Hyperparameter Notes

* Generally, convergence is easier when the asset is more volatile, or the time step is longer, as this enhances the effect of actions and makes the trend easier to detect.

# DQN

* There is no need for excessively large networks, and indeed using them requires a larger batch size, which slows down the algorithm considerably.
* There is a trade-off between Learning Rate, Epoch and batch size, changing one invariably means the others need to be altered
  + A small batch size requires a small learning rate, as any anomalies will be exaggerated, and the network must not over fit to them (and vice versa)
  + A high epoch also requires a smaller learning rate, as the same data gets fitted towards multiple times. Typically using a high epoch and low learning rate will allow for better fitting, as overshoot is less likely. On the other hand, it is more likely that a local solution will be found.
* Sigmoid is always used as it is more stable than Relu due to the bounded output.
* Triangular networks appear to perform better than square networks off the same flexibility.
* Greater retrain frequency is usually advantageous, however it increases runtime, and if batch size is large and the model is retrained every episode the network will be fitted to the same dataset at each fitting.

# Actor Critic

* Both the actor and critic neural networks use sigmoid activation function as it is more stable when it experiences extreme inputs.
* Sigma is decayed from 2 to 0.05, as a large value is required to begin to speed the agent’s convergence towards the correct solution, and a much smaller value is required later on to allow the agent to fine tune its prediction.

## Learning Rate

* The critic needs to quickly converge to the value function of the newest iteration of the policy, thus it uses a high epoch and learning rate combination.
* The actor has a much lower learning rate & epoch combination as pushing it too far causes instability and overshooting.

## Retrain Frequency

* The actor and critic are retrained at the same frequency (although the critic is fitted to a greater extent each time)
* The frequency may not be set too low as the effect of anomalous experience will be exaggerated (especially since AC operated using Monte Carlo update targets), and it may not be too high as convergence is slowed.

## Network Size

* Both the actor and critic managed to solve single and multi (2) asset problems with a network size of 4,4. This is mainly due to the structure of the agent; each network is responsible for predicting only a fraction of the total problem.

## Regularization

* The Actor has a significantly higher regularization term than the critic, as it is important in this setting that the actor appreciate the independence of action on wealth.