02-712 Week 1

Biological Modeling and Simulation

Aidan Jan

August 26, 2025

Course Topics

Models for Optimization

- Ex. Sequence Similarity: Given a DNA or a protein, find other DNA or proteins that are similar.
  - We can model this problem as a sequence alignment problem, which can be optimized.
  - We can also model this problem as an alignment free problem, for example, dividing the DNA or protein into K-mers and comparing them.
- This class focuses not on how to align the DNA or protein, but instead focuses on why we would use one representation of the problem over another.
  - Why do we pick one representation over another?
  - What are the parameters for the problem?

Simulation and Sampling

Suppose A, B, and C are chemicals, such that $A + B \to C$, with rate k.

- How can you simulate the concentrations of A, B, or C over time?
- Maybe use a graph, or perhaps differential equations (mass action model).
- Or, we can do a discrete system. Start with some amount of A, and B, and roll a dice, and if it is greater than a certain number, make a C.
- Why would we pick one representation of the problem over another?

Model Inference and AI Modeling

Suppose we have chemicals A, B, and C, and $A + B \rightleftharpoons C$. Forward reaction with rate k_1, and backwards reaction with rate k_2. We did the reaction and recorded the concentrations over time (discrete), and plotted them on a graph.

- How do we pick parameters of the model to match the data?
- Maybe find the values of k_1 and k_2 that best fit the system?
- How do we even know the correct equation is $A + B \rightleftharpoons C$? It could be $A + 2B \rightleftharpoons C$, etc. (Structural Inference)

Model Building

Consider the evolutionary tree building problem. You have species human, gorilla, cow, chicken, tuna, and fly. How do we model this, or how do we design a program to model this?

Questions to consider:

1. How are we going to answer this?
2. What information do we have to work with? (e.g., phenotypes, genetic information, etc.)
3. Which properties matter a lot, and which do not?
4. What assumptions can we make?