Week 9 HW: We will create a DQN agent using Keras to master the CartPole-v0 environment a
nd take several hundred episodes to eventually balance the pole

**Configure appropriate CartPole-v0 Environment Variables to benchmark the ability of rein
forcement learning agent**

env= <TimeLimit<CartPoleEnv<CartPole-v0>>>
nb_actions= 2

First build simple 3 hidden layers neural network model with 16 neurons each

Model Summary

```
Model: "sequential_2"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
flatten_2 (Flatten)          (None, 4)                 0
_________________________________________________________________
dense_5 (Dense)              (None, 16)                80
_________________________________________________________________
activation_4 (Activation)    (None, 16)                0
_________________________________________________________________
dense_6 (Dense)              (None, 16)                272
_________________________________________________________________
activation_5 (Activation)    (None, 16)                0
_________________________________________________________________
dense_7 (Dense)              (None, 16)                272
_________________________________________________________________
activation_6 (Activation)    (None, 16)                0
_________________________________________________________________
dense_8 (Dense)              (None, 2)                 34
=================================================================
Total params: 658
Trainable params: 658
Non-trainable params: 0
_________________________________________________________________
None
```

Model Layer Output Shape

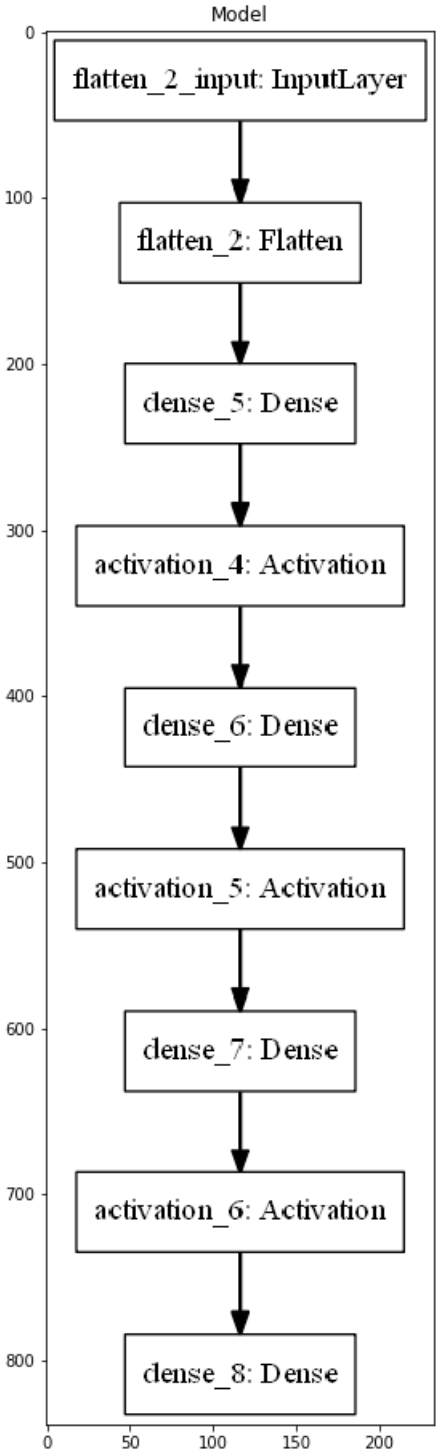
```
Tensor("dense_8/BiasAdd:0", shape=(?, 2), dtype=float32)
(None, 4)
(None, 16)
(None, 16)
(None, 16)
(None, 16)
(None, 16)
(None, 16)
(None, 2)


Model Layer Output Shape layer.get_output_at(0).get_shape().as_list()
[None, None]
[None, 16]
[None, 16]
[None, 16]
[None, 16]
[None, 16]
[None, 16]
```

[None, 2]

Model Layer Output Shape l.output_shape
(None, 4)
(None, 16)
(None, 16)
(None, 16)
(None, 16)
(None, 16)
(None, 16)
(None, 2)

Plot the Model and its Layers

/


Model
flatten_2_input: InputLayer
flatten_2: Flatten
dense_5: Dense
activation_4: Activation
dense_6: Dense
activation_5: Activation
dense_7: Dense
activation_6: Activation
dense_8: Dense
0
100
200
300
400
500
600
700
800
0
50
100
150
200

/

Create a deep Q network Agent

Compile the deep Q network Agent

Use Keras-RL callbacks for convenient model checkpointing and logging

Train the deep Q network Agent for 5000 steps

```
Training for 5000 steps ...
   10/5000: episode: 1, duration: 0.291s, episode steps: 10, steps per second: 34, episod
e reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000], m
ean observation: 0.132 [-1.967, 3.014], loss: --, mae: --, mean_q: --
   18/5000: episode: 2, duration: 1.725s, episode steps: 8, steps per second: 5, episode
 reward: 8.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000], mea
n observation: 0.178 [-1.519, 2.562], loss: 0.454291, mae: 0.598499, mean_q: 0.145811
   28/5000: episode: 3, duration: 0.079s, episode steps: 10, steps per second: 127, episo
de reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000],
 mean observation: 0.167 [-1.905, 3.089], loss: 0.326441, mae: 0.509657, mean_q: 0.310235
   41/5000: episode: 4, duration: 0.101s, episode steps: 13, steps per second: 129, episo
de reward: 13.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.154 [0.000, 1.000],
 mean observation: 0.109 [-1.785, 2.852], loss: 0.214880, mae: 0.412804, mean_q: 0.539174
   52/5000: episode: 5, duration: 0.089s, episode steps: 11, steps per second: 124, episo
de reward: 11.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000],
 mean observation: 0.152 [-2.111, 3.316], loss: 0.141770, mae: 0.317146, mean_q: 0.812477
   61/5000: episode: 6, duration: 0.072s, episode steps: 9, steps per second: 125, episod
e reward: 9.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000], me
an observation: 0.154 [-1.753, 2.794], loss: 0.122628, mae: 0.272088, mean_q: 1.010123
   71/5000: episode: 7, duration: 0.080s, episode steps: 10, steps per second: 125, episo
de reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000],
 mean observation: 0.155 [-1.913, 3.101], loss: 0.101965, mae: 0.233659, mean_q: 1.139329
   82/5000: episode: 8, duration: 0.088s, episode steps: 11, steps per second: 126, episo
de reward: 11.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.091 [0.000, 1.000],
 mean observation: 0.118 [-1.764, 2.754], loss: 0.092331, mae: 0.237285, mean_q: 1.254707
   91/5000: episode: 9, duration: 0.074s, episode steps: 9, steps per second: 121, episod
e reward: 9.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000], me
an observation: 0.159 [-1.716, 2.815], loss: 0.069547, mae: 0.253702, mean_q: 1.282520
  100/5000: episode: 10, duration: 0.074s, episode steps: 9, steps per second: 122, episo
de reward: 9.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000], m
ean observation: 0.151 [-1.779, 2.894], loss: 0.059420, mae: 0.271782, mean_q: 1.404533
  109/5000: episode: 11, duration: 0.079s, episode steps: 9, steps per second: 114, episo
de reward: 9.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000], m
ean observation: 0.134 [-1.787, 2.754], loss: 0.057834, mae: 0.293884, mean_q: 1.497505
  119/5000: episode: 12, duration: 0.080s, episode steps: 10, steps per second: 125, epis
ode reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.100 [0.000, 1.000],
mean observation: 0.138 [-1.526, 2.518], loss: 0.062822, mae: 0.325236, mean_q: 1.529075
  129/5000: episode: 13, duration: 0.089s, episode steps: 10, steps per second: 112, epis
ode reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.100 [0.000, 1.000],
mean observation: 0.145 [-1.717, 2.672], loss: 0.064777, mae: 0.366230, mean_q: 1.669357
  139/5000: episode: 14, duration: 0.109s, episode steps: 10, steps per second: 92, episo
de reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000],
 mean observation: 0.151 [-1.939, 3.082], loss: 0.058430, mae: 0.411845, mean_q: 1.596746
  150/5000: episode: 15, duration: 0.117s, episode steps: 11, steps per second: 94, episo
de reward: 11.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.091 [0.000, 1.000],
 mean observation: 0.109 [-1.768, 2.828], loss: 0.055849, mae: 0.459281, mean_q: 1.748447
  158/5000: episode: 16, duration: 0.085s, episode steps: 8, steps per second: 94, episod
e reward: 8.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000], me
an observation: 0.142 [-1.558, 2.523], loss: 0.059324, mae: 0.506196, mean_q: 1.730640
  168/5000: episode: 17, duration: 0.104s, episode steps: 10, steps per second: 97, episo
de reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000],
```

```
 mean observation: 0.126 [-1.937, 3.018], loss: 0.051501, mae: 0.543777, mean_q: 1.781355
  178/5000: episode: 18, duration: 0.109s, episode steps: 10, steps per second: 91, episo
de reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000],
 mean observation: 0.141 [-1.965, 3.056], loss: 0.045447, mae: 0.584823, mean_q: 1.858207
  188/5000: episode: 19, duration: 0.106s, episode steps: 10, steps per second: 94, episo
de reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000],
 mean observation: 0.173 [-1.932, 3.121], loss: 0.065308, mae: 0.633815, mean_q: 1.838645
  197/5000: episode: 20, duration: 0.097s, episode steps: 9, steps per second: 93, episod
e reward: 9.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000], me
an observation: 0.153 [-1.756, 2.807], loss: 0.065111, mae: 0.690170, mean_q: 1.940017
  207/5000: episode: 21, duration: 0.106s, episode steps: 10, steps per second: 95, episo
de reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.200 [0.000, 1.000],
 mean observation: 0.159 [-1.160, 2.070], loss: 0.065605, mae: 0.718963, mean_q: 1.919881
  216/5000: episode: 22, duration: 0.095s, episode steps: 9, steps per second: 95, episod
e reward: 9.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000], me
an observation: 0.145 [-1.773, 2.813], loss: 0.064549, mae: 0.772839, mean_q: 2.042578
  226/5000: episode: 23, duration: 0.108s, episode steps: 10, steps per second: 93, episo
de reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.100 [0.000, 1.000],
 mean observation: 0.142 [-1.533, 2.599], loss: 0.055372, mae: 0.820693, mean_q: 2.105072
  235/5000: episode: 24, duration: 0.096s, episode steps: 9, steps per second: 94, episod
e reward: 9.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000], me
an observation: 0.158 [-1.746, 2.798], loss: 0.055449, mae: 0.859219, mean_q: 2.096381
  246/5000: episode: 25, duration: 0.112s, episode steps: 11, steps per second: 98, episo
de reward: 11.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.091 [0.000, 1.000],
 mean observation: 0.109 [-1.775, 2.775], loss: 0.053857, mae: 0.927196, mean_q: 2.210764
  254/5000: episode: 26, duration: 0.098s, episode steps: 8, steps per second: 81, episod
e reward: 8.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000], me
an observation: 0.143 [-1.551, 2.549], loss: 0.057833, mae: 0.961933, mean_q: 2.209025
  264/5000: episode: 27, duration: 0.094s, episode steps: 10, steps per second: 106, epis
ode reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000],
mean observation: 0.118 [-1.964, 3.004], loss: 0.047974, mae: 1.006883, mean_q: 2.279693
  273/5000: episode: 28, duration: 0.075s, episode steps: 9, steps per second: 121, episo
de reward: 9.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000], m
ean observation: 0.138 [-1.810, 2.873], loss: 0.046399, mae: 1.050996, mean_q: 2.356630
  283/5000: episode: 29, duration: 0.090s, episode steps: 10, steps per second: 111, epis
ode reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000],
mean observation: 0.175 [-1.910, 3.113], loss: 0.051155, mae: 1.081558, mean_q: 2.332756
  292/5000: episode: 30, duration: 0.073s, episode steps: 9, steps per second: 123, episo
de reward: 9.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000], m
ean observation: 0.146 [-1.717, 2.796], loss: 0.047170, mae: 1.122321, mean_q: 2.434658
  302/5000: episode: 31, duration: 0.088s, episode steps: 10, steps per second: 114, epis
ode reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000],
mean observation: 0.127 [-1.985, 3.032], loss: 0.043530, mae: 1.138402, mean_q: 2.405537
```

```
  311/5000: episode: 32, duration: 0.090s, episode steps: 9, steps per second: 101, episo
de reward: 9.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000], m
ean observation: 0.141 [-1.807, 2.880], loss: 0.036995, mae: 1.197158, mean_q: 2.527449
  321/5000: episode: 33, duration: 0.083s, episode steps: 10, steps per second: 121, epis
ode reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.100 [0.000, 1.000],
mean observation: 0.131 [-1.544, 2.503], loss: 0.028095, mae: 1.270133, mean_q: 2.675335
  330/5000: episode: 34, duration: 0.072s, episode steps: 9, steps per second: 125, episo
de reward: 9.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000], m
ean observation: 0.156 [-1.712, 2.836], loss: 0.033210, mae: 1.253716, mean_q: 2.577719
  338/5000: episode: 35, duration: 0.069s, episode steps: 8, steps per second: 116, episo
de reward: 8.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000], m
ean observation: 0.144 [-1.590, 2.565], loss: 0.045628, mae: 1.302855, mean_q: 2.647108
  348/5000: episode: 36, duration: 0.082s, episode steps: 10, steps per second: 122, epis
ode reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.000 [0.000, 0.000],
mean observation: 0.134 [-1.977, 3.069], loss: 0.040504, mae: 1.339111, mean_q: 2.715513
  373/5000: episode: 37, duration: 0.215s, episode steps: 25, steps per second: 116, epis
ode reward: 25.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.480 [0.000, 1.000],
mean observation: -0.088 [-1.037, 0.426], loss: 0.045445, mae: 1.437801, mean_q: 2.807569
  398/5000: episode: 38, duration: 0.199s, episode steps: 25, steps per second: 126, epis
ode reward: 25.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.520 [0.000, 1.000],
mean observation: -0.062 [-1.014, 0.604], loss: 0.042003, mae: 1.558427, mean_q: 3.050294
  410/5000: episode: 39, duration: 0.102s, episode steps: 12, steps per second: 118, epis
ode reward: 12.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.667 [0.000, 1.000],
mean observation: -0.108 [-1.503, 0.791], loss: 0.064395, mae: 1.624939, mean_q: 3.130578
  424/5000: episode: 40, duration: 0.117s, episode steps: 14, steps per second: 120, epis
ode reward: 14.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.714 [0.000, 1.000],
mean observation: -0.107 [-1.995, 1.167], loss: 0.046754, mae: 1.746137, mean_q: 3.384033
  438/5000: episode: 41, duration: 0.122s, episode steps: 14, steps per second: 115, epis
ode reward: 14.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.786 [0.000, 1.000],
mean observation: -0.074 [-2.458, 1.569], loss: 0.066493, mae: 1.762538, mean_q: 3.404691
  448/5000: episode: 42, duration: 0.081s, episode steps: 10, steps per second: 123, epis
ode reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.900 [0.000, 1.000],
mean observation: -0.139 [-2.498, 1.518], loss: 0.093468, mae: 1.826491, mean_q: 3.492524
  458/5000: episode: 43, duration: 0.084s, episode steps: 10, steps per second: 119, epis
ode reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 1.000 [1.000, 1.000],
mean observation: -0.099 [-2.945, 1.990], loss: 0.090584, mae: 1.866340, mean_q: 3.574726
  467/5000: episode: 44, duration: 0.076s, episode steps: 9, steps per second: 118, episo
de reward: 9.000, mean reward: 1.000 [1.000, 1.000], mean action: 1.000 [1.000, 1.000], m
ean observation: -0.135 [-2.830, 1.801], loss: 0.143844, mae: 1.877949, mean_q: 3.588778
  477/5000: episode: 45, duration: 0.084s, episode steps: 10, steps per second: 119, epis
ode reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.900 [0.000, 1.000],
mean observation: -0.112 [-2.471, 1.606], loss: 0.157385, mae: 1.950302, mean_q: 3.686626
  490/5000: episode: 46, duration: 0.105s, episode steps: 13, steps per second: 124, epis
ode reward: 13.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.769 [0.000, 1.000],
mean observation: -0.078 [-2.412, 1.600], loss: 0.094953, mae: 2.047654, mean_q: 3.943588
  502/5000: episode: 47, duration: 0.100s, episode steps: 12, steps per second: 120, epis
ode reward: 12.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.833 [0.000, 1.000],
mean observation: -0.098 [-2.561, 1.608], loss: 0.100058, mae: 2.014257, mean_q: 3.832979
  512/5000: episode: 48, duration: 0.082s, episode steps: 10, steps per second: 122, epis
ode reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.900 [0.000, 1.000],
mean observation: -0.160 [-2.530, 1.534], loss: 0.136010, mae: 2.074991, mean_q: 3.956948
  522/5000: episode: 49, duration: 0.080s, episode steps: 10, steps per second: 124, epis
ode reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.800 [0.000, 1.000],
mean observation: -0.127 [-2.171, 1.363], loss: 0.154409, mae: 2.108635, mean_q: 4.003610
  531/5000: episode: 50, duration: 0.080s, episode steps: 9, steps per second: 113, episo
de reward: 9.000, mean reward: 1.000 [1.000, 1.000], mean action: 1.000 [1.000, 1.000], m
ean observation: -0.170 [-2.857, 1.741], loss: 0.147762, mae: 2.166375, mean_q: 4.143505
  540/5000: episode: 51, duration: 0.079s, episode steps: 9, steps per second: 114, episo
de reward: 9.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.778 [0.000, 1.000], m
```

/

```
ean observation: -0.127 [-2.125, 1.403], loss: 0.131347, mae: 2.160215, mean_q: 4.102093
  548/5000: episode: 52, duration: 0.068s, episode steps: 8, steps per second: 118, episo
de reward: 8.000, mean reward: 1.000 [1.000, 1.000], mean action: 1.000 [1.000, 1.000], m
ean observation: -0.150 [-2.537, 1.534], loss: 0.093786, mae: 2.168877, mean_q: 4.158699
  558/5000: episode: 53, duration: 0.086s, episode steps: 10, steps per second: 116, epis
ode reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.900 [0.000, 1.000],
mean observation: -0.134 [-2.591, 1.528], loss: 0.228492, mae: 2.264523, mean_q: 4.213691
  568/5000: episode: 54, duration: 0.089s, episode steps: 10, steps per second: 112, epis
ode reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 1.000 [1.000, 1.000],
mean observation: -0.160 [-3.135, 1.952], loss: 0.170119, mae: 2.269305, mean_q: 4.302257
  578/5000: episode: 55, duration: 0.092s, episode steps: 10, steps per second: 109, epis
ode reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 1.000 [1.000, 1.000],
mean observation: -0.149 [-3.083, 1.975], loss: 0.110511, mae: 2.316682, mean_q: 4.450932
  588/5000: episode: 56, duration: 0.084s, episode steps: 10, steps per second: 119, epis
ode reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.800 [0.000, 1.000],
mean observation: -0.149 [-2.024, 1.151], loss: 0.157980, mae: 2.377328, mean_q: 4.551291
  598/5000: episode: 57, duration: 0.084s, episode steps: 10, steps per second: 119, epis
ode reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 1.000 [1.000, 1.000],
mean observation: -0.151 [-3.091, 1.942], loss: 0.171920, mae: 2.406657, mean_q: 4.574345
  608/5000: episode: 58, duration: 0.085s, episode steps: 10, steps per second: 117, epis
ode reward: 10.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.800 [0.000, 1.000],
mean observation: -0.129 [-1.972, 1.208], loss: 0.165615, mae: 2.493935, mean_q: 4.740290
  627/5000: episode: 59, duration: 0.154s, episode steps: 19, steps per second: 124, epis
ode reward: 19.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.684 [0.000, 1.000],
mean observation: -0.074 [-2.215, 1.335], loss: 0.154227, mae: 2.472045, mean_q: 4.667439
  638/5000: episode: 60, duration: 0.091s, episode steps: 11, steps per second: 121, epis
ode reward: 11.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.727 [0.000, 1.000],
mean observation: -0.128 [-1.752, 0.945], loss: 0.125701, mae: 2.494137, mean_q: 4.740989
  663/5000: episode: 61, duration: 0.200s, episode steps: 25, steps per second: 125, epis
ode reward: 25.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.560 [0.000, 1.000],
mean observation: -0.092 [-1.455, 0.597], loss: 0.175838, mae: 2.621624, mean_q: 4.947594
  684/5000: episode: 62, duration: 0.168s, episode steps: 21, steps per second: 125, epis
ode reward: 21.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.571 [0.000, 1.000],
mean observation: -0.093 [-1.382, 0.612], loss: 0.146122, mae: 2.700436, mean_q: 5.113858
```

```
  733/5000: episode: 63, duration: 0.387s, episode steps: 49, steps per second: 127, epis
ode reward: 49.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.531 [0.000, 1.000],
mean observation: -0.075 [-1.552, 0.653], loss: 0.152839, mae: 2.768546, mean_q: 5.256532
  759/5000: episode: 64, duration: 0.219s, episode steps: 26, steps per second: 119, epis
ode reward: 26.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.538 [0.000, 1.000],
mean observation: -0.096 [-1.359, 0.466], loss: 0.135866, mae: 2.894452, mean_q: 5.521513
  787/5000: episode: 65, duration: 0.222s, episode steps: 28, steps per second: 126, epis
ode reward: 28.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.500 [0.000, 1.000],
mean observation: -0.120 [-0.903, 0.194], loss: 0.102334, mae: 3.032893, mean_q: 5.852368
  824/5000: episode: 66, duration: 0.289s, episode steps: 37, steps per second: 128, epis
ode reward: 37.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.486 [0.000, 1.000],
mean observation: -0.109 [-0.834, 0.275], loss: 0.109647, mae: 3.120635, mean_q: 6.040859
  866/5000: episode: 67, duration: 0.335s, episode steps: 42, steps per second: 125, epis
ode reward: 42.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.476 [0.000, 1.000],
mean observation: -0.133 [-0.882, 0.195], loss: 0.129854, mae: 3.306193, mean_q: 6.403864
  942/5000: episode: 68, duration: 0.610s, episode steps: 76, steps per second: 125, epis
ode reward: 76.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.487 [0.000, 1.000],
mean observation: 0.017 [-0.666, 0.599], loss: 0.140557, mae: 3.520003, mean_q: 6.805383
 1000/5000: episode: 69, duration: 0.545s, episode steps: 58, steps per second: 106, epis
ode reward: 58.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.448 [0.000, 1.000],
mean observation: -0.187 [-1.086, 0.404], loss: 0.141794, mae: 3.747918, mean_q: 7.274477
 1175/5000: episode: 70, duration: 1.430s, episode steps: 175, steps per second: 122, epi
sode reward: 175.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.469 [0.000, 1.00
0], mean observation: -0.380 [-2.429, 0.600], loss: 0.127730, mae: 4.186992, mean_q: 8.22
6146
 1375/5000: episode: 71, duration: 1.715s, episode steps: 200, steps per second: 117, epi
sode reward: 200.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.490 [0.000, 1.00
0], mean observation: -0.252 [-1.605, 0.443], loss: 0.145075, mae: 4.881478, mean_q: 9.70
4865
 1575/5000: episode: 72, duration: 1.685s, episode steps: 200, steps per second: 119, epi
sode reward: 200.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.490 [0.000, 1.00
0], mean observation: -0.243 [-1.565, 0.839], loss: 0.209172, mae: 5.706455, mean_q: 11.3
67964
 1775/5000: episode: 73, duration: 1.813s, episode steps: 200, steps per second: 110, epi
sode reward: 200.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.490 [0.000, 1.00
0], mean observation: -0.315 [-1.997, 0.533], loss: 0.272271, mae: 6.476453, mean_q: 12.9
27590
 1975/5000: episode: 74, duration: 1.712s, episode steps: 200, steps per second: 117, epi
sode reward: 200.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.485 [0.000, 1.00
0], mean observation: -0.278 [-1.807, 0.619], loss: 0.295935, mae: 7.230465, mean_q: 14.4
62371
 2175/5000: episode: 75, duration: 2.051s, episode steps: 200, steps per second: 98, epis
ode reward: 200.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.490 [0.000, 1.00
0], mean observation: -0.292 [-1.835, 0.569], loss: 0.277529, mae: 7.951419, mean_q: 15.9
65154
 2375/5000: episode: 76, duration: 1.789s, episode steps: 200, steps per second: 112, epi
sode reward: 200.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.490 [0.000, 1.00
0], mean observation: -0.273 [-1.772, 0.586], loss: 0.241732, mae: 8.808797, mean_q: 17.7
19254
 2575/5000: episode: 77, duration: 1.750s, episode steps: 200, steps per second: 114, epi
sode reward: 200.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.490 [0.000, 1.00
0], mean observation: -0.357 [-2.156, 0.565], loss: 0.494572, mae: 9.572337, mean_q: 19.2
34221
 2775/5000: episode: 78, duration: 1.921s, episode steps: 200, steps per second: 104, epi
sode reward: 200.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.490 [0.000, 1.00
0], mean observation: -0.240 [-1.567, 0.787], loss: 0.622973, mae: 10.331852, mean_q: 20.
729204
 2975/5000: episode: 79, duration: 1.912s, episode steps: 200, steps per second: 105, epi
sode reward: 200.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.490 [0.000, 1.00
```

/

```
0], mean observation: -0.360 [-2.200, 0.621], loss: 0.665681, mae: 11.075564, mean_q: 22.
260723
 3175/5000: episode: 80, duration: 1.750s, episode steps: 200, steps per second: 114, epi
sode reward: 200.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.490 [0.000, 1.00
0], mean observation: -0.319 [-1.958, 0.618], loss: 0.827838, mae: 11.774877, mean_q: 23.
641541
 3375/5000: episode: 81, duration: 2.332s, episode steps: 200, steps per second: 86, epis
ode reward: 200.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.490 [0.000, 1.00
0], mean observation: -0.316 [-1.949, 0.879], loss: 0.739013, mae: 12.510545, mean_q: 25.
136274
 3575/5000: episode: 82, duration: 1.746s, episode steps: 200, steps per second: 115, epi
sode reward: 200.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.485 [0.000, 1.00
0], mean observation: -0.332 [-2.097, 0.639], loss: 0.726417, mae: 13.064812, mean_q: 26.
322561
 3775/5000: episode: 83, duration: 2.466s, episode steps: 200, steps per second: 81, epis
ode reward: 200.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.490 [0.000, 1.00
0], mean observation: -0.290 [-1.888, 0.538], loss: 0.786486, mae: 13.809456, mean_q: 27.
798203
 3975/5000: episode: 84, duration: 1.914s, episode steps: 200, steps per second: 105, epi
sode reward: 200.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.490 [0.000, 1.00
0], mean observation: -0.254 [-1.691, 0.561], loss: 0.863271, mae: 14.344907, mean_q: 28.
912426
 4175/5000: episode: 85, duration: 1.685s, episode steps: 200, steps per second: 119, epi
sode reward: 200.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.485 [0.000, 1.00
0], mean observation: -0.246 [-1.715, 0.558], loss: 0.926563, mae: 15.054695, mean_q: 30.
388418
 4375/5000: episode: 86, duration: 1.817s, episode steps: 200, steps per second: 110, epi
sode reward: 200.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.490 [0.000, 1.00
0], mean observation: -0.234 [-1.543, 0.765], loss: 0.975346, mae: 15.764216, mean_q: 31.
802553
 4575/5000: episode: 87, duration: 1.839s, episode steps: 200, steps per second: 109, epi
sode reward: 200.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.490 [0.000, 1.00
0], mean observation: -0.341 [-2.085, 0.698], loss: 1.058545, mae: 16.353354, mean_q: 33.
008862
 4775/5000: episode: 88, duration: 1.866s, episode steps: 200, steps per second: 107, epi
sode reward: 200.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.490 [0.000, 1.00
0], mean observation: -0.299 [-1.866, 0.617], loss: 0.908102, mae: 17.031160, mean_q: 34.
404156
 4975/5000: episode: 89, duration: 2.505s, episode steps: 200, steps per second: 80, epis
ode reward: 200.000, mean reward: 1.000 [1.000, 1.000], mean action: 0.490 [0.000, 1.00
0], mean observation: -0.318 [-2.041, 0.581], loss: 1.327677, mae: 17.480297, mean_q: 35.
184235
done, took 48.513 seconds
```

Note: After the first 250 episodes, we see that the total rewards for the episode approa
ch 200 and the episode steps also approach 200. This means that the agent has learned to
balance the pole on the cart until the environment ends at a maximum of 200 steps.

Test the deep Q network Agent for 5 Episodes

```
Testing for 5 episodes ...
Episode 1: reward: 200.000, steps: 200
Episode 2: reward: 200.000, steps: 200
Episode 3: reward: 200.000, steps: 200
Episode 4: reward: 200.000, steps: 200
Episode 5: reward: 200.000, steps: 200
```

Out[2]: <keras.callbacks.callbacks.History at 0x23c8b771240>