Introduction(week 1, part 1)

2 Jan 2023

Table of contents

## integrity

* Stack Overflow and ChatGPT
* group work, copying

## technical skills/tools

* reproducibility: version control (Github)
* machinery
  + R, Julia
  + VSCode
  + reproducibility: Quarto or Sweave or Jupyter notebook
* *command line bullshittery* (“bullshit (read: diagnosing and debugging weird things) is a part of life in the world of computers”) (Adar 2015)

## Goals

* understand theory behind (novel) methods
* read papers
* (choose methods)
* implement methods
* read/understand/improve existing methods

## About me

* weird background (physics u/g, Zoology PhD, epidemiological modeling)
* biases:
* interested in scientific answers (but see Navarro (2019))

## Themes

* preventing overfitting
  + shrinkage
  + regularization
  + penalization
  + dropouts (NN)
  + constraining tree depth
  + learning rate (boosting)
  + early stopping
  + priors
* basis construction/feature engineering
  + splines
  + GPs
  + tree splits
  + wavelets
  + Fourier bases
  + neural network architecture
* optimization
  + gradient descent, SGD
  + IRLS
  + BFGS etc.
* loss functions
* assessment and diagnostics
  + cross-validation (blocked, etc.)
  + bootstrap
* sparsity
* continuous vs discrete structures

Adar, Eytan. 2015. “On the Value of Command-Line ‘Bullshittery’.” *Medium*. <https://medium.com/@eytanadar/on-the-value-of-command-line-bullshittery-94dc19ec8c61>.

Navarro, Danielle. 2019. “Science and Statistics.” Aarhus University. <https://slides.com/djnavarro/scienceandstatistics>.