EBIO 5460
Fundamentals of Data Science for Ecology & Evolution

Dr Brett Melbourne
Professor, EBIO
brett.melbourne@colorado.edu
Office hours: Any time by appointment
Office: Ramaley N336 and Zoom
Pronouns: he, him, his

Git & GitHub

- Class Github organization
- Bookmark this:
- https://github.com/EBIO5460Fall2025
- Organization, syllabus, timetable
- Slides, code, homework
- You'll also submit your work here

Slides for today

- github.com/EBIO5460Fall2025
- Go to repositories
- Open class-materials
- 01_1_slides_thu_intro_welcome

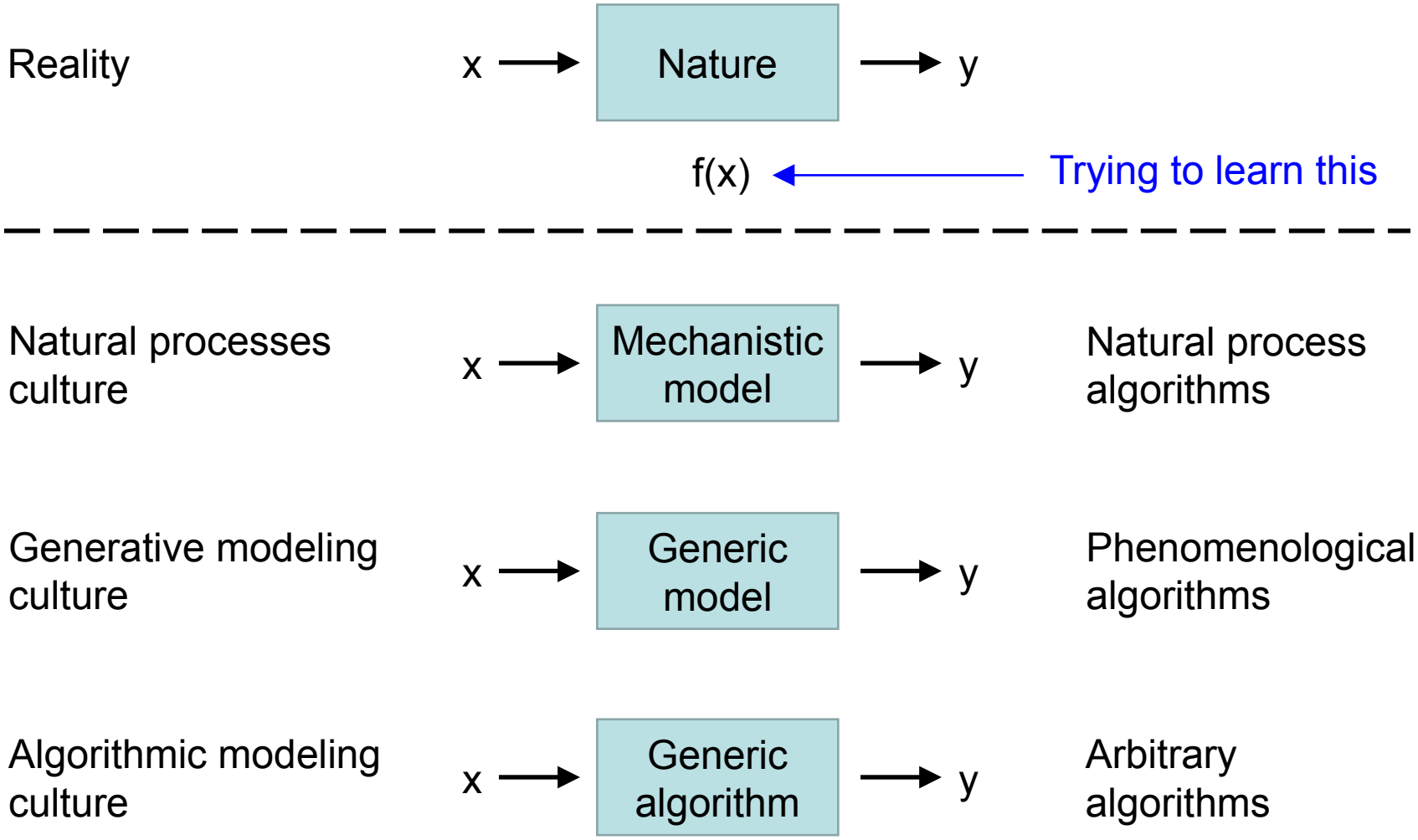
Today

- What is data science?
- Introductions
- Syllabus & how we'll do the class

What is data science?

- Workflows and algorithms to learn from data
- Learning goal
  - Confident to use a range of skills and concepts to plan for, acquire, manage, analyze, infer or predict from, and report about datasets of any size in your area of biological research
  - Become awesome quantitative scientists!

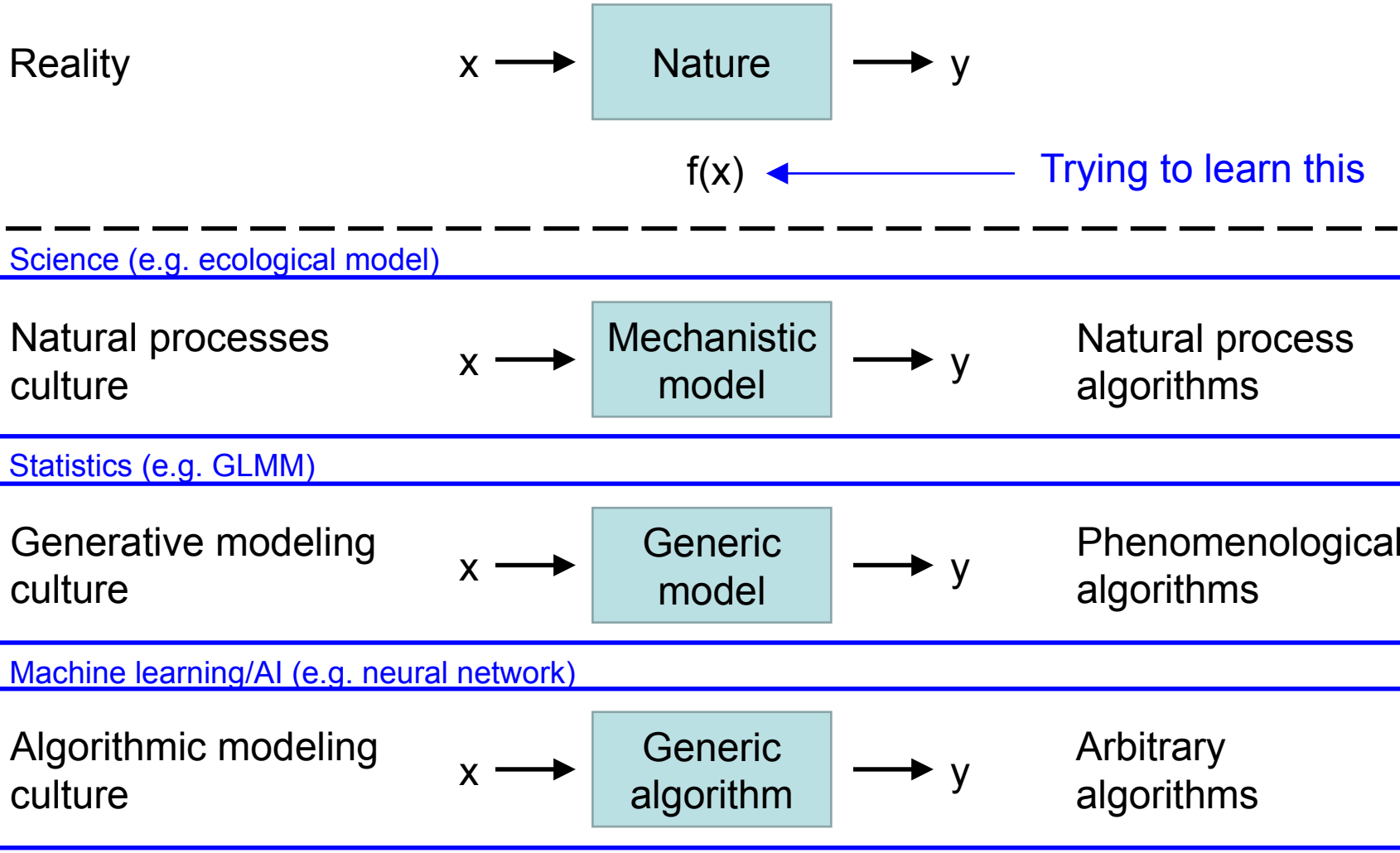
Data science cultures



f can mean different things in different cultures

Breiman (2001)
Denoho (2017)

Data science cultures



f can mean different things in different cultures

Breiman (2001)
Denoho (2017)

Model

Definition: a representation of nature

My philosophy

~~What analysis should I run on my data?~~
~~What package should I run my data through?~~

How can I best model nature?

Algorithm

- Procedure for solving a problem in terms of actions to execute and order to execute them
- Code

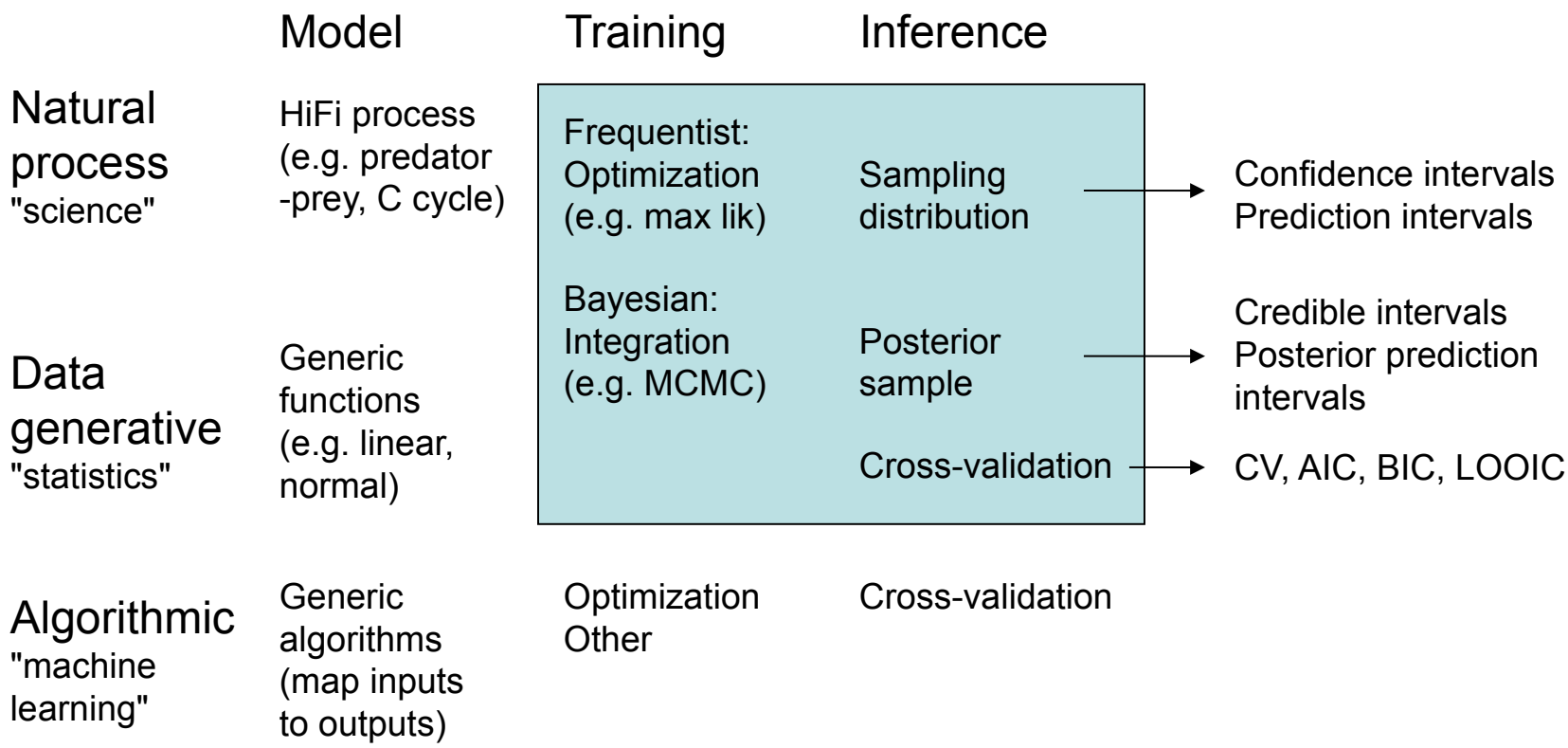
Algorithms in data science

- Model algorithm
- Training algorithm
- Inference (reliability) algorithm

Modeling with data

Algorithm classes

Modeling culture



Data generating processes

- How will/did my data come to be?
- Simulating data
  - explore the biology (alternative models)
  - design studies
  - check understanding & methods
- Key to great science IMO

Introductions

- Name (and pronouns)
- Masters or PhD (what year)?
- Advisor
- Department
- What fascinates you (your research)?
- Hopes for the course

Algorithms and models

- Understand the broad classes
- Frequentist, Bayesian, likelihood, information theory, machine learning
- Ecol/evo, generative (generic), algorithmic
- We'll start with
  - 1) learn how to program algorithms
  - 2) consider simple models from each perspective above

Learning format

- Flipped, often. Short video lectures. Sometimes short live lectures.
- Collaborative learning. Work in small groups or share in small groups.
- Piazza: collaboratively discuss the homework. Collaborative learning is not only allowed but **encouraged** in this class! FERPA compliant.

Staying healthy

- If you’re sick, consider joining via Zoom
- I’ll record lectures
- Let’s all stay well!

Languages

- C for fundamentals of programming
- R and Python for data science

AI tools

- Use them to help you learn to code
- Not code for you! (vibe coding)
- Examples
  - ChatGPT (OpenAI)
  - Copilot (Microsoft/OpenAI, var ChatGPT) CU
  - Gemini (Google) CU
  - Claude (Anthropic)
  - Others: open source, ...

Strategies, limitations

Grading

- GitHub portfolio
- 65% continuous Github code commits
- 35% individual assignment
  - 15% presentation in exam week
  - 20% submitted materials

Homework this week

- Posted to GitHub
  - e.g. “hw4tue”
- Set up GitHub
- Set up Positron IDE
- Set up C tooling
- Update R
- Intro to programming