Deep-Q Learning

Salvin Chowdhury

February 25, 2025

1 Decision Problems

In finance, *optimization* and associated techniques play a central role. Finance is nothing but the systemic application of optimization techniques to problems of a financial context. Here are the different optimization problems:

- *Discrete vs Continuous Action Space:* the quantites or action to be chosen through optimization can be from a set of finite, discrete options (*optimal choice*) or from a set of inifnite, continuous options (*optimal control*)
- *Static vs Dynamic Problems:* static problems are one-off optimization problems, dynamic problems are characterized by a typically large number of sequential and connected optimization problems
- *Finite vs Infinite Horizon:* dynamic optimization problems have a finite or infinite horizon. Playing a game of chess has a finite horizon. Climate policy is a descision problem with infinite hoirzons.
- *Discrete vs Continuous Time:* some dynamic problems only require discrete decisions and optimizations at different points in time, such as chess. Other dynamic problems require continuous decisions and optimizations. For example, driving a car.

2 Q-Learning

QL is based on an agent interacting with the environment and learning from the ensuing experiences through rewards and penalties. A QL agent takes actions based on two principles:

- *Exploitation:* refers to the actions taken by the QL agent under the current optimal policy QL
- *Exploration:* refers to actions taken by a Ql agent that are random. The purpose is to explore random actions and their associated values beyond what the current optimal policy would dictate

A QL agent usually follows a ϵ-greedy strategy, where ϵ is defined as the ratio with which the agent relies on exploration as compared to exploitation. During training, it is assumed that the ϵ decreases with a increasing number of training units.

2.1 Deep Q-Learning

In DQL, the policy Q is regularly updated through *replay*. For replay, the agent stores passed experiences such as (state, actions, rewards and next states and etc.) and use small batches from the memorized experiences to retrain the DNN.