EBIO 5460
Machine Learning for Ecology

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/EBIO5460Spring2026
- Organization, syllabus, timetable
- Slides, code, homework
- You'll also submit your work here
- Main resource links: README.md

Slides for today

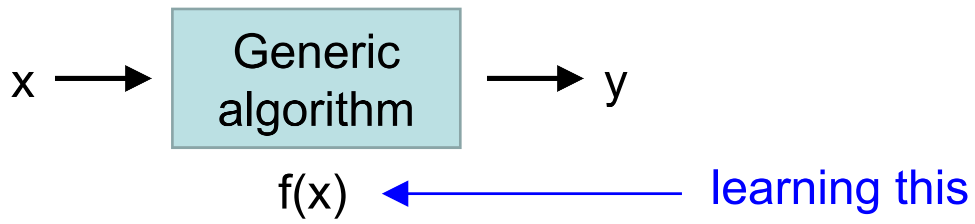
- github.com/EBIO5460Spring2026
- Go to repositories
- Open class-materials
- 01_1_slides_thu_intro

Today

- What is machine learning? (2 mins)
- Introductions (20 mins)
- Syllabus & how we'll do the class (20 mins)
- Where does machine learning fit in to data science & algorithms?

What is machine learning?

- Working definition
- Using generic algorithms to predict outputs y from inputs x
- Emphasis: prediction, predictive skill



Examples in ecology

- Species distribution models (SDMs)
  - predicting the spatial distribution of a species from environmental variables
- Counting penguins in all of Antarctica from satellite imagery
- Identifying mammal species in camera trap images in the Serengeti
- Identifying bird species from audio recordings
- Do you have any examples?

Introductions

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

Syllabus

- We worked through it here:
- 00_syllabus.md

Learning goals

- Understand the fundamental concepts and algorithms that underpin most of machine learning
- Become confident to use machine learning algorithms in your research
- Gain a broad overview of how ecologists are currently using machine learning algorithms to revolutionize ecological research

Learning format

- Coding demonstration in live lectures. Sometimes short videos.
- Collaborative learning. Work in small groups or share in small groups.
- Piazza: collaboratively discuss lectures and assignments. Link @ README.md. FERPA compliant. Collaborative learning is not only allowed but **encouraged** in this class!

Computing

- Install/Update R ... and/or
- Install/Update Python (suggest: via conda)
- Install/Update IDE(s)
    - e.g. Positron, Rstudio, VSCode

Text

- James et al (2021). An Introduction to Statistical Learning: With Applications in R (or Python), 2nd ed.
- https://www.statlearning.com/
- Free download

Grading

- For completion
- Assignments 35%
- Discussions 20%
- Lead discussion 10%
- Individual project 35%

Week 1-9 assignments

- Will be posted to GitHub
- 01_3_homework_to_get_started.md
  - Learn Git
  - Review algorithms
  - Set up GitHub
- Week 1 HW is not part of grade but needs to be done by Tuesday Week 2
- 4 assignments: 3 x 7%, 1 x 14% = 35%

Week 10-15 literature & project

- Each person will lead a discussion
  - paper of your choice
  - ca 2% per discussion
- Individual project
  - data project or literature review
  - data set or topic of your choice
  - presentation in finals week 15%
  - code or paper submission 20%

Data Science

- Workflows and algorithms to learn from data
- Part 1 (e.g. Fall semester 2025): Fundamental algorithms and concepts
  - https://github.com/EBIO5460Fall2025/class-materials
- Part 2 (this class) Machine learning

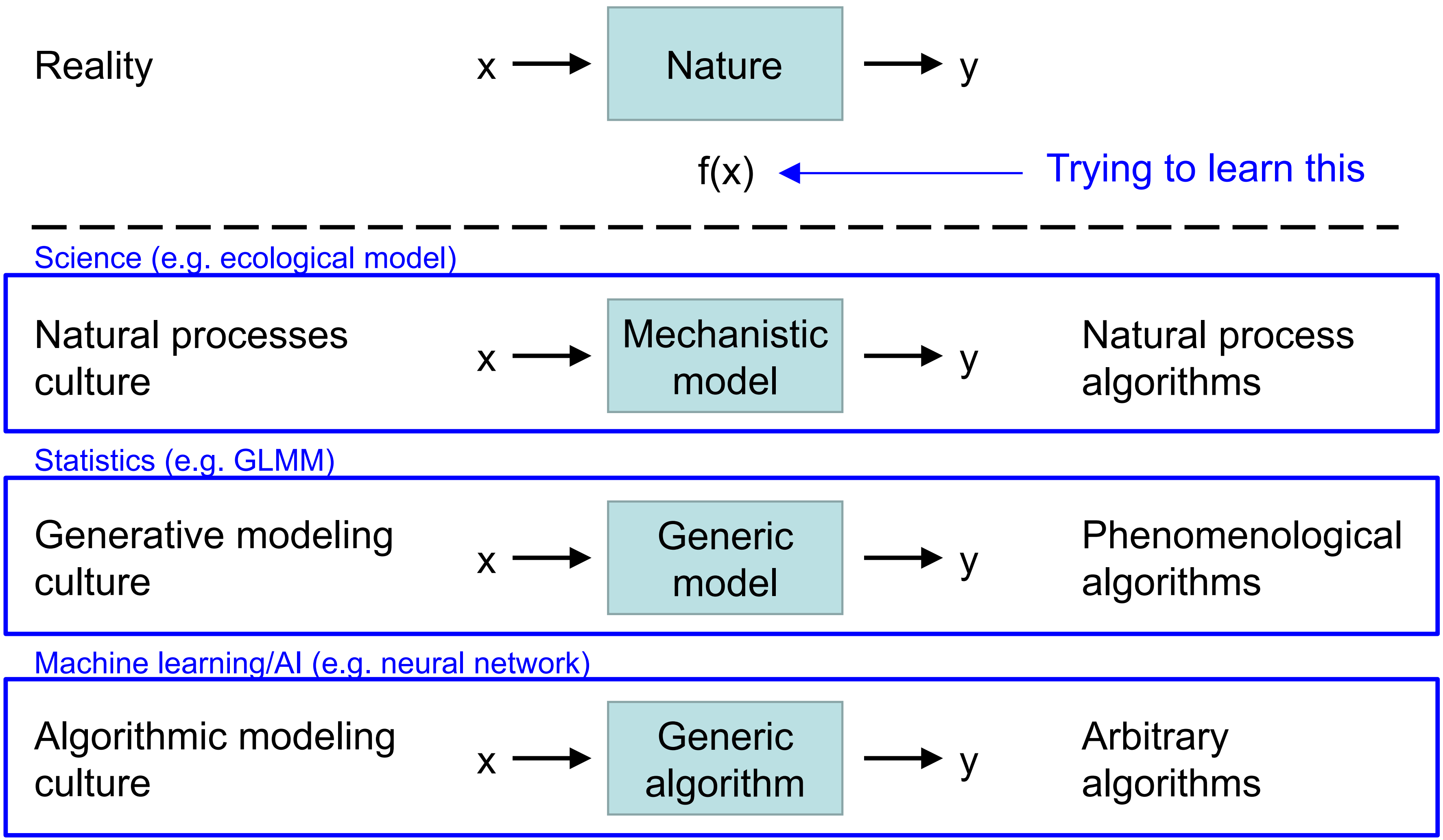
Data science cultures

Reality	x → Nature → y	
	f(x) ← Trying to learn this	
Natural processes culture	x → Mechanistic model → y	Natural process algorithms
Generative modeling culture	x → Generic model → y	Phenomenological algorithms
Algorithmic modeling culture	x → Generic algorithm → y	Arbitrary algorithms

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)

Algorithm

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

- Algorithms are fundamental: most math in statistics is a solution or approximation to a data-generating algorithm

Algorithms in data science

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

Algorithms in data science

- Model algorithm
  - The function f(x)!
  - Often equations, sometimes rules
  - Usually has parameters
  - e.g. $y = a + b x$
- Training algorithm
- Inference (reliability) algorithm

Algorithms in data science

- Model algorithm
- Training algorithm
  - Algorithm to train a model algorithm on data
  - syn. model fitting, calibration, parameter estimation
  - e.g. Nelder-Mead simplex optimization, gradient descent
- Inference (reliability) algorithm

Algorithms in data science

- Model algorithm
- Training algorithm
- Inference (reliability) algorithm
  - first, what kind of inference?

Statistical inference

- Judge the accuracy of an estimation or prediction algorithm
  - Efron & Hastie 2016
- Reliability
- Uncertainty

ISO definition of accuracy: the closeness of a measurement to the true value
Two components: bias, variance

Different inference problems

Estimation
Infer a property of a population (e.g. mean) from a sample

Model comparison
Infer the data generating process from among a set of candidate data-generating processes

Hypothesis test (association)
Infer that y is associated with x

Causation
Infer that x causes y
Infer the size of an effect due to an experimental intervention (estimation)
Infer that an experimental intervention had an effect (H-test)

Prediction

Machine learning

Predict the value of a new observation or population state (extrapolation or interpolation)
Predict the population state in the future (forecast/extrapolation)

Algorithms in data science

- Model algorithm
- Training algorithm
- Inference (reliability) algorithm
  - looking back: consider all the ways data could have happened (mechanistic, generative)
  - looking forward: predict new data and test against them (mechanistic, generative, algorithmic)

Machine learning

Machine learning doesn't care about the possible ways data could have happened. It just cares about how well an algorithm predicts.

Algorithms review

algorithms4ds_review.md

Modeling with data

Algorithm classes

Modeling culture

	Model	Training	Inference	
Natural process "science"	HiFi process (e.g. predator -prey, C cycle)	Frequentist: Optimization (e.g. max lik)	Sampling distribution	→ Confidence intervals Prediction intervals
Data generative "statistics"	Generic functions (e.g. linear, normal)	Bayesian: Integration (e.g. MCMC)	Posterior sample	→ Credible intervals Posterior prediction intervals
			Cross-validation	→ CV, AIC, BIC, LOOIC
Algorithmic "machine learning"	Generic algorithms (map inputs to outputs)	Optimization Other	Cross-validation	→ CV, AUC, ROC

Machine Learning

- Supervised learning (this semester)
  - labeled response data
  - compare prediction to labeled (“known”) response
- Unsupervised learning
  - unlabeled response data, discover patterns
  - aka traditional topic: “multivariate analysis”
  - clustering, ordination etc