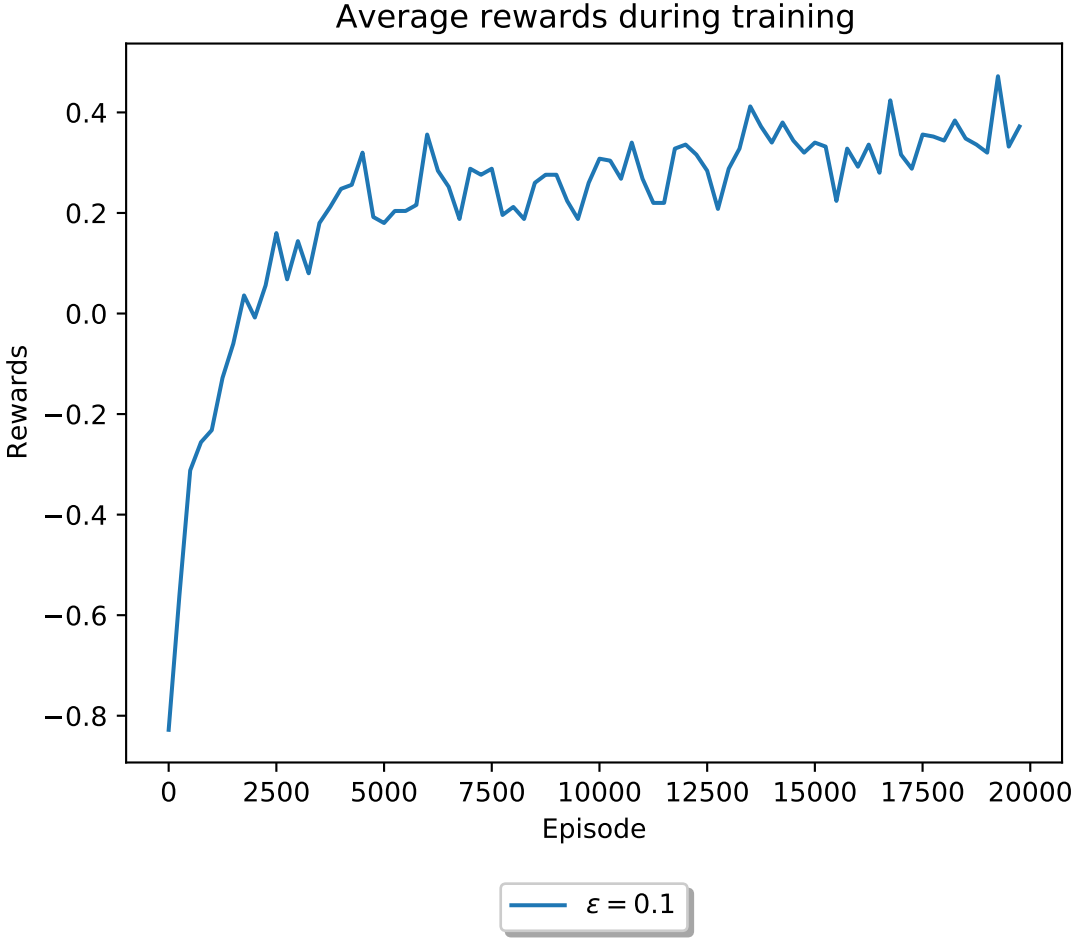
Average rewards during training
Rewards
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
0
2500
5000
7500
10000
12500
15000
17500
20000
Episode
$\varepsilon = 0.1$


Average loss during training
Loss
0.060
0.055
0.050
0.045
0.040
0.035
0.030
0
2500
5000
7500
10000
12500
15000
17500
20000
Episode
$\varepsilon = 0.1$