Homework 4: We become data scientists

ENM1050, UPenn

Due date: October 30th by midnight (11:59pm)

This is an **individual assignment**. Submit your answers on Canvas using the instructions at the end of the handout. Late submissions will be accepted until midnight of the following Wednesday (11:59pm), but will be penalized by 10% for each partial or full date late. After the late deadline, no further assignments may be submitted; post a private message on Ed to request an extension if you need one due to a special situation such as illness.

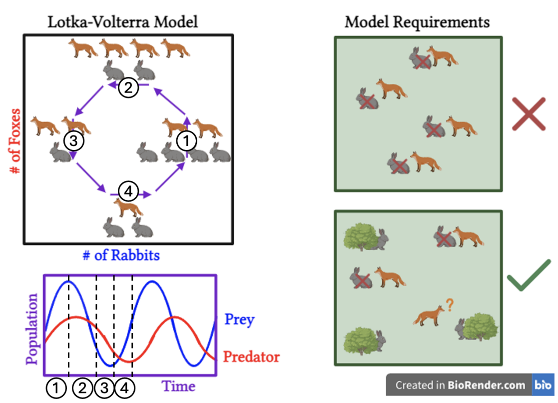

Figure 1: In population dynamics, we assume that we can derive a model for how predator and prey species interact, and that we can use that model to say things about the long term .

You may talk with other students about this assignment, ask the teaching team questions, use a calculator and other tools, and consult outside sources such as the Internet or a LLM. When you get stuck, post a question on Ed or go to office hours! **Warning.** This homework consists of multiple parts, and will require you to start it early. Make sure you begin early enough that you can get support at office hours - we designed this to force you to go to OH so you can get 1-1 interaction with the teaching team.

Population dynamics and Lotka-Volterra

Many systems involve *agents* which interact with each other through an assumed system of interactions or laws. Systems modeling is used throughout science, economics, and engineering; the equations we will use were developed by Lotka in 1910 to study chemical reactions, and he extended them to study interactions between herbivores and plant species in 1920. Volterra independently derived an identical model in 1926 to describe the interaction of predator and prey fish species during World War 1. These equations are often called *predator-prey* relationships, or the *Lotka-Volterra* model. They describe any system where interacting agents are consumed relative to each other; this could be fuel/oxidizers in chemistry or the spread of a virus in epidemiology (like how scientists studied COVID).

In the typical model,

$$\frac{dx}{dt} = ax - bxy$$
$$\frac{dy}{dt} = -cy + dxy$$

where: $x(t)$ represents the population of the prey species at time t. $y(t)$ represents the population of the predator species at time t. a is the intrinsic growth rate of the prey species. b is the predation rate. c is the natural death rate of the predator species. d is the efficiency of converting prey into predator offspring.

The *model parameters* (a, b, c, d) are tuned to match historical data of a given population. Let's take a look through each term on the right hand side. The first term (ax) corresponds to how quickly prey is born. The second term $(-bxy)$ determines how quickly they are hunted. Note that its proportional to both x and y; if there are no prey or no predators, not hunting can occur, and the more of each, the more opportunities there will be for them to run into each other. On the predator dynamics $(\dot{y})$, a death rate is assumed $(-cy)$, and a birth rate which depends again on how frequently they can eat the prey (dxy).

This system of differential equations describes the dynamics of a predator-prey relationship, where the prey population grows exponentially in the absence of predators, and the predator population declines exponentially in the absence of prey. The interaction term, $-bxy$, describes the decrease in prey population due to predation, while the interaction term, dxy, describes the increase in predator population due to predation.

Your assignment.

Your task is to collect a dataset and fit a scientific model to describe it. We'll break this down into three parts:

1. **Data collection and cleanup.** Run the bouncy rock paper scissors code, and collect data into a format that we can easily work with.

2. **Fit a model.** Build a model in PyTorch to describe the data, and use gradient descent to train the model.

3. **Reason about model.** Run the model to perform experiments and build hypotheses about the model system.

Each of these individual components will be a chunk of effort, so make sure you spread this assignment out and don't attempt it all at once. You'll assemble all the pieces into a report at the end.

Part 1 - Data collection and cleanup.

- Run the bouncy rock paper scissors code. The code generates a running plot in the corner of the screen, showing what portion of the population consists of a rock, papers, and scissors.

- When the predators (e.g. scissors) come into contact with prey (e.g. paper) they both switch to the same type (e.g. both become predators). Discuss in your writeup how this differs from the original Lotka-Volterra model. Of the four terms on the right-hand side, which are applicable to describing this setting? Which do you think don't make sense?

- Use pickle (as directed in the in-class exercise) to pull out 20 runs. Make sure that you save the output after an amount of time that makes sense - e.g. after there is only one species remaining, or some other threshold.

- At the completion of this part, you should have a directory containing a list of 20 rock/paper/scissor population percentages that you can load into pytorch.

Part 2 - Model fitting

- In PyTorch implement a model of the following form:

$$\dot{x} = \alpha\, x * y + \beta\, x * z$$

$$\dot{y} = \gamma\, y * x + \delta\, y * z$$

$$\dot{z} = \epsilon\, z * x + \zeta\, z * y$$

- Take the example from class on Wednesday showing how to fit a model to data, and modify it to load in your curated dataset.

- Train the model to find parameters $(\alpha, \beta, \gamma, \delta, \epsilon, \zeta)$ which best describe your dataset.

- Make sure in your writeup that you explain how you picked hyperparameters.

Part 3 - The actual science In your report, discuss the following questions.

- What values did you obtain for the 6 parameters? Describe anything about symmetries or the values that they take that you find interesting.

- Use odeint to solve the model for many different initial conditions (i.e. pick different initial states for x, y, z satisfying $0 \leq x, y, z \leq 1$. What steady state behavior do you observe? Do you observe multiple peaks in a species population? Do they die off and then come back? Is there always one winner? Briefly summarize your experiments in a paragraph, using figures to support your discussion.

- Briefly describe (one paragraph) in words what you think would happen if you changed the attractive/repulsive forces between scissors and paper. What if you randomly "birthed" rocks by adding them sporadically into the domain? Use simulations from odeint to justify your hypotheses.

Note: Turn in all python files generated in your project, as well as a single pdf (generated in word or however you'd like) describing your experiments. Pay careful attention to how you present information, you will be graded both on technical correctness and clarity of presentation.