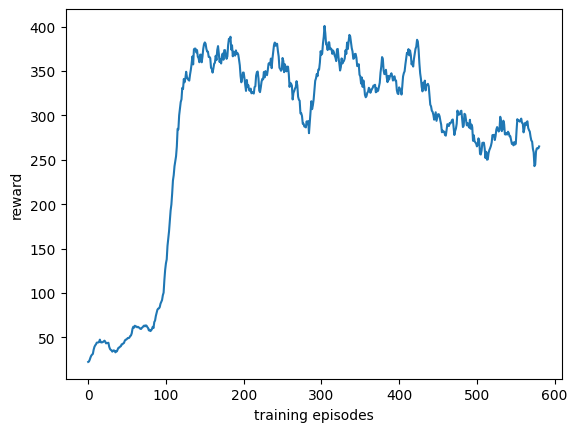
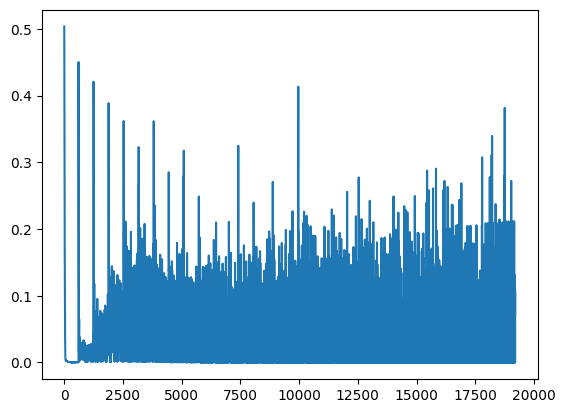
avg = -378

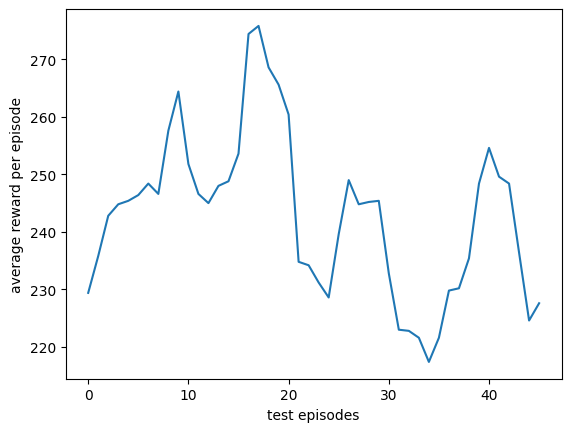
Cartpole  
  
- without target network

avg = 128  
Update every target network  
  
 

avg=122

-update every 20 training calls  
  
 

avg = 241