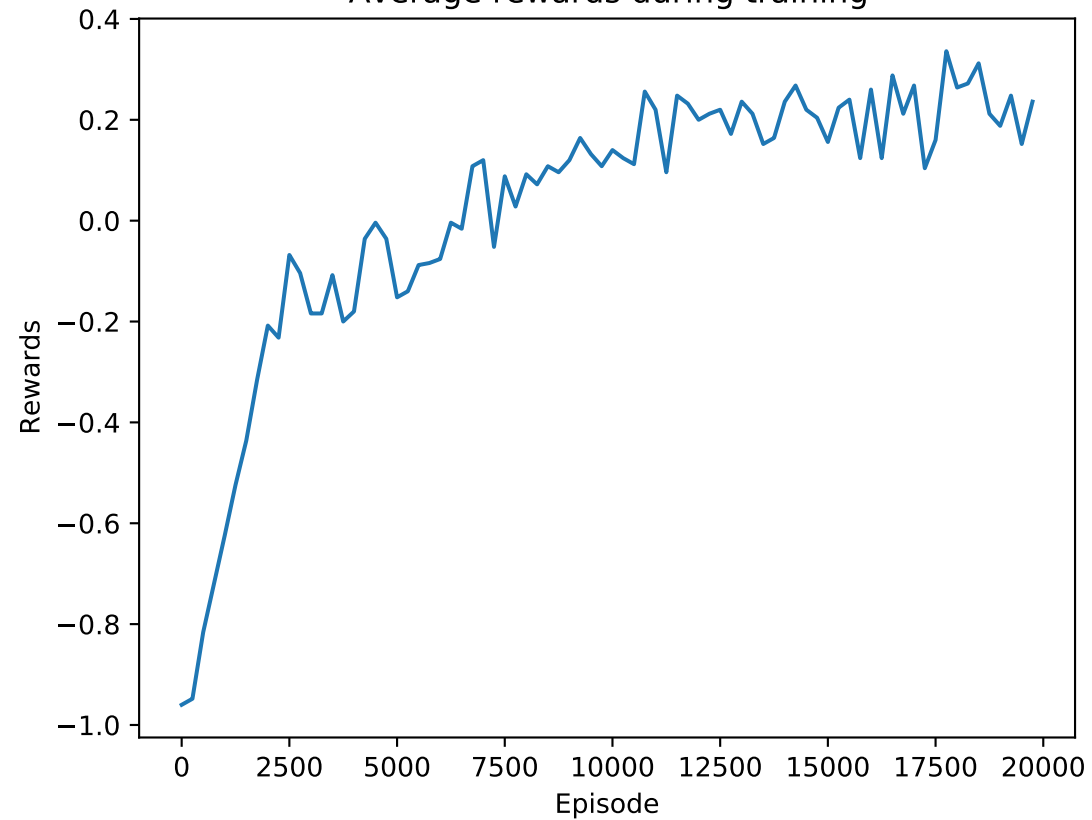
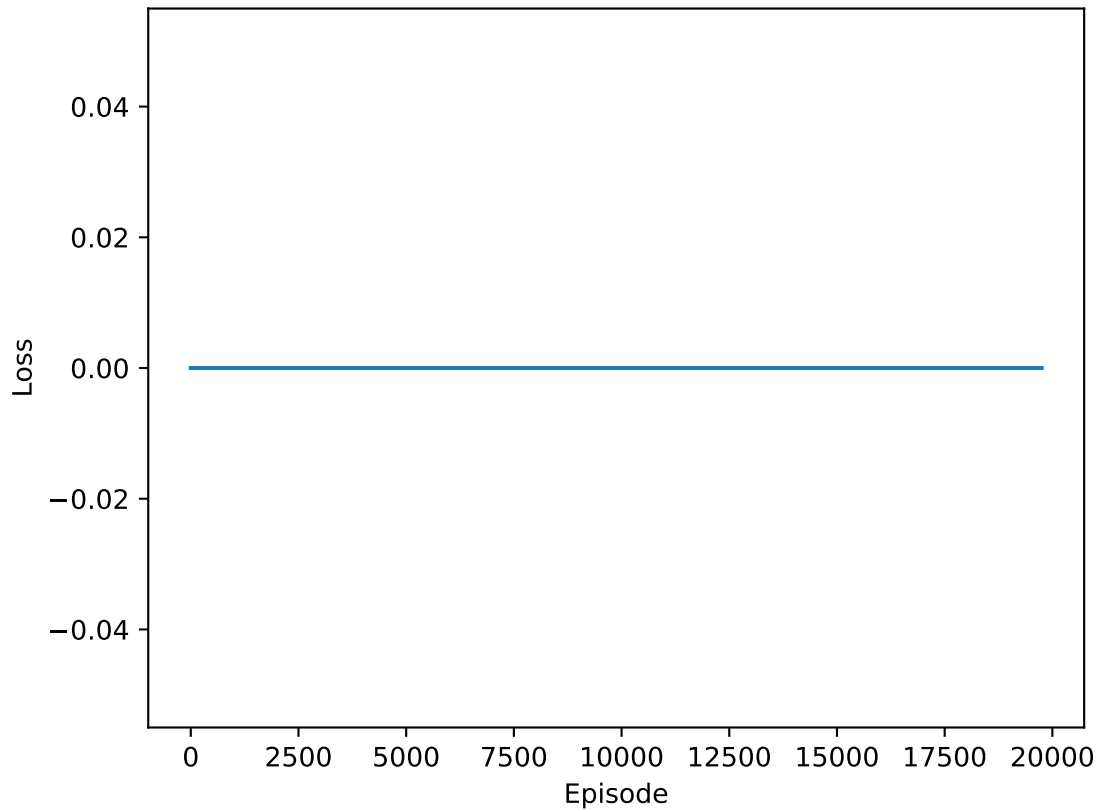


Average rewards during training

 $\epsilon = 0.1$

Average loss during training

 $\epsilon = 0.1$