RL Project Proposal
Do NoisyNet and RND exploration offer complementary benefits for exploration or are their effects redundant?

RL_MA
github.com/automl-classroom/rl-week-8-rl_ma

1 Motivation

Exploration is a key challenge in reinforcement learning, especially in sparse-reward environments. Its also a topic that many groups, including us, struggled with in the exercise. That is why we want to dive deeper into exploration methods: While Noisy Networks and intrinsic motivation methods like Random Network Distillation (RND) have both been shown to improve exploration individually, it is unclear whether they provide complementary benefits or exhibit redundant behavior when combined. This project aims to investigate the interaction between NoisyNets and RND in the context of DQN, comparing their individual and combined effects on exploration efficiency.

2 Related Topics

This project relates to the lecture topics of exploration (Lecture 7) and DQN (Lecture 4), as well as to broader themes in intrinsic motivation and uncertainty-driven exploration. It builds directly on the work of Fortunato et al. (2017), who introduced NoisyNets for exploration, and Burda et al. (2018), who proposed RND as a scalable intrinsic motivation method.

- Fortunato et al., "Noisy Networks for Exploration" (2017)[1]
- Burda et al., "Exploration by Random Network Distillation" (2018)[2]

3 Idea

Our goal is to empirically evaluate whether Noisy Networks and Random Network Distillation (RND) offer complementary exploration benefits when used together or whether their effects are redundant. We will analyze both learning performance and exploration behavior (e.g., state visitation diversity) of DQN agents using different exploration strategies as listed in the experiments section. Our hypothesis is that combining NoisyNets and RND will lead to improved exploration.

4 Experiments

We want to know roughly what you are planning to show in terms of experiments.

Environments & Metrics We plan to evaluate all methods on a MiniGrid environment:

- **MiniGrid:** Specifically, sparse-reward tasks like `MiniGrid-Empty-8x8`

[1] https://arxiv.org/abs/1706.10295
[2] https://arxiv.org/abs/1810.12894

Metrics will include:

- Total episode return over time (learning curve)
- Time to first reward (exploration success)
- State visitation heatmaps (qualitative analysis)
- Variance across seeds (stability)

Experimental Scope We will test the following 4 agent variants in the same MiniGrid environment:

1. DQN + ϵ-greedy (as baseline)
2. DQN + NoisyNet
3. DQN + RND
4. DQN + NoisyNet + RND

Each variant will be run with 10 random seeds. We will perform light hyperparameter tuning on RND reward scale and NoisyNet initialization (based on literature defaults).

Estimated Computational Load Each experiment (1 variant, 1 seed) is expected to run for 1-3 hours on a single CPU or basic GPU machine. In total, this results in approximately 4 variants $\times$ 10 seeds $\times$ 1 environment $= 40$ runs, totaling roughly 40–120 compute hours.

5 Timeline

- **Research & Design (1 day):** Deep dive into NoisyNet and RND implementations, finalize experiment plan, and review related work.
- **Implementation (2-3 days):** Integrate NoisyNet and RND into a DQN codebase.
- **Experiments (6 days):** Run all training variants and collect metrics.
- **Analysis (2 days):** Generate plots, heatmaps, and compute statistics.
- **Reporting (1 da):** Write final report and design poster, including result discussion and limitations.