**Deep Reinforcement Learning with Double Q-learning**

Hado van Hasselt, Arthur Guez, David Silver

Carefully read the paper (pages 1-7). We will discuss the questions below.

Note: the online network is also called the policy network in literature.

---

Questions for Discussion:

1. What is the main problem that this paper addresses?

Overestimation of the action values in Q learning and in DQN. This overestimation was leading to poorer policies. The solution proposed is based on the tabular Double Q-learning approach and applied to neural networks. This results in a new novel algorithm know as **Double DQN.**

1. How does the algorithm solve this problem? Hint: How do the neural network weights play a role?

It solves this problem by having a second network, known as the target network with different weights that are used to evaluate the value of the policy. The target networks parameters are copied from the online network periodically.

1. Look at the update equation for the target. How does it differ from DQN?

Like Q learning, the selection of the action in argmax is still due to the online weights. We are still estimating the value of the greedy policy according to the current parameter values in theta. However, we use the second set of weights in theta prime to fairly evaluate the value of this policy.

Double DQN uses online network for selection, and the target for evaluation.

DQN uses the target network for both selecting and evaluating.