Project (AY25/26 T1)

Reinforcement Learning is about thinking long-term for the future. And what better way to illustrate the benefits of long-term planning, than to think about fishing. (And, the ocean and water are cool!)

Assume that only *two species* exist in the ocean - Salmon and Sharks. The salmon population increases as it reproduces and new salmon are spawned, while it decreases as it dies from old age or disease, as well as be eaten by sharks and caught by human. Whether there is a net increase or decrease is determined by the relative balance.

As for sharks, their net rate of increase generally depends on whether there is an abundance of prey. If there are plenty of prey (salmon), the shark population will have a tendency to increase (with all else being equal). On the other hand, when salmon is scarce relative to the shark population, the shark population is expected to decrease.

We can therefore expect some sort of natural 'correction' effect and a dynamic equilibrium. If the prey population increases, so too will the predator, hence keeping the prey population in check eventually. On the other end of things, if the prey drastically reduces in numbers, so too will the predators after a few generations, hence allowing the prey to repopulate itself without being hunted to extinction.

But bring in humans to the picture, and things change. Humans, unlike sharks, do not die when any of its prey approaches extinction (there are plenty of land animals). At least, not yet. Humans catch salmon for food (let's ignore the fact that humans catch sharks too). Therefore, the salmon has two predators - sharks and humans.

A key difference is that Humans has a choice of how much salmon to hunt. For simplicity, let's assume just a single human (e.g. the Government) that decides the *effort* (eg. money, manpower, or other resources – all resented as a single quantitative number) to dedicate towards catching salmon. With all else being equal, a higher effort would lead to a higher number of salmon being caught from the ocean.

You are given a function update_populations:

```
def update_populations(salmon_t, shark_t, fishing_effort_t, month_t):
    """
    Update salmon and shark populations for one time step (month)
    """
    ### Cool stuff inside. I can share it with you all after the project is completed.
    return salmon_caught_t, salmon_t_plus_1, shark_t_plus_1
```

This function has four arguments as input. You can acquire observations as many times as you like. Given (salmon_t, shark_t, fishing_effort_t, month_t), the function returns (salmon_caught_t, salmon_t_plus_1, shark_t_plus_1).

salmon_t and shark_t are the number of salmon and sharks, respectively, in the ocean at the beginning of month_t, while fishing_effort_t is the effort dedicated towards catching salmon for that particular month. Note that effort is not linear; putting 10 times the effort may not translate to getting 10 times the number of salmon. (Try setting the effort to an extremely large number and you will see that algorithm is also robust enough to account for the fact that it's impossible to catch every single salmon in the ocean.)

The reproduction rate of salmon and sharks may be affected by seasonality as well as long-term super-cycles. This will be computed internally based on the month_t, but you are not privy to the inner workings and exact formula.

Build and train an RL agent that, given a state (salmon_t, shark_t, month_t), returns an action fishing_effort_t, which can be any non-negative number. There is no upper bound (eg. imagine dedicating an entire country's manpower and resources into catching salmon, which would be as good as infinite effort for all intends and purposes). In the next time-step, the month will be incremented by one (from t to t+1), and the number of salmon and sharks at the start of the new month will correspond to salmon_caught_t, salmon_t_plus_1, respectively.

The reward for each timestep will be

$$r_t = K_1 \,\text{salmon\_caught\_t} - K_2 \,\text{fishing\_effort\_t}$$

The entire scenario will terminate after 900 months, ie. 75 years. Upon reaching the terminal state, there is an additional sustainability reward corresponding to the log of the number of salmon and sharks remaining. Assume (*may be quite a stretch*) that humans are extremely long-sighted, and there will not be any discounting. Therefore, the total return will be

$$G = \sum_{t=1}^{900} r_t + K_3 \log(\text{salmon\_t\_plus\_1}|_{t=900}) + K_4 \log(\text{shark\_t\_plus\_1}|_{t=900})$$

where $K_1 = 0.001, K_2 = 0.01, K_3 = 100, K_4 = 100$.

You are also given a 'starter kit' to test out the calling of the function (distributed as a wheel – refer here for usage guide) as well as the computation of rewards. Ensure that your agent is compatible with the given script. There is an internal random factor that influence the reproduction rate, but during inference I will have a separate script that extracts this value, so that the exact same factor is applied to all groups for fairness.

Total 25 marks, corresponding to 25% of the overall course grade. The grading criteria and guidelines are as follows:

Proposal – 5 marks
Describe what your group plans to do, in terms of both theory and coding.

- What algorithms will you try, and why?
- What are the libraries and compute resources that would be used? How do you intend to track and present your experiments?
- How would you know if you moving in the correct direction? Have you considered the ways in which you will be able to score marks in the subsequent sections?
- What are the key technical challenges that you foresee?
- Who will be doing what? (ie. distribution of workload among all team members)

Code – 10 marks
of which

- 8 marks for performance
- 2 marks for code readability and documentations

- Scoring higher than a simple benchmark: 2 marks
- Relative ranking: 6 marks for the top group, and pro-rated for all other groups

Note that if the agent adopts a rule-based method in any way, total performance marks will be capped at 2 out of 8.

Final Report (& presentation) – 10 marks
of which

- 5 marks for presentation
- 5 marks for written report

- Demonstrate that you are clear about what's going on, and that the work is not blindly copy-and-paste from somewhere.
- How informative are your visual aids? Is it rich in contents, with plenty of useful information?
- How do you actively apply what you have learnt in the course?
- How do you fill the gaps and acquire additional knowledge through self-learning?
- What are the insights you have gained along the way? If certain things do not work, can you explain why?