DDPG Agent – Continuous Control Problem – Udacity
Rens ter Weijde, March 2019

Environment

The toy environment used to train and display the algorithm is the Unity Reacher environment. This is technically a multi-joint arm that shows a continuous control problem, for which DDPG is ideal. The goal is to ensure that the arm stays in the right target area. The observation space has 33 variables, the action space has 4 possible actions (on two joints in the arm).

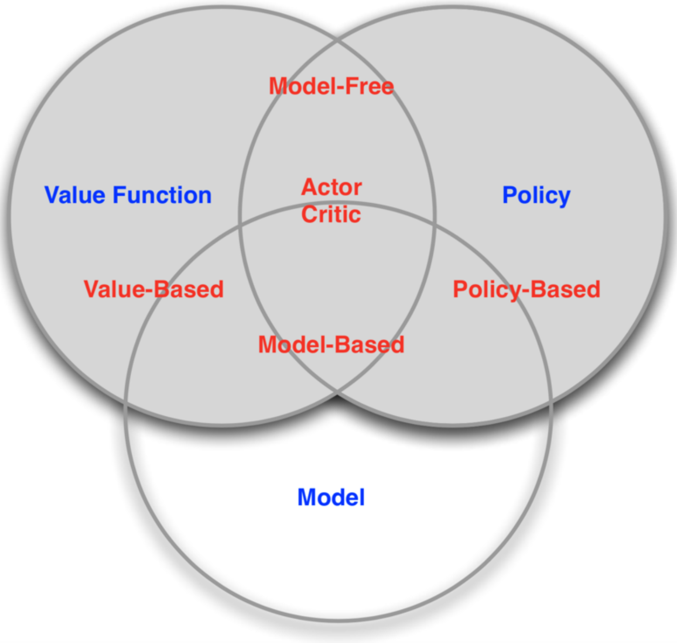
Algorithm

The DDPG algorithm is the focused on improving performance in continuous action environments, and is presented by the authors as actor-critic, for having both an actor network (selecting the right action) and a critic network (calculating the Q-values that are then used to update the actor). In that sense, it falls between classic Q-value-methods and policy-gradient methods. For a visual explanation of where Actor-Critic models fall, see the visual below.

In essence: Monte Carlo episodes have the problem of high variance (but low bias), where policy gradient methods like Temporal Difference learning have a low variance (but are biased). Combining these two models to minimize the central weakness in each approach is what actor-critic models stand for. The actor is now not trained on the actual values after a full episode, but on the values that it receives from the critic.

Implementation

Note that both actor and critic are relatively simple neural nets (e.g. 256, 128 nodes in hidden layers); where the Critic has 4 layers in the implementation (another 128 node FC layer) and the Actor 3. As is often the case, argued by Karpathy, ~3 layers should be enough to build a model that works reasonably well.



Taken from presentation by David Silver

The algorithm implementation furthermore applies leaky ReLUs, also popular with GANs implementation, to fix the 'dying ReLU' problem. Also, it uses Experience Replay to store tuples of experiences (S, A, R, S') and randomly sample from them for learning; this is done to minimize correlation between sequential experiences and provides for more stable learning in general. As a final note: I've chosen a relatively small batch size of 128 because the training was very slow; still took ~3 hours for 2,000 episodes with a 128 batch size on GPU. This also made the training comparison of different hyperparameter setups hard; with a larger batch size (e.g. 1024) the training would have taken ~5-6 hours (and Udacity workspaces go idle after 30 minutes). In addition, the implementation uses 'soft updates', where during each step a small % of the weights are shared with the target network. This stands in contract to DQN, where after x episodes all the weights are copied.

As a final note: I picked a smaller number of epochs (1000 vs 2000 in original setup) to speed up the learning; this indeed increased the speed of learning but the learning steps were too small to converge to 30. I therefore changed this to 1800.

The parameters used in the training setup are quite similar to the Udacity original script:

- Replay buffer size: 1e6
- Batch size: 128 (small than I liked, but for speed reasons)
- Gamma: 0.99
- Tau: 1e-3
- Actor Learning rate (alpha): 1e-4
- Critic Learning Rate (alpha): 3e-4
- N_episodes: 2000 (needed for convergence towards +30)
- Epochs: 1800 (after trying 1000)
- Leakage for leaky ReLU (negative slope): 0.01
- Update interval: 20

Results

The training was much slower than I anticipated, as it took multiple hours on GPU (0800-1400 = 6 hours on GPU for 1800 episodes, in which I had to stay near to the computer in order to not let the Udacity environment go idle!). This made experimentation harder. I therefore hope the final results are ok this way – the agent reached ~35, but would need a bit more time to stay there in a stable fashion.

What helped was the fact that although you could still see high variance in the 10-episode rounds (e.g. there are some serious drops in performance at times), the Average Score would continue to rise quite linearly so I could tell the algorithm worked as planned. What worried me mostly was the low efficiency with which it ran; getting to an average score of 30 would probably have taken 3 more hours at this speed.

The detailed results of the training are shown below:

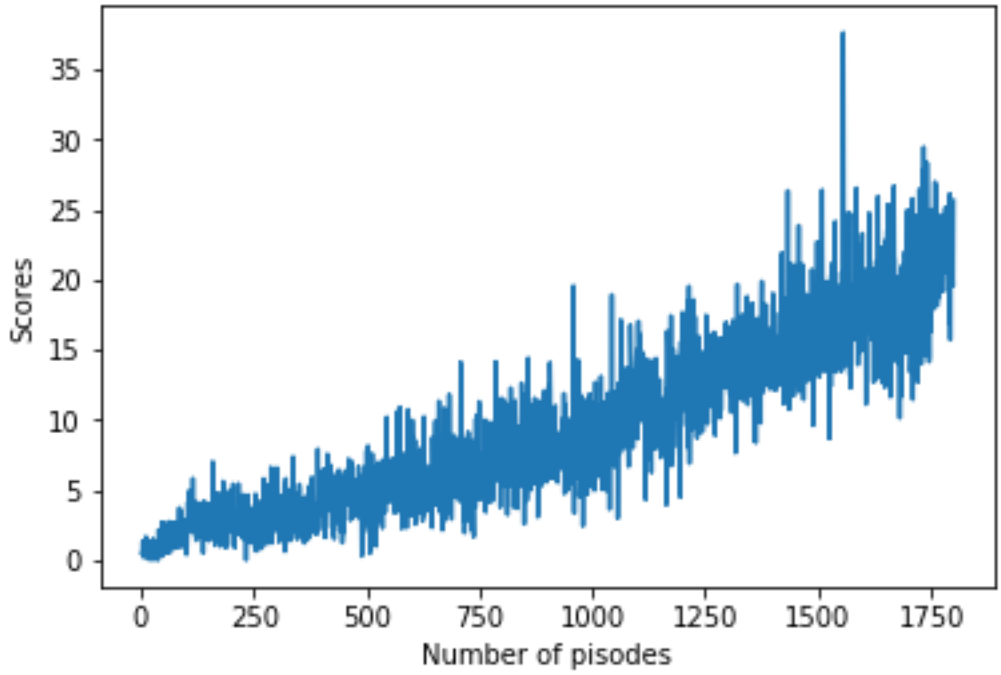
```
Training on cuda:0 started...
Episode: 10    Average Score: 0.70    Current Score: 1.01
```

```
Episode: 20    Average Score: 0.71   Current Score: 0.30
Episode: 30    Average Score: 0.73   Current Score: 0.39
Episode: 40    Average Score: 0.79   Current Score: 1.61
Episode: 50    Average Score: 0.87   Current Score: 1.93
Episode: 60    Average Score: 0.98   Current Score: 1.59
Episode: 70    Average Score: 1.08   Current Score: 1.86
Episode: 80    Average Score: 1.16   Current Score: 2.58
Episode: 90    Average Score: 1.25   Current Score: 3.54
Episode: 100   Average Score: 1.35   Current Score: 2.77
Episode: 110   Average Score: 1.55   Current Score: 2.52
Episode: 120   Average Score: 1.81   Current Score: 1.94
Episode: 130   Average Score: 1.96   Current Score: 1.97
Episode: 140   Average Score: 2.11   Current Score: 4.22
Episode: 150   Average Score: 2.24   Current Score: 2.89
Episode: 160   Average Score: 2.38   Current Score: 1.53
Episode: 170   Average Score: 2.46   Current Score: 2.09
Episode: 180   Average Score: 2.55   Current Score: 2.41
Episode: 190   Average Score: 2.64   Current Score: 2.91
Episode: 200   Average Score: 2.74   Current Score: 3.08
Episode: 210   Average Score: 2.73   Current Score: 2.60
Episode: 220   Average Score: 2.70   Current Score: 3.90
Episode: 230   Average Score: 2.78   Current Score: 1.43
Episode: 240   Average Score: 2.81   Current Score: 3.40
Episode: 250   Average Score: 2.83   Current Score: 3.29
Episode: 260   Average Score: 2.80   Current Score: 3.26
Episode: 270   Average Score: 2.82   Current Score: 4.02
Episode: 280   Average Score: 2.90   Current Score: 3.98
Episode: 290   Average Score: 2.94   Current Score: 3.07
Episode: 300   Average Score: 3.01   Current Score: 4.92
Episode: 310   Average Score: 3.11   Current Score: 3.77
Episode: 320   Average Score: 3.08   Current Score: 5.17
Episode: 330   Average Score: 3.09   Current Score: 1.37
Episode: 340   Average Score: 3.14   Current Score: 3.17
Episode: 350   Average Score: 3.20   Current Score: 1.96
Episode: 360   Average Score: 3.26   Current Score: 3.01
Episode: 370   Average Score: 3.24   Current Score: 2.33
Episode: 380   Average Score: 3.28   Current Score: 4.44
Episode: 390   Average Score: 3.24   Current Score: 1.87
Episode: 400   Average Score: 3.36   Current Score: 4.38
Episode: 410   Average Score: 3.47   Current Score: 4.85
Episode: 420   Average Score: 3.75   Current Score: 6.69
Episode: 430   Average Score: 3.80   Current Score: 4.56
Episode: 440   Average Score: 3.91   Current Score: 2.68
Episode: 450   Average Score: 3.95   Current Score: 1.96
Episode: 460   Average Score: 4.15   Current Score: 7.23
Episode: 470   Average Score: 4.32   Current Score: 3.05
Episode: 480   Average Score: 4.41   Current Score: 4.80
```

```
Episode: 490   Average Score: 4.56    Current Score: 3.22
Episode: 500   Average Score: 4.49    Current Score: 7.84
Episode: 510   Average Score: 4.43    Current Score: 5.93
Episode: 520   Average Score: 4.38    Current Score: 3.62
Episode: 530   Average Score: 4.44    Current Score: 5.35
Episode: 540   Average Score: 4.54    Current Score: 6.93
Episode: 550   Average Score: 4.74    Current Score: 6.071
Episode: 560   Average Score: 4.77    Current Score: 4.90
Episode: 570   Average Score: 4.96    Current Score: 6.122
Episode: 580   Average Score: 5.01    Current Score: 2.952
Episode: 590   Average Score: 5.16    Current Score: 8.69
Episode: 600   Average Score: 5.20    Current Score: 3.224
Episode: 610   Average Score: 5.45    Current Score: 7.600
Episode: 620   Average Score: 5.47    Current Score: 4.19
Episode: 630   Average Score: 5.67    Current Score: 5.734
Episode: 640   Average Score: 5.62    Current Score: 5.29
Episode: 650   Average Score: 5.65    Current Score: 8.10
Episode: 660   Average Score: 5.84    Current Score: 11.35
Episode: 670   Average Score: 5.78    Current Score: 6.05
Episode: 680   Average Score: 5.90    Current Score: 6.515
Episode: 690   Average Score: 6.00    Current Score: 8.682
Episode: 700   Average Score: 6.17    Current Score: 8.70
Episode: 710   Average Score: 6.29    Current Score: 7.607
Episode: 720   Average Score: 6.26    Current Score: 4.40
Episode: 730   Average Score: 6.18    Current Score: 5.46
Episode: 740   Average Score: 6.25    Current Score: 8.09
Episode: 750   Average Score: 6.40    Current Score: 5.311
Episode: 760   Average Score: 6.25    Current Score: 6.27
Episode: 770   Average Score: 6.35    Current Score: 6.522
Episode: 780   Average Score: 6.33    Current Score: 6.83
Episode: 790   Average Score: 6.50    Current Score: 9.473
Episode: 800   Average Score: 6.68    Current Score: 7.806
Episode: 810   Average Score: 6.79    Current Score: 6.102
Episode: 820   Average Score: 7.10    Current Score: 5.549
Episode: 830   Average Score: 7.31    Current Score: 9.829
Episode: 840   Average Score: 7.42    Current Score: 6.49
Episode: 850   Average Score: 7.48    Current Score: 11.78
Episode: 860   Average Score: 7.69    Current Score: 9.203
Episode: 870   Average Score: 7.78    Current Score: 5.951
Episode: 880   Average Score: 7.95    Current Score: 7.727
Episode: 890   Average Score: 7.89    Current Score: 6.058
Episode: 900   Average Score: 7.95    Current Score: 8.530
Episode: 910   Average Score: 7.98    Current Score: 6.333
Episode: 920   Average Score: 8.04    Current Score: 9.183
Episode: 930   Average Score: 7.99    Current Score: 7.35
Episode: 940   Average Score: 8.18    Current Score: 8.975
Episode: 950   Average Score: 8.18    Current Score: 10.08
```

```
Episode: 960   Average Score: 8.17    Current Score: 3.348
Episode: 970   Average Score: 8.29    Current Score: 8.096
Episode: 980   Average Score: 8.21    Current Score: 5.864
Episode: 990   Average Score: 8.27    Current Score: 7.585
Episode: 1000  Average Score: 8.20    Current Score: 7.70
Episode: 1010  Average Score: 8.11    Current Score: 8.480
Episode: 1020  Average Score: 8.17    Current Score: 7.952
Episode: 1030  Average Score: 8.32    Current Score: 5.690
Episode: 1040  Average Score: 8.34    Current Score: 11.33
Episode: 1050  Average Score: 8.58    Current Score: 10.35
Episode: 1060  Average Score: 8.73    Current Score: 8.991
Episode: 1070  Average Score: 8.94    Current Score: 7.975
Episode: 1080  Average Score: 9.23    Current Score: 6.673
Episode: 1090  Average Score: 9.61    Current Score: 8.000
Episode: 1100  Average Score: 10.04   Current Score: 8.648
Episode: 1110  Average Score: 10.52   Current Score: 10.72
Episode: 1120  Average Score: 10.76   Current Score: 12.54
Episode: 1130  Average Score: 11.07   Current Score: 12.45
Episode: 1140  Average Score: 11.22   Current Score: 7.947
Episode: 1150  Average Score: 11.31   Current Score: 10.52
Episode: 1160  Average Score: 11.37   Current Score: 10.78
Episode: 1170  Average Score: 11.43   Current Score: 10.29
Episode: 1180  Average Score: 11.48   Current Score: 8.047
Episode: 1190  Average Score: 11.51   Current Score: 12.51
Episode: 1200  Average Score: 11.51   Current Score: 17.69
Episode: 1210  Average Score: 11.39   Current Score: 12.61
Episode: 1220  Average Score: 11.57   Current Score: 11.18
Episode: 1230  Average Score: 11.67   Current Score: 10.84
Episode: 1240  Average Score: 11.78   Current Score: 10.00
Episode: 1250  Average Score: 11.81   Current Score: 13.89
Episode: 1260  Average Score: 12.20   Current Score: 16.11
Episode: 1270  Average Score: 12.31   Current Score: 16.26
Episode: 1280  Average Score: 12.50   Current Score: 14.96
Episode: 1290  Average Score: 12.58   Current Score: 14.04
Episode: 1300  Average Score: 12.77   Current Score: 14.19
Episode: 1310  Average Score: 12.93   Current Score: 15.35
Episode: 1320  Average Score: 13.03   Current Score: 13.20
Episode: 1330  Average Score: 13.17   Current Score: 11.45
Episode: 1340  Average Score: 13.49   Current Score: 15.62
Episode: 1350  Average Score: 13.76   Current Score: 16.64
Episode: 1360  Average Score: 13.79   Current Score: 8.885
Episode: 1370  Average Score: 13.86   Current Score: 17.08
Episode: 1380  Average Score: 14.08   Current Score: 14.51
Episode: 1390  Average Score: 14.31   Current Score: 12.40
Episode: 1400  Average Score: 14.23   Current Score: 19.04
Episode: 1410  Average Score: 14.34   Current Score: 14.56
Episode: 1420  Average Score: 14.48   Current Score: 21.93
```

```
Episode: 1430  Average Score: 14.52   Current Score: 13.20
Episode: 1440  Average Score: 14.73   Current Score: 21.14
Episode: 1450  Average Score: 14.96   Current Score: 12.47
Episode: 1460  Average Score: 15.18   Current Score: 13.56
Episode: 1470  Average Score: 15.52   Current Score: 17.34
Episode: 1480  Average Score: 15.60   Current Score: 15.25
Episode: 1490  Average Score: 15.61   Current Score: 19.00
Episode: 1500  Average Score: 15.96   Current Score: 15.14
Episode: 1510  Average Score: 16.25   Current Score: 20.07
Episode: 1520  Average Score: 16.38   Current Score: 19.89
Episode: 1530  Average Score: 16.36   Current Score: 14.14
Episode: 1540  Average Score: 16.53   Current Score: 17.54
Episode: 1550  Average Score: 16.57   Current Score: 19.03
Episode: 1560  Average Score: 16.83   Current Score: 17.58
Episode: 1570  Average Score: 16.94   Current Score: 13.22
Episode: 1580  Average Score: 17.16   Current Score: 24.74
Episode: 1590  Average Score: 17.66   Current Score: 22.26
Episode: 1600  Average Score: 17.83   Current Score: 19.77
Episode: 1610  Average Score: 17.85   Current Score: 19.77
Episode: 1620  Average Score: 17.99   Current Score: 15.71
Episode: 1630  Average Score: 18.19   Current Score: 12.92
Episode: 1640  Average Score: 18.25   Current Score: 22.29
Episode: 1650  Average Score: 18.36   Current Score: 22.85
Episode: 1660  Average Score: 18.29   Current Score: 11.63
Episode: 1670  Average Score: 18.55   Current Score: 21.35
Episode: 1680  Average Score: 18.49   Current Score: 17.96
Episode: 1690  Average Score: 18.20   Current Score: 18.27
Episode: 1700  Average Score: 18.25   Current Score: 21.69
Episode: 1710  Average Score: 18.25   Current Score: 11.45
Episode: 1720  Average Score: 18.43   Current Score: 14.91
Episode: 1730  Average Score: 18.77   Current Score: 17.82
Episode: 1740  Average Score: 18.93   Current Score: 17.47
Episode: 1750  Average Score: 19.11   Current Score: 21.06
Episode: 1760  Average Score: 19.48   Current Score: 21.32
Episode: 1770  Average Score: 19.57   Current Score: 20.43
Episode: 1780  Average Score: 20.02   Current Score: 24.53
Episode: 1790  Average Score: 20.56   Current Score: 25.63
Episode: 1800  Average Score: 20.80   Current Score: 25.66
```



Ideas for improvement

Some of the key ideas, like leaky ReLU instead of ReLU, experience buffers etc. are already there. Other ideas that I have to make it better:

- Let it run even longer; in order to get to a stable 30 it would take more hours.
- Increase the batch_size (now 128) to minimize the variance a bit. I made the batch_size smaller to improve training speed, but there seems to be a downside to it as well.
- Experiment with the learning rates for actor (1e-4) and critic (3e-4); both are small. Maybe training would be faster if they started at a higher point.
- Experiment with different network architectures. The networks are basic and small; maybe another layer (or more nodes, e.g. 512) could be interesting to see what would happen. Also include dropout in those layers. This new architecture with extra layers comes with the risk of making the training even slower.
- Try the same environment with other algorithms, in particular PPO or A3C (quite some trained Unity environments with PPO on Youtube).
- Improve the Replay Buffer with Prioritized Replay. This has also shown to be beneficial for performance.