NC STATE UNIVERSITY

Machine Learning Student Symposium 2025

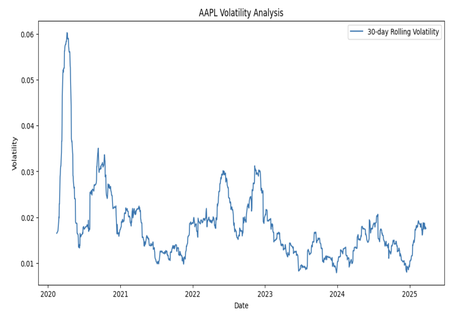
Group 113

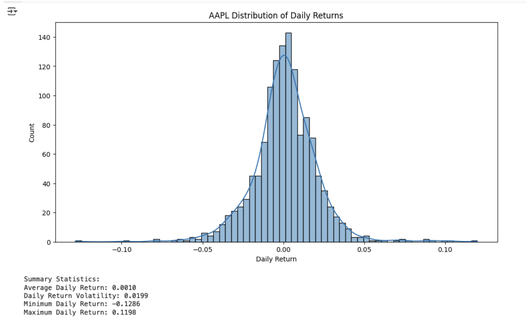
Adversarially Robust Deep Q-Network for Algorithmic Trading

Sreejeet Maity, Sai Kavya Marthala, Rajesh Debnath

Problem Statement : Deep Q-Networks (DQNs) for algorithmic trading are vulnerable to adversarial attacks and noisy, high-dimensional market data. Even small, targeted data perturbations can cause significant trading losses. Our goal is to develop a robust DQN that maintains reliable performance despite adversarial contamination and market uncertainties.

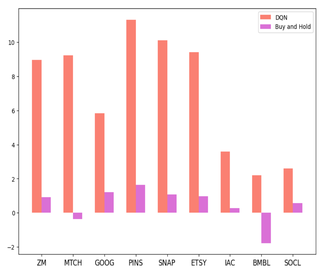
Data : We use the Yahoo Finance Dataset (2018–2023), which provides daily open, high, low, and close prices for equities, ETFs, and indexes. Data is cleaned to remove missing values from weekends and holidays, then normalized for consistency.

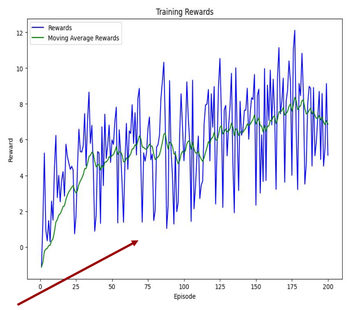




Baseline

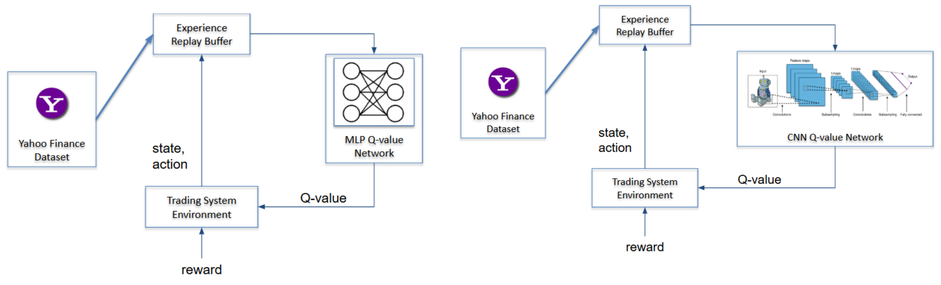
- The following is a DQN implementation of the Algorithmic Trading Problem under ***no data corruption.***
- For the Neural Approximation part, we have used a MLP, and a CNN (shown below).





Increasing Reward

Neural Architecture

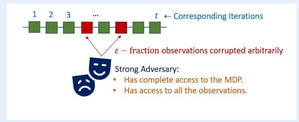


MDP Modelling

We will model this as a *Markov Decision Process*
States defined using lagged returns of length K.
At each time step t, the state is given as follows
$s_t = [r_{t-K+1}\ r_{t-K+2}\ , \ldots, r_t\ ]$.
The action space represents possible positions the trader can take on each day. That means the trader can take long, take short or stay neutral, but only for a fixed magnitude. a_t ={-1,0,1}.
Rewards can be modelled as $R_t = a_t r_t$.

Corruption Model

- Observe the entire reward set $\{R_t(s,a)\}_{(s,a)\in S\times A}$ in each iteration t.
- Perturb $\varepsilon \in \left[0, \frac{1}{2}\right)$ fraction of these observed rewards up to t.



Result 1: Provable Vulnerability of DQN

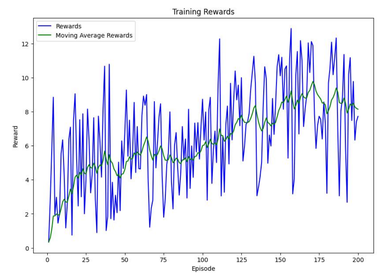
Under the corruption model, assume an adversary *corrupts the reward at each iteration with probability* ϵ (Huber Contamination).

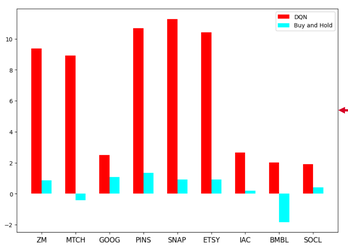
Theorem 1: Under Huber contamination, and a suitable step size α, with probability 1, $Q_t \to \widetilde{Q_c^*}$, where $\widetilde{Q_c^*}$ is the unique fixed point of the perturbed Bellman operator $\mathcal{T}_c^*$, satisfying:

$$(\mathcal{T}_c^* Q)(s,a) = \mathrm{R}_c(\mathrm{s},\mathrm{a}) + \gamma \sum_{s' \in S} P(s'|s,a)\ \max_{\mathrm{a}' \in \mathrm{A}} Q(s',a')$$

Result 2: Adversarially Robust DQN using robust estimators [1],[2]

Sharp Decline in Rewards due to adversary





Using robust estimators for rewards as per [1], [2] we have constructed an adversarially robust DQN.

Main References

- Robust Q under Corrupted Rewards (Maity, Mitra CDC 2024)
- Adversarially Robust TD Learning (Maity, Mitra AISTATS 2025)