A/ MOTIVATION: (insights gotten from http://www.wildml.com/2018/02/introduction-to-learning-to-trade-with-reinforcement-learning/)

- Price Prediction model can forecast a price, however the policy is very simple (if predicted price above threshold then buy for example). **That said, Reinforcement Learning learns the optimal action rather than models the market.**

- One quick fix for Price Prediction Model is to encode more information we want into the loss function at each timestep(e.g the commision cost for each trade). However, by making use of Reinforcement Learning, we can use many better techniques such as future discounted cumulative rewards rather than instantaneous benefit, actor critic pair models...

B/ TARGET PROBLEMS

- Optimizing for only one stocks (3 options: buy, sell, hold)

- Portfolio optimization (can make the actions become continuous, let's define it here)

C/ COMMON METHODS

- Mostly Deep - Q Network (DQN), with discounted future rewards. A drawback of Q-learning is that the action space cannot be continuous. (This belongs to value-iteration approaches).

Proposed experiments:
1/ Make use of latest update in DQN like Dueling/Double DQN, Prioritized Experience Replay for one stock trading strategy.

2/ Use actor-critic models to perform portfolio management. Can try out more advanced actor-critic such as : https://arxiv.org/abs/1712.08987
Note: Actor-critic is quite similar to GAN: **in GANs, the generator is trying to approximate the data distribution, and the discriminator is trying to evaluate the distribution of the generator; in actor-critic, the actor is trying to approximate the policy, that is, the distribution $P(a|s)$, and the critic is trying to evaluate this policy.**

D/ Baseline:

- **Note**: Reinforcement Learning usually makes use of neural networks as a component inside. So the way to compare the result is by training a neural network, using the price prediction to perform some simple policy, then compare with the Reinforcement Learning approach to see the difference in the profits made. Should I use HAN+SFM or a simple RNN? I suggest using a simple RNN is enough because the point is to prove RL works.

- **Baseline 1**: Implemented a Deep-Q network which makes use of LSTM inside and an uniform “experience replay” to derive trading strategy for only one stock (DONE)

- **Experiment 1**: Update the baseline 1 with Fixed-Q target, Dual/Double DQN, Prioritized Experience Replay.

- **Experiment 2**: Figure out the portfolio management with actor critic