

# EE 526 – Final Project

## Double Deep Q Learning

### using the Acrobot-v1 Environment



Submitted by:

Anjana Deva Prasad

Email: [anjana@iastate.edu](mailto:anjana@iastate.edu)

## Description of the Problem

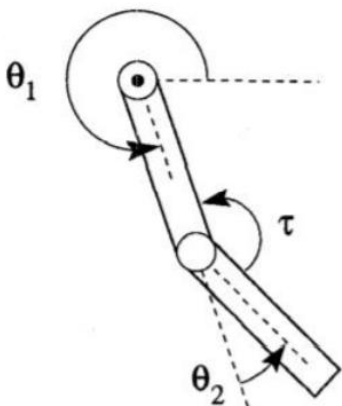
The project problem was to design a reinforcement learning algorithm using double Q-learning algorithm with function approximation. This was supposed to be implemented using two multi-layer neural networks which are to act as Q function approximators. The environment required was the Acrobot-v1 taken from the OpenAI Gym package.

## What is OpenAI Gym?

Gym is a toolkit for developing and comparing reinforcement learning algorithms. It makes no assumptions about the structure of the agent and is compatible with a wide range of numerical libraries. Gym comes with a diverse suite of environments that range from easy to difficult and involve different kinds of data. One set of these environments is the classic control theory problems environment. Acrobot-v1 is one problem in this environment space. It is beneficial to use gym because the existing open-source collections of RL environments don't have enough variety, and they are often difficult to even set up and use. Gym also offers a standardization for the various environments and makes it easier to compare results.

## What is Acrobot-v1 ?

The acrobot-v1 is a system includes two joints and two links, where the joint between the two links is actuated. Initially, the links are hanging downwards, and the goal is to swing the end of the lower link up to a given height.



This system is analogous to a gymnast swinging on a high bar. Such problems are extremely challenging problems of non-linear control because they cannot be commanded to follow a certain trajectory.

The acrobot-v1 system is characterized by the following features:

States: The Acrobot system is defined by 6 state values which are joint angular velocities ( $\omega_1$  at free swinging shoulder and  $\omega_2$  at the actuated link) and two rotational joint angles defined by:  $\sin(\theta_1)$ ,  $\cos(\theta_1)$ ,  $\sin(\theta_2)$  and  $\cos(\theta_2)$

Action: The system has three torques which act as the actions, +1, 0 or -1 torque on the actuated elbow links.

Rewards: The reward system is designed as -1 if it is not a terminal state and 0 if terminal.

Goal: The goal is to swing the endpoint (the feet) above the bar by an amount equal to one of the links.

# Double Deep Q Learning- Pseudo Code

In Double Deep Q Learning, we have a model  $Q$  and a target model  $Q'$  declared as two independent models. We use the  $Q'$  for action selection and  $Q$  for action evaluation.

## Pseudo Code:

Declare two models,  $Q^A$  and  $Q^B$  with their neural-network structure clearly defined.

for each episode do:

    Repeat for each step:

        Choose an initial action randomly based on  $Q^A(s, \cdot)$  and  $Q^B(s, \cdot)$ , observe next state ( $s'$ ), reward, and if terminal or not.

        if choosing to Update A then: (This is what I did)

            Define  $a^* = \text{argmax}(Q^A(s, a))$

            Predict for  $Q^B(s', a)$

            Update  $Q^A(s, a)$  as  $Q^A(s, a) + \text{reward} + \gamma Q^B(s', a^*)$

        if choosing to Update B then:

            Define  $b^* = \text{argmax}(Q^B(s, a))$

            Predict for  $Q^A(s', a)$

            Update  $Q^B(s, a)$  as  $Q^B(s, a) + \text{reward} + \gamma Q^A(s', b^*)$

    end if

Update state with next state

Until end

end of total episodes

# Performance of the Algorithm

## Code:

For some of the logic of implementing a DDQ Network, I referred to the following link:

<https://github.com/datalearningpr/ReinforcementLearning-OpenAI-Gym/blob/master/CartPole-v0%20-%20DDQN/DDQN.py>

## Neural Network, Performance and Results:

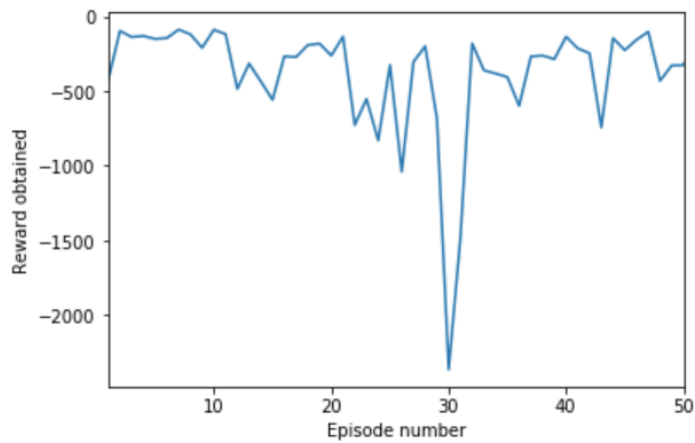
### Result 1:

Initially, I tried DDQN with a Neural network of 3 layers, with no\_of\_neurons\_in\_layer1=16 and no\_of\_neurons\_in\_layer2=16. I tried this for the number of episodes=50,100,200.

For the number of episodes=50, even though the convergence of no of steps to terminal state for each episode were not very good, most of them were still under 500 steps. This took 30 minutes to train.

The hyper-parameters are:

No of episodes=50  
Learning rate (alpha)=0.001  
Layer1=16, Layer2=16, Output=3  
Memory size=5000  
Batch\_size = 32  
Gamma=1  
Max\_exploration = 1.0  
Min\_exploration = 0.02  
Exploration\_decay = 0.995



Maximum best total reward obtained was -89.0

---

For episodes=100 and 200, after a certain point of converging for the no of steps, the no of steps started increasing to values in the 1000s occasionally. The mean of the rewards for all the episodes was still below 500. The train time was 1 hr.

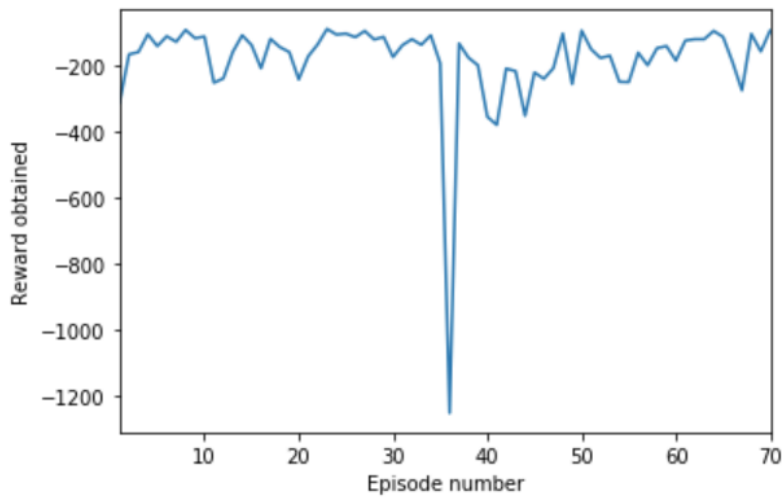
The best reward I got for 16,16 with no of episodes=50 was -89.

## Result 2:

I kept all other hyperparameters the same, except no of episodes=70 and no of neurons in layer 1=64, no of neurons in layer 2=64. The model seemed to converge to rewards of mostly below 200 as is shown in the plot below. The best reward obtained was -89. Train time was 25 minutes.

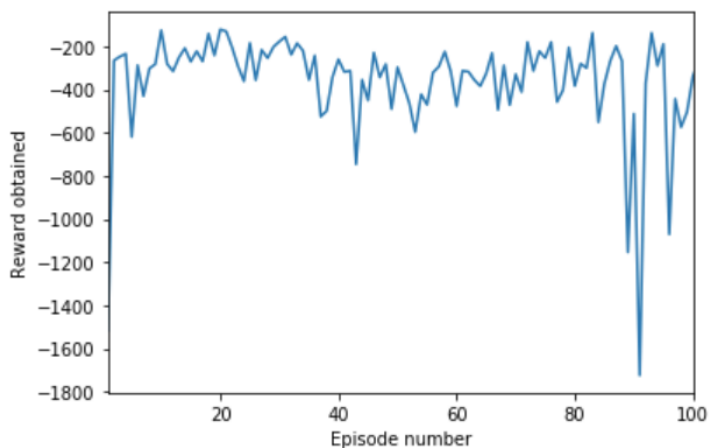
The hyper-parameters are:

No of episodes=70  
 Learning rate (alpha)=0.001  
 Layer1=64, Layer2=64, Output=3  
 Memory size=5000  
 Batch\_size = 32  
 Gamma=1  
 Max\_exploration = 1.0  
 Min\_exploration = 0.02  
 Exploration\_decay = 0.995



Best total reward obtained was -89.0

For the same number of neurons, when I tried for no\_of\_episodes=100, the performance got worse after around 80 episodes. It took one hour to train. Train time was 1.5hrs. The plot is as shown below:



Maximum best total reward obtained was -121.0

## Other Results:

For  $n_1=8$ ,  $n_2=8$ , The execution time was too long, and I therefore did not plot the graph. I am attaching some of the values obtained for reference.

The episode is: 1,	Exploration is: 0.02,	No of steps is: 834	,Total Reward is: -833.0
The episode is: 2,	Exploration is: 0.02,	No of steps is: 289	,Total Reward is: -288.0
The episode is: 3,	Exploration is: 0.02,	No of steps is: 399	,Total Reward is: -398.0
The episode is: 4,	Exploration is: 0.02,	No of steps is: 503	,Total Reward is: -502.0
The episode is: 5,	Exploration is: 0.02,	No of steps is: 259	,Total Reward is: -258.0
The episode is: 6,	Exploration is: 0.02,	No of steps is: 218	,Total Reward is: -217.0
The episode is: 7,	Exploration is: 0.02,	No of steps is: 471	,Total Reward is: -470.0
The episode is: 8,	Exploration is: 0.02,	No of steps is: 557	,Total Reward is: -556.0
The episode is: 9,	Exploration is: 0.02,	No of steps is: 800	,Total Reward is: -799.0
The episode is: 10,	Exploration is: 0.02,	No of steps is: 681	,Total Reward is: -680.0
The episode is: 11,	Exploration is: 0.02,	No of steps is: 355	,Total Reward is: -354.0
The episode is: 12,	Exploration is: 0.02,	No of steps is: 513	,Total Reward is: -512.0
The episode is: 13,	Exploration is: 0.02,	No of steps is: 411	,Total Reward is: -410.0
The episode is: 14,	Exploration is: 0.02,	No of steps is: 469	,Total Reward is: -468.0
The episode is: 15,	Exploration is: 0.02,	No of steps is: 548	,Total Reward is: -547.0
The episode is: 16,	Exploration is: 0.02,	No of steps is: 353	,Total Reward is: -352.0
The episode is: 17,	Exploration is: 0.02,	No of steps is: 456	,Total Reward is: -455.0
The episode is: 18,	Exploration is: 0.02,	No of steps is: 309	,Total Reward is: -308.0
The episode is: 19,	Exploration is: 0.02,	No of steps is: 376	,Total Reward is: -375.0
The episode is: 20,	Exploration is: 0.02,	No of steps is: 1640	,Total Reward is: -1639.0
The episode is: 21,	Exploration is: 0.02,	No of steps is: 208	,Total Reward is: -207.0
The episode is: 22,	Exploration is: 0.02,	No of steps is: 471	,Total Reward is: -470.0
The episode is: 23,	Exploration is: 0.02,	No of steps is: 171	,Total Reward is: -170.0
The episode is: 24,	Exploration is: 0.02,	No of steps is: 301	,Total Reward is: -300.0
The episode is: 25,	Exploration is: 0.02,	No of steps is: 129	,Total Reward is: -128.0
The episode is: 26,	Exploration is: 0.02,	No of steps is: 377	,Total Reward is: -376.0
The episode is: 27,	Exploration is: 0.02,	No of steps is: 1181	,Total Reward is: -1180.0
The episode is: 28,	Exploration is: 0.02,	No of steps is: 393	,Total Reward is: -392.0
The episode is: 29,	Exploration is: 0.02,	No of steps is: 439	,Total Reward is: -438.0
The episode is: 30,	Exploration is: 0.02,	No of steps is: 289	,Total Reward is: -288.0
The episode is: 31,	Exploration is: 0.02,	No of steps is: 2119	,Total Reward is: -2118.0
The episode is: 32,	Exploration is: 0.02,	No of steps is: 1182	,Total Reward is: -1181.0
The episode is: 33,	Exploration is: 0.02,	No of steps is: 436	,Total Reward is: -435.0
The episode is: 34,	Exploration is: 0.02,	No of steps is: 460	,Total Reward is: -459.0
The episode is: 35,	Exploration is: 0.02,	No of steps is: 151	,Total Reward is: -150.0
The episode is: 36,	Exploration is: 0.02,	No of steps is: 1183	,Total Reward is: -1182.0
The episode is: 37,	Exploration is: 0.02,	No of steps is: 549	,Total Reward is: -548.0
The episode is: 38,	Exploration is: 0.02,	No of steps is: 555	,Total Reward is: -554.0
The episode is: 39,	Exploration is: 0.02,	No of steps is: 348	,Total Reward is: -347.0
The episode is: 40,	Exploration is: 0.02,	No of steps is: 277	,Total Reward is: -276.0
The episode is: 41,	Exploration is: 0.02,	No of steps is: 317	,Total Reward is: -316.0
The episode is: 42,	Exploration is: 0.02,	No of steps is: 298	,Total Reward is: -297.0
The episode is: 43,	Exploration is: 0.02,	No of steps is: 218	,Total Reward is: -217.0
The episode is: 44,	Exploration is: 0.02,	No of steps is: 393	,Total Reward is: -392.0
The episode is: 45,	Exploration is: 0.02,	No of steps is: 2258	,Total Reward is: -2257.0
The episode is: 46,	Exploration is: 0.02,	No of steps is: 1119	,Total Reward is: -1118.0
The episode is: 47,	Exploration is: 0.02,	No of steps is: 411	,Total Reward is: -410.0
The episode is: 48,	Exploration is: 0.02,	No of steps is: 3027	,Total Reward is: -3026.0
The episode is: 49,	Exploration is: 0.02,	No of steps is: 1182	,Total Reward is: -1181.0
The episode is: 50,	Exploration is: 0.02,	No of steps is: 1972	,Total Reward is: -1971.0
The episode is: 51,	Exploration is: 0.02,	No of steps is: 3519	,Total Reward is: -3518.0
The episode is: 52,	Exploration is: 0.02,	No of steps is: 196	,Total Reward is: -195.0
The episode is: 53,	Exploration is: 0.02,	No of steps is: 4196	,Total Reward is: -4195.0
The episode is: 54,	Exploration is: 0.02,	No of steps is: 1716	,Total Reward is: -1715.0
The episode is: 55,	Exploration is: 0.02,	No of steps is: 194	,Total Reward is: -193.0
The episode is: 56,	Exploration is: 0.02,	No of steps is: 161	,Total Reward is: -160.0



The episode is: 57,	Exploration is: 0.02,	No of steps is: 97	,Total Reward is: -96.0
The episode is: 58,	Exploration is: 0.02,	No of steps is: 133	,Total Reward is: -132.0
The episode is: 59,	Exploration is: 0.02,	No of steps is: 118	,Total Reward is: -117.0
The episode is: 60,	Exploration is: 0.02,	No of steps is: 114	,Total Reward is: -113.0
The episode is: 61,	Exploration is: 0.02,	No of steps is: 101	,Total Reward is: -100.0
The episode is: 62,	Exploration is: 0.02,	No of steps is: 107	,Total Reward is: -106.0
The episode is: 63,	Exploration is: 0.02,	No of steps is: 126	,Total Reward is: -125.0
The episode is: 64,	Exploration is: 0.02,	No of steps is: 434	,Total Reward is: -433.0
The episode is: 65,	Exploration is: 0.02,	No of steps is: 1045	,Total Reward is: -1044.0
The episode is: 66,	Exploration is: 0.02,	No of steps is: 809	,Total Reward is: -808.0
The episode is: 67,	Exploration is: 0.02,	No of steps is: 373	,Total Reward is: -372.0
The episode is: 68,	Exploration is: 0.02,	No of steps is: 317	,Total Reward is: -316.0
The episode is: 69,	Exploration is: 0.02,	No of steps is: 469	,Total Reward is: -468.0
The episode is: 70,	Exploration is: 0.02,	No of steps is: 1046	,Total Reward is: -1045.0
The episode is: 71,	Exploration is: 0.02,	No of steps is: 5261	,Total Reward is: -5260.0
The episode is: 72,	Exploration is: 0.02,	No of steps is: 1468	,Total Reward is: -1467.0
The episode is: 73,	Exploration is: 0.02,	No of steps is: 359	,Total Reward is: -358.0
The episode is: 74,	Exploration is: 0.02,	No of steps is: 543	,Total Reward is: -542.0
The episode is: 75,	Exploration is: 0.02,	No of steps is: 296	,Total Reward is: -295.0
The episode is: 76,	Exploration is: 0.02,	No of steps is: 206	,Total Reward is: -205.0
The episode is: 77,	Exploration is: 0.02,	No of steps is: 215	,Total Reward is: -214.0
The episode is: 78,	Exploration is: 0.02,	No of steps is: 166	,Total Reward is: -165.0
The episode is: 79,	Exploration is: 0.02,	No of steps is: 867	,Total Reward is: -866.0
The episode is: 80,	Exploration is: 0.02,	No of steps is: 322	,Total Reward is: -321.0
The episode is: 81,	Exploration is: 0.02,	No of steps is: 1288	,Total Reward is: -1287.0
The episode is: 82,	Exploration is: 0.02,	No of steps is: 267	,Total Reward is: -266.0
The episode is: 83,	Exploration is: 0.02,	No of steps is: 2283	,Total Reward is: -2282.0
The episode is: 84,	Exploration is: 0.02,	No of steps is: 546	,Total Reward is: -545.0
The episode is: 85,	Exploration is: 0.02,	No of steps is: 684	,Total Reward is: -683.0
The episode is: 86,	Exploration is: 0.02,	No of steps is: 436	,Total Reward is: -435.0
The episode is: 87,	Exploration is: 0.02,	No of steps is: 563	,Total Reward is: -562.0
The episode is: 88,	Exploration is: 0.02,	No of steps is: 565	,Total Reward is: -564.0
The episode is: 89,	Exploration is: 0.02,	No of steps is: 647	,Total Reward is: -646.0
The episode is: 90,	Exploration is: 0.02,	No of steps is: 306	,Total Reward is: -305.0
The episode is: 91,	Exploration is: 0.02,	No of steps is: 196	,Total Reward is: -195.0
The episode is: 92,	Exploration is: 0.02,	No of steps is: 358	,Total Reward is: -357.0
The episode is: 93,	Exploration is: 0.02,	No of steps is: 167	,Total Reward is: -166.0
The episode is: 94,	Exploration is: 0.02,	No of steps is: 545	,Total Reward is: -544.0
The episode is: 95,	Exploration is: 0.02,	No of steps is: 182	,Total Reward is: -181.0
The episode is: 96,	Exploration is: 0.02,	No of steps is: 196	,Total Reward is: -195.0
The episode is: 97,	Exploration is: 0.02,	No of steps is: 141	,Total Reward is: -140.0
The episode is: 98,	Exploration is: 0.02,	No of steps is: 128	,Total Reward is: -127.0
The episode is: 99,	Exploration is: 0.02,	No of steps is: 201	,Total Reward is: -200.0
The episode is: 100,	Exploration is: 0.02,	No of steps is: 346	,Total Reward is: -345.0
The episode is: 101,	Exploration is: 0.02,	No of steps is: 192	,Total Reward is: -191.0
The episode is: 102,	Exploration is: 0.02,	No of steps is: 217	,Total Reward is: -216.0
The episode is: 103,	Exploration is: 0.02,	No of steps is: 115	,Total Reward is: -114.0
The episode is: 104,	Exploration is: 0.02,	No of steps is: 333	,Total Reward is: -332.0
The episode is: 105,	Exploration is: 0.02,	No of steps is: 372	,Total Reward is: -371.0
The episode is: 106,	Exploration is: 0.02,	No of steps is: 219	,Total Reward is: -218.0
The episode is: 107,	Exploration is: 0.02,	No of steps is: 121	,Total Reward is: -120.0
The episode is: 108,	Exploration is: 0.02,	No of steps is: 123	,Total Reward is: -122.0
The episode is: 109,	Exploration is: 0.02,	No of steps is: 132	,Total Reward is: -131.0
The episode is: 110,	Exploration is: 0.02,	No of steps is: 137	,Total Reward is: -136.0
The episode is: 111,	Exploration is: 0.02,	No of steps is: 200	,Total Reward is: -199.0
The episode is: 112,	Exploration is: 0.02,	No of steps is: 155	,Total Reward is: -154.0
The episode is: 113,	Exploration is: 0.02,	No of steps is: 316	,Total Reward is: -315.0
The episode is: 114,	Exploration is: 0.02,	No of steps is: 210	,Total Reward is: -209.0

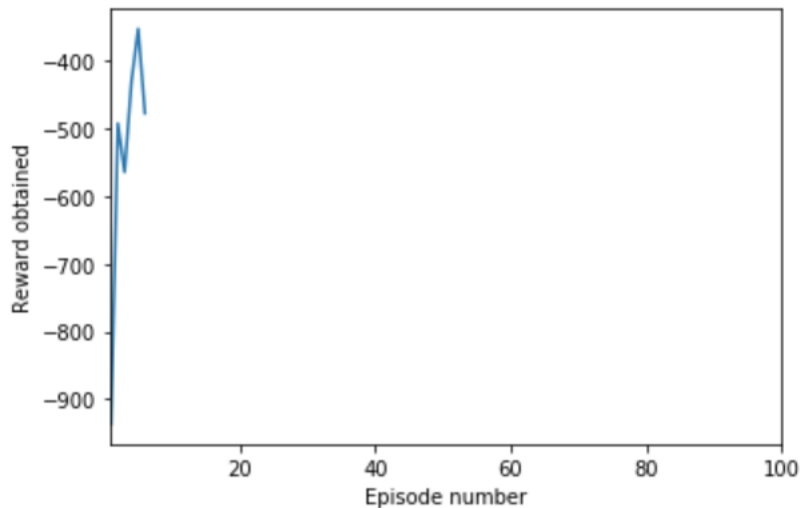
The episode is: 115,	Exploration is: 0.02,	No of steps is: 367	,Total Reward is: -366.0
The episode is: 116,	Exploration is: 0.02,	No of steps is: 315	,Total Reward is: -314.0
The episode is: 117,	Exploration is: 0.02,	No of steps is: 202	,Total Reward is: -201.0
The episode is: 118,	Exploration is: 0.02,	No of steps is: 193	,Total Reward is: -192.0
The episode is: 119,	Exploration is: 0.02,	No of steps is: 264	,Total Reward is: -263.0
The episode is: 120,	Exploration is: 0.02,	No of steps is: 115	,Total Reward is: -114.0
The episode is: 121,	Exploration is: 0.02,	No of steps is: 274	,Total Reward is: -273.0
The episode is: 122,	Exploration is: 0.02,	No of steps is: 256	,Total Reward is: -255.0
The episode is: 123,	Exploration is: 0.02,	No of steps is: 196	,Total Reward is: -195.0
The episode is: 124,	Exploration is: 0.02,	No of steps is: 243	,Total Reward is: -242.0
The episode is: 125,	Exploration is: 0.02,	No of steps is: 206	,Total Reward is: -205.0
The episode is: 126,	Exploration is: 0.02,	No of steps is: 163	,Total Reward is: -162.0
The episode is: 127,	Exploration is: 0.02,	No of steps is: 206	,Total Reward is: -205.0
The episode is: 128,	Exploration is: 0.02,	No of steps is: 196	,Total Reward is: -195.0
The episode is: 129,	Exploration is: 0.02,	No of steps is: 261	,Total Reward is: -260.0
The episode is: 130,	Exploration is: 0.02,	No of steps is: 114	,Total Reward is: -113.0
The episode is: 131,	Exploration is: 0.02,	No of steps is: 196	,Total Reward is: -195.0
The episode is: 132,	Exploration is: 0.02,	No of steps is: 191	,Total Reward is: -190.0
The episode is: 133,	Exploration is: 0.02,	No of steps is: 468	,Total Reward is: -467.0
The episode is: 134,	Exploration is: 0.02,	No of steps is: 201	,Total Reward is: -200.0
The episode is: 135,	Exploration is: 0.02,	No of steps is: 403	,Total Reward is: -402.0
The episode is: 136,	Exploration is: 0.02,	No of steps is: 386	,Total Reward is: -385.0
The episode is: 137,	Exploration is: 0.02,	No of steps is: 172	,Total Reward is: -171.0
The episode is: 138,	Exploration is: 0.02,	No of steps is: 273	,Total Reward is: -272.0
The episode is: 139,	Exploration is: 0.02,	No of steps is: 235	,Total Reward is: -234.0
The episode is: 140,	Exploration is: 0.02,	No of steps is: 332	,Total Reward is: -331.0
The episode is: 141,	Exploration is: 0.02,	No of steps is: 144	,Total Reward is: -143.0
The episode is: 142,	Exploration is: 0.02,	No of steps is: 168	,Total Reward is: -167.0
The episode is: 143,	Exploration is: 0.02,	No of steps is: 136	,Total Reward is: -135.0
The episode is: 144,	Exploration is: 0.02,	No of steps is: 266	,Total Reward is: -265.0
The episode is: 145,	Exploration is: 0.02,	No of steps is: 229	,Total Reward is: -228.0
The episode is: 146,	Exploration is: 0.02,	No of steps is: 427	,Total Reward is: -426.0
The episode is: 147,	Exploration is: 0.02,	No of steps is: 281	,Total Reward is: -280.0
The episode is: 148,	Exploration is: 0.02,	No of steps is: 246	,Total Reward is: -245.0
The episode is: 149,	Exploration is: 0.02,	No of steps is: 203	,Total Reward is: -202.0
The episode is: 150,	Exploration is: 0.02,	No of steps is: 235	,Total Reward is: -234.0
The episode is: 151,	Exploration is: 0.02,	No of steps is: 140	,Total Reward is: -139.0
The episode is: 152,	Exploration is: 0.02,	No of steps is: 260	,Total Reward is: -259.0
The episode is: 153,	Exploration is: 0.02,	No of steps is: 100	,Total Reward is: -99.0
The episode is: 154,	Exploration is: 0.02,	No of steps is: 300	,Total Reward is: -299.0
The episode is: 155,	Exploration is: 0.02,	No of steps is: 126	,Total Reward is: -125.0
The episode is: 156,	Exploration is: 0.02,	No of steps is: 197	,Total Reward is: -196.0
The episode is: 157,	Exploration is: 0.02,	No of steps is: 96	,Total Reward is: -95.0
The episode is: 158,	Exploration is: 0.02,	No of steps is: 276	,Total Reward is: -275.0
The episode is: 159,	Exploration is: 0.02,	No of steps is: 423	,Total Reward is: -422.0
The episode is: 160,	Exploration is: 0.02,	No of steps is: 282	,Total Reward is: -281.0
The episode is: 161,	Exploration is: 0.02,	No of steps is: 494	,Total Reward is: -493.0
The episode is: 162,	Exploration is: 0.02,	No of steps is: 189	,Total Reward is: -188.0
The episode is: 163,	Exploration is: 0.02,	No of steps is: 353	,Total Reward is: -352.0
The episode is: 164,	Exploration is: 0.02,	No of steps is: 278	,Total Reward is: -277.0
The episode is: 165,	Exploration is: 0.02,	No of steps is: 467	,Total Reward is: -466.0
The episode is: 166,	Exploration is: 0.02,	No of steps is: 151	,Total Reward is: -150.0
The episode is: 167,	Exploration is: 0.02,	No of steps is: 455	,Total Reward is: -454.0
The episode is: 168,	Exploration is: 0.02,	No of steps is: 839	,Total Reward is: -838.0
The episode is: 169,	Exploration is: 0.02,	No of steps is: 185	,Total Reward is: -184.0
The episode is: 170,	Exploration is: 0.02,	No of steps is: 444	,Total Reward is: -443.0
The episode is: 171,	Exploration is: 0.02,	No of steps is: 217	,Total Reward is: -216.0
The episode is: 172,	Exploration is: 0.02,	No of steps is: 590	,Total Reward is: -589.0

The episode is: 173,	Exploration is: 0.02,	No of steps is: 265	,Total Reward is: -264.0
The episode is: 174,	Exploration is: 0.02,	No of steps is: 3590	,Total Reward is: -3589.0
The episode is: 175,	Exploration is: 0.02,	No of steps is: 202	,Total Reward is: -201.0
The episode is: 176,	Exploration is: 0.02,	No of steps is: 329	,Total Reward is: -328.0
The episode is: 177,	Exploration is: 0.02,	No of steps is: 184	,Total Reward is: -183.0
The episode is: 178,	Exploration is: 0.02,	No of steps is: 305	,Total Reward is: -304.0
The episode is: 179,	Exploration is: 0.02,	No of steps is: 300	,Total Reward is: -299.0
The episode is: 180,	Exploration is: 0.02,	No of steps is: 135	,Total Reward is: -134.0
The episode is: 181,	Exploration is: 0.02,	No of steps is: 233	,Total Reward is: -232.0
The episode is: 182,	Exploration is: 0.02,	No of steps is: 345	,Total Reward is: -344.0
The episode is: 183,	Exploration is: 0.02,	No of steps is: 297	,Total Reward is: -296.0
The episode is: 184,	Exploration is: 0.02,	No of steps is: 182	,Total Reward is: -181.0
The episode is: 185,	Exploration is: 0.02,	No of steps is: 187	,Total Reward is: -186.0
The episode is: 186,	Exploration is: 0.02,	No of steps is: 295	,Total Reward is: -294.0
The episode is: 187,	Exploration is: 0.02,	No of steps is: 406	,Total Reward is: -405.0
The episode is: 188,	Exploration is: 0.02,	No of steps is: 186	,Total Reward is: -185.0
The episode is: 189,	Exploration is: 0.02,	No of steps is: 457	,Total Reward is: -456.0
The episode is: 190,	Exploration is: 0.02,	No of steps is: 846	,Total Reward is: -845.0
The episode is: 191,	Exploration is: 0.02,	No of steps is: 439	,Total Reward is: -438.0
The episode is: 192,	Exploration is: 0.02,	No of steps is: 816	,Total Reward is: -815.0
The episode is: 193,	Exploration is: 0.02,	No of steps is: 518	,Total Reward is: -517.0
The episode is: 194,	Exploration is: 0.02,	No of steps is: 534	,Total Reward is: -533.0
The episode is: 195,	Exploration is: 0.02,	No of steps is: 229	,Total Reward is: -228.0
The episode is: 196,	Exploration is: 0.02,	No of steps is: 367	,Total Reward is: -366.0
The episode is: 197,	Exploration is: 0.02,	No of steps is: 231	,Total Reward is: -230.0
The episode is: 198,	Exploration is: 0.02,	No of steps is: 188	,Total Reward is: -187.0
The episode is: 199,	Exploration is: 0.02,	No of steps is: 167	,Total Reward is: -166.0
The episode is: 200,	Exploration is: 0.02,	No of steps is: 202	,Total Reward is: -201.0
The episode is: 201,	Exploration is: 0.02,	No of steps is: 238	,Total Reward is: -237.0
The episode is: 202,	Exploration is: 0.02,	No of steps is: 219	,Total Reward is: -218.0
The episode is: 203,	Exploration is: 0.02,	No of steps is: 329	,Total Reward is: -328.0
The episode is: 204,	Exploration is: 0.02,	No of steps is: 378	,Total Reward is: -377.0
The episode is: 205,	Exploration is: 0.02,	No of steps is: 368	,Total Reward is: -367.0
The episode is: 206,	Exploration is: 0.02,	No of steps is: 255	,Total Reward is: -254.0
The episode is: 207,	Exploration is: 0.02,	No of steps is: 279	,Total Reward is: -278.0
The episode is: 208,	Exploration is: 0.02,	No of steps is: 489	,Total Reward is: -488.0
The episode is: 209,	Exploration is: 0.02,	No of steps is: 402	,Total Reward is: -401.0
The episode is: 210,	Exploration is: 0.02,	No of steps is: 267	,Total Reward is: -266.0
The episode is: 211,	Exploration is: 0.02,	No of steps is: 424	,Total Reward is: -423.0
The episode is: 212,	Exploration is: 0.02,	No of steps is: 346	,Total Reward is: -345.0
The episode is: 213,	Exploration is: 0.02,	No of steps is: 394	,Total Reward is: -393.0
The episode is: 214,	Exploration is: 0.02,	No of steps is: 402	,Total Reward is: -401.0
The episode is: 215,	Exploration is: 0.02,	No of steps is: 627	,Total Reward is: -626.0
The episode is: 216,	Exploration is: 0.02,	No of steps is: 211	,Total Reward is: -210.0

For  $n_1=128$ ,  $n_2=128$ , The training was too long to carry out and the performance was also poor.

For reference, attaching a video of  $n_1=16$ ,  $n_2=16$ , and  $no\_of\_episodes=5$ .

The training time was 2:30 mins.



## Key Takeaways and Lessons learned

1. Continuous State problems, Unresolved environments and Non-linear Control problems:  
Since Acrobot-v1 fits all these three categories, I have realized that for such problems, it is not practical to store state-action pairs in a table and determine the values for every possible state-action pair. This is where, using neural networks as function approximators has made a significant difference in training these kinds of environments.
2. Reward: In the Acrobot-v1 environment, the reward criteria is defined on Github as: -1 for non-terminal state and 0 otherwise. The penalty for going into a non-terminal state is not significant, and the reward for reaching a terminal state is also not high. In such problems, the performance of the algorithm might therefore depend on how well the reward system has been defined.
3. While performance for my Algorithm, did increase when I increased the number of neurons till a certain value, after a certain number, even for large number of neurons, the performance started to deteriorate. The time required to learn also increased as the number of neurons in each layer was increased. Therefore performance of such a network, would require a compromise between number of neurons in each layer and the number of episodes.
4. Though we model the network, to make the arms cross the required level in lesser time, since this system has continuous states, the no of steps required to reach the terminal state may not always converge as the number of episodes increase. There is still a possibility for

oscillations about a mean value, or for the mean value for the number of steps to increase as the number of episodes increases.

5. Just because the number of neurons are less, does not mean the training time will be less. For cases where each layer had only 8 neurons, it still took a lot of time to execute because the performance was poor and the number of steps required to reach terminal state were very high.

## Improvements

1. Prioritized Experience Replay: The fundamental idea is that some experiences might be more important than the others for our training, but their occurrences are very infrequent. In this approach we try to give priority to the experiences where there is a big difference between our prediction and the TD target because we can learn a lot from those experiences. This would allow us to reach the terminal state faster.
2. Clipped Double Q-learning: In Clipped Double Q-learning, the idea is to have two independent estimates of the true Q value. Here, for computing the update targets, the minimum of the two next-state action values produced by our two Q networks is taken; When the Q estimate from one is greater than the other, it is reduced to minimum to avoid overestimation.
3. For this problem, the reward system can be designed, to give higher values for reaching a terminal state, and maybe stricter penalties for not reaching a terminal state.

## References:

<https://gym.openai.com/docs/>

<https://medium.com/@ameetsd97/deep-double-q-learning-why-you-should-use-it-bedf660d5295>

<https://www.freecodecamp.org/news/improvements-in-deep-q-learning-dueling-double-dqn-prioritized-experience-replay-and-fixed-58b130cc5682/>

<https://towardsdatascience.com/double-deep-q-networks-905dd8325412>

<https://arxiv.org/pdf/1509.06461.pdf>

<https://ieeexplore.ieee.org/document/606789>

<https://github.com/datalearningpr/ReinforcementLearning-OpenAI-Gym/blob/master/CartPole-v0%20-%20DDQN/DDQN.py>