

EBIO 5460

Machine Learning for Ecology

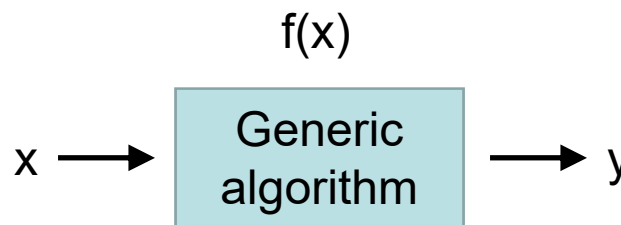
Dr Brett Melbourne
Associate Professor, EBIO
brett.melbourne@colorado.edu
Office hours: Any time by appointment
Office: Ramaley N336 and Zoom
Pronouns: he, him, his

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

Git & GitHub

- Class Github organization
- Bookmark this:
- <https://github.com/EBIO5460Spring2024>
- Organization, syllabus, timetable
- Code, homework

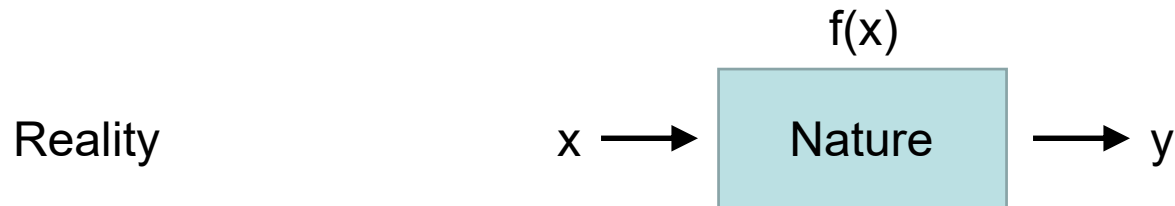
Week 1-8 assignments

- Will be posted to GitHub
- First tasks:
 - Update R & R studio
 - Set up GitHub

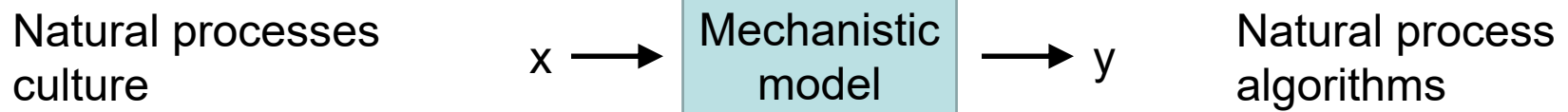
Data Science

- Workflows and **algorithms** to learn from data
- Part 1 (e.g. Fall semester 2022):
Generalized linear models up to Bayesian
multilevel models
 - <https://github.com/EBIO5460Fall2022/class-materials>
- Part 2 (this class) Machine learning

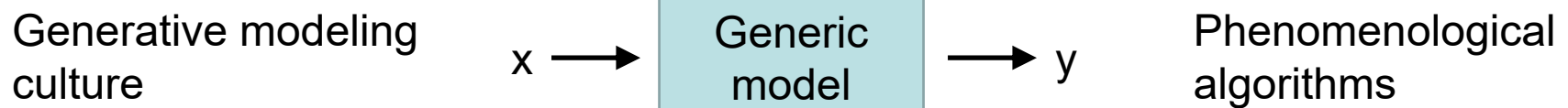
Data science cultures



Part 3: Ecological models



Part 1: GLMMs



Part 2: Machine learning



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

Algorithms in data science

- Model algorithm
 - Often equations, sometimes rules
 - Often 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
 - e.g. Nelder-Mead simplex optimization
- 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

Algorithms review

01_2_algorithms4ds_review.md

Modeling with data

Algorithm classes

Modeling culture

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