

Zachary Pearson

D439H842

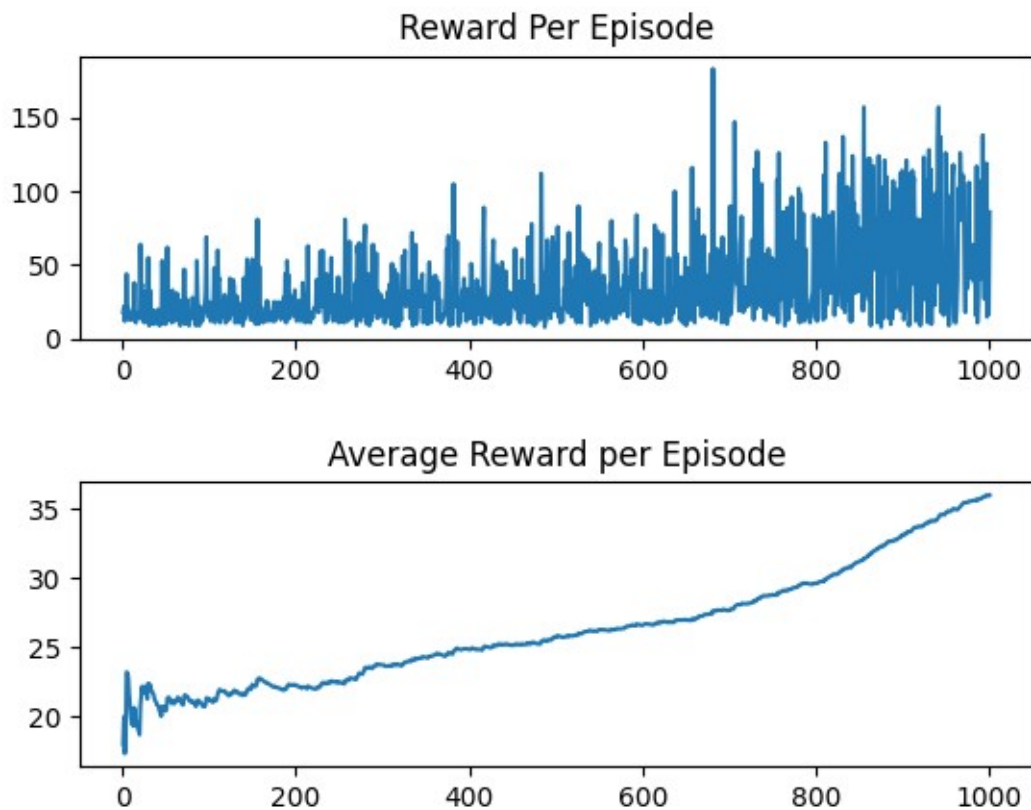
CS898

Deep Learning

Assignment 5

q-learning

**Architecture:** I used a q-table implemented as a dictionary. I rounded every input to a single decimal place and focused only on the pole position and velocity.



Random q-table values:

**Before Training**

state: (0.0 0.9)  
action: [0. 0.]

----

state: (0.8 0.1)  
action: [0. 0.]

----

state: (0.6 0.3)  
action: [0. 0.]

----

state: (0.4 0.8)  
action: [0. 0.]

----

state: (0.9 0.0)  
action: [0. 0.]

**After Training**

state: (-0.0, 0.3)  
action: [0.50203905 1.94572771]

----

state: (-0.0, 0.0)  
action: [2.05550284 0.66692894]

----

state: (-0.0, 0.6)  
action: [0.13373252 0.44655833]

----

state: (-0.0, 0.9)  
action: [0.08439671 0.04054707]