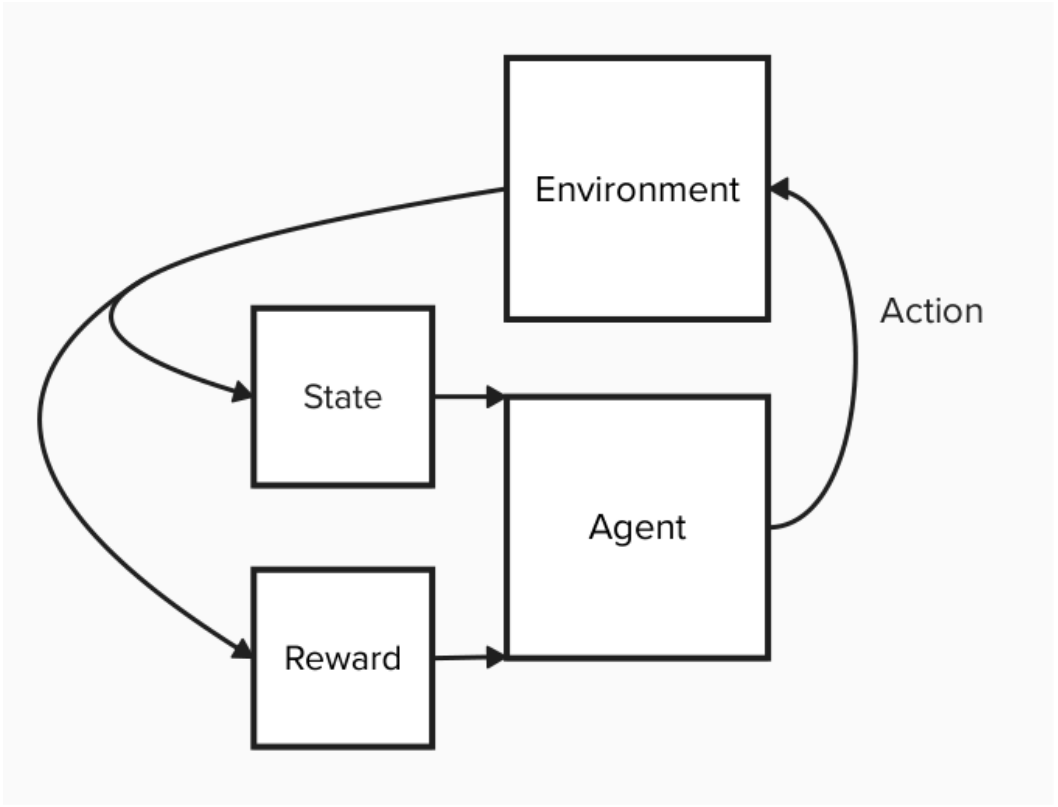
Task 3

DQN

- Supervised Learning Algorithm
- Agent tries to get the highest cumulative reward
- Takes actions based on Q-values
  - Expected cumulative reward (looks into future)
- Q-values are calculated by the neural network acting as a function approximator
  - Q-values calculated for each state-action pair
  - Uses Epsilon greedy exploration
    - Likelihood decays with time
- Neural network is updated based on the rewards from the agent
  - Randomly sampled from a replay memory
  - To update the policy network

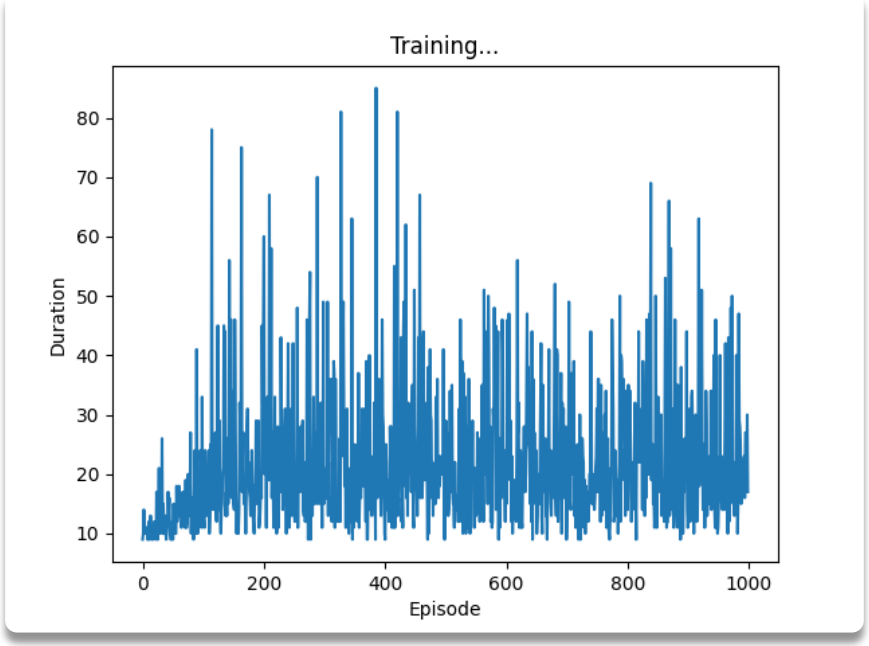


Network Updates

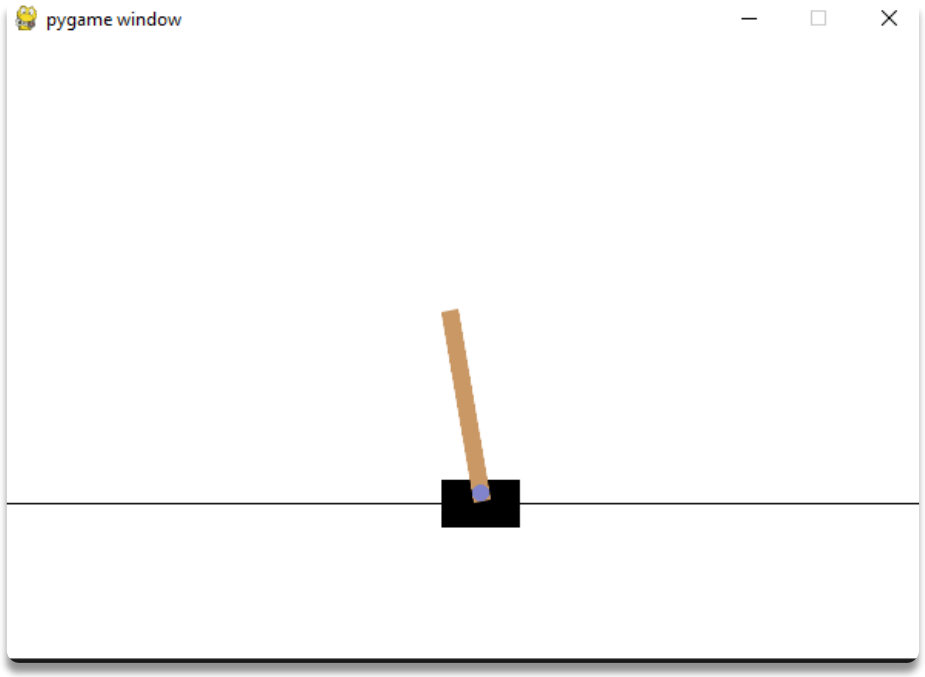
- Policy network updates come from minimising the loss function
  - $L(\theta) = E_{(S_t, A_t, R_{t+1}, S_{t+1})\sim\text{Replay Memory}}\left[\left(R_{t+1} + \gamma max_a Q(S_{t+1}, A_{t+1}; \theta^-) - Q(S_t, A_t; \theta)\right)^2\right]$
  - $L(\theta)$ is the network parameters
  - R_{t+1} is the reward from transitioning from state S_t to S_{t+1} using action A_t
  - Where γ is the discount factor and α is the learning rate
- Changes from policy network are reflected in the target network via
  - targetState = policyState * τ + targetState * (1.0 - τ)
  - Where the states are the state dictionaries of the models

Training

- Run with parameters
  - Batch Size : 128
  - Replay Memory Size : 1000
  - Episodes : 1000
  - Gamma : 0.99
  - Learning Rate : 0.0001
  - Epsilon Start : 0.9
  - Epsilon End : 0.01
  - Epsilon Decay : 1000
  - Hidden Layer Size : 64



Outputs



- Model “cartPole.pth” is saved to be reloaded
  - This is the model after all the training not the best run
  - Best run could be saved by only saving when a model reaches best performance or by using checkpoints
  - Using savedModelDemo.py the model can be displayed
- Prints the best score out that the model achieved in its run
  - 85 from the training conducted with previous parameters

Conclusion

- Highest duration is rather low
  - Indicates further parameter tuning is needed
  - Would suggest changing:
    - Replay memory
    - Learning rate
    - Epsilon decay
    - Hidden layers