EBIO 5460
Data Science for Biological Research

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/EBIO5460Fall2024
- Organization, syllabus, timetable
- Slides, code, homework
- You'll also submit your work here

Slides for today

- github.com/EBIO5460Fall2024
- Go to repositories
- Open class-materials
- 01_1_welcome_to_data_science_5460

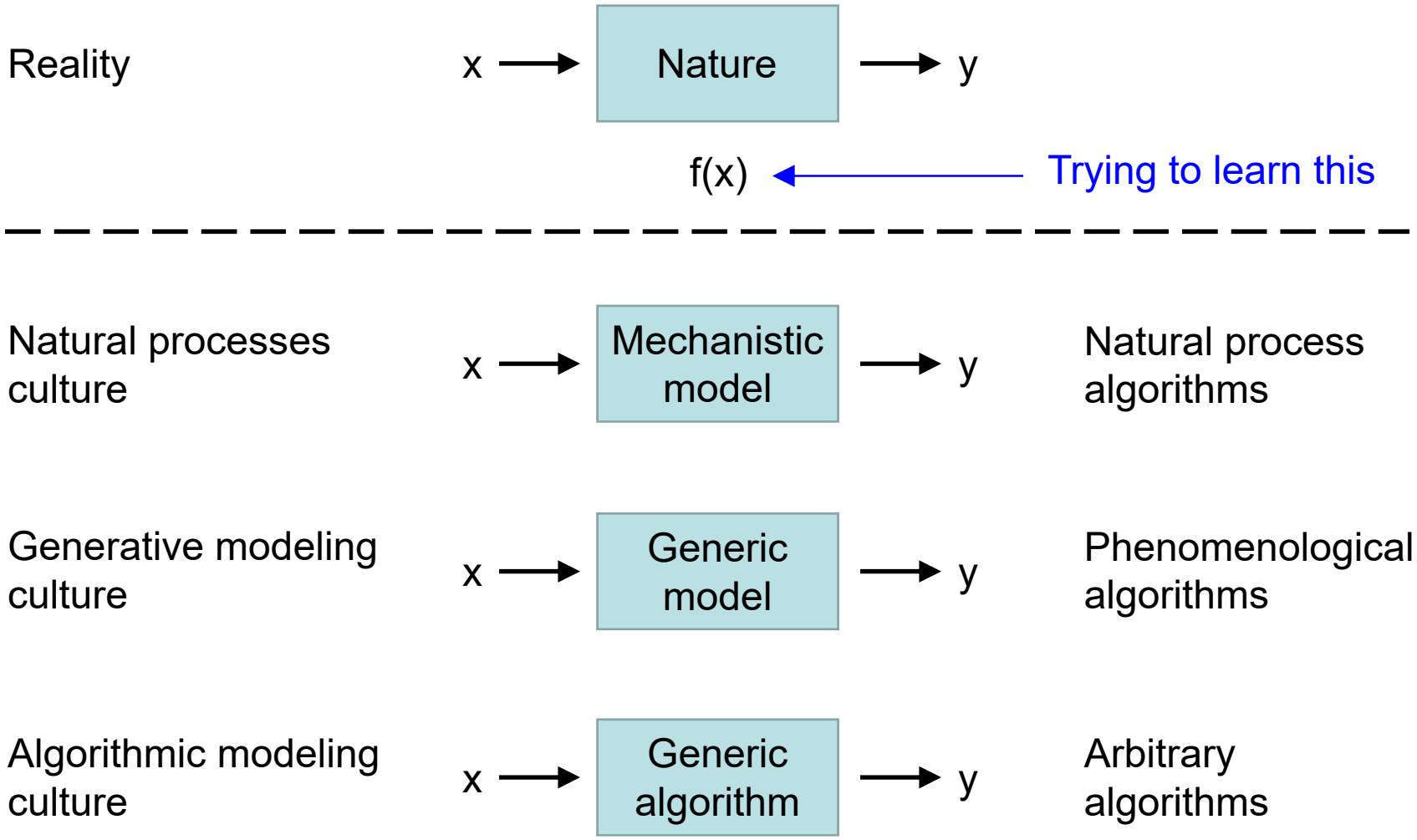
Today

- 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

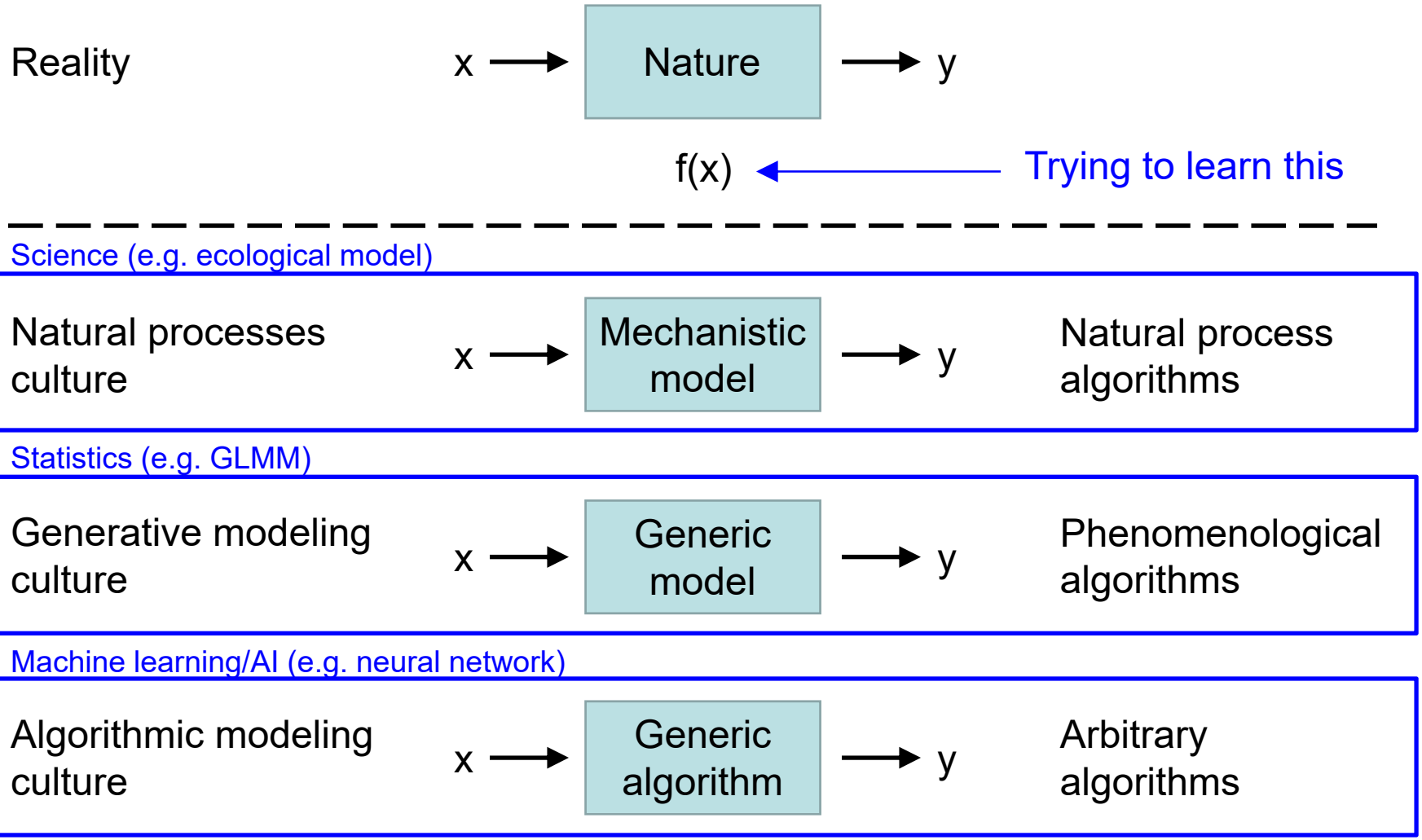
Data science cultures



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

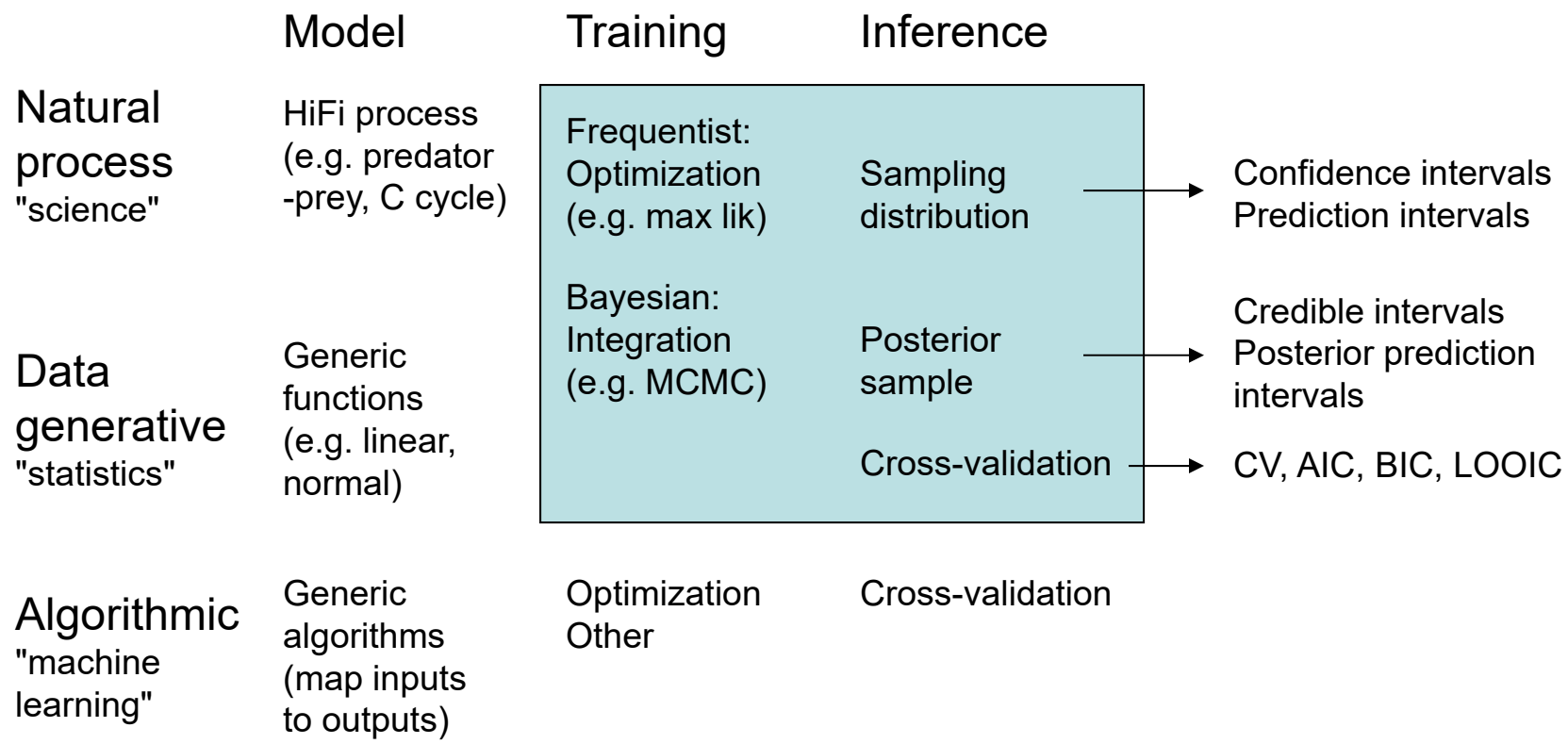
Algorithms in data science

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

Modeling with data

Algorithm classes

Modeling culture



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, predictionist
- Emphasize multi-level linear models
- We'll start by
  - 1) learning how to program algorithms
  - 2) considering 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 preclass work. Collaborative learning is not only allowed but **encouraged** in this class! FERPA compliant.

Texts

- All on Google Drive or open source
- I'll provide all materials.
- This one is worth obtaining eventually:
- McElreath, R (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan.* 2nd Ed.

Grading

- GitHub portfolio
- 50% continuous Github code commits
- 50% individual assignment

Homework

- Posted to GitHub
  - “preclass4wed”
- Update R & R studio
- Set up GitHub
- Intro to programming