Workshop 2

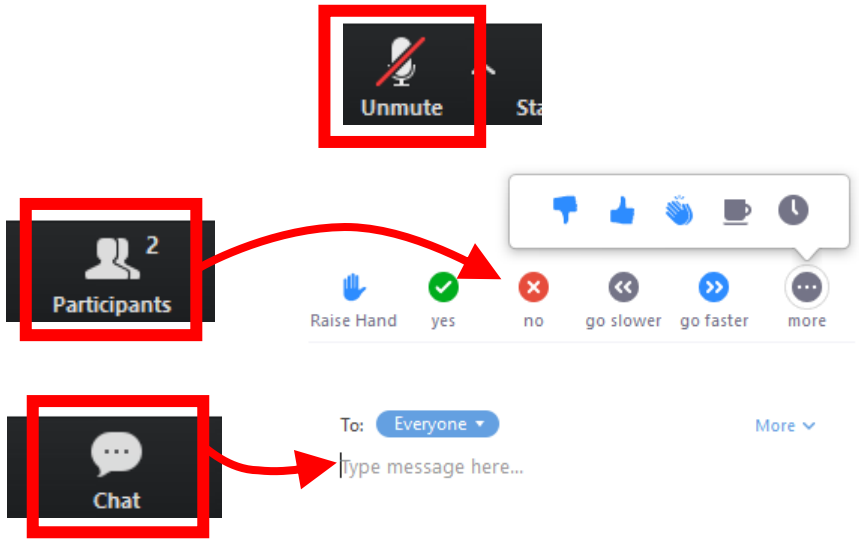
COMP90051 Machine Learning

Semester 1, 2021

Zoom in workshops

To make the most of online workshops, please:

- ensure your mic is muted unless you have the floor
- make use of non-verbal feedback (e.g. raise hand)
- use chat to communicate discretely (e.g. send a pm)
- participate in class discussion



Agenda

1. Icebreaker
2. Python ecosystem for ML
3. Refresher: Bayes' theorem
4. Worksheet on Bayesian inference

Icebreaker

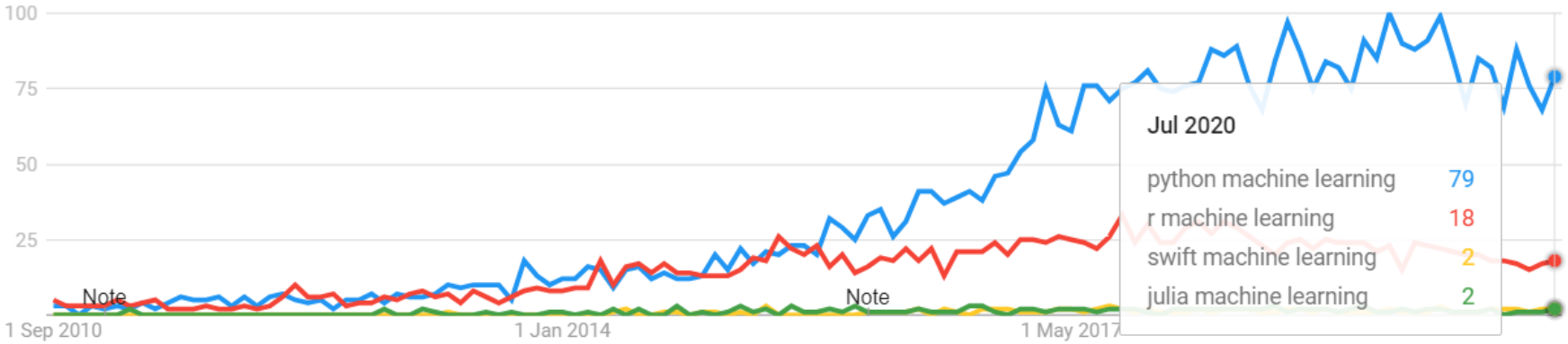
Is your system ready to go?

- You should have installed Anaconda on your system before today's workshop. If not, please install it now.
- Anaconda is a Python distribution tailored for scientific computing
- Most of the packages we need are installed by default
- Worksheets will be distributed as Jupyter Notebooks

Python ML ecosystem

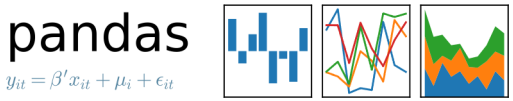
Why Python for ML?

- It's popular in academia and industry
- There's a huge collection of open-source packages/libraries
- Acts as a glue between low-level libraries so can be relatively fast



Top 5 libraries for beginners to master

- Library for working with large multidimensional arrays
- High-level functions for arrays



- Library for analysis and manipulation of tabular data
- Provides similar functionality to DataFrames and dplyr in R

- Machine learning library
- Includes implementations of most models covered in this course (exception: neural nets)

- Scientific computing library
- Functionality includes: statistics/random number generation, linear algebra, optimisation, special functions, integration

- 2D plotting library
- Provides similar interface to MATLAB

We'll see some of these libraries later…

Deep learning frameworks

Probabilistic programming frameworks

Bayesian inference

Recall from Lecture 2c

Tools of probabilistic inference

Bayes

- Bayesian probabilistic inference
  - Start with prior $\mathrm{P}(\theta)$ and likelihood $\mathrm{P}(X|\theta)$
  - Observe data $X = x$
  - Update prior to posterior $\mathrm{P}(\theta|X = x)$
- Primary tools to obtain the posterior
  - Bayes Rule: reverses order of conditioning

$$P(\theta|X = x) = \frac{P(X = x|\theta)P(\theta)}{P(X = x)}$$

  - Marginalisation: eliminates unwanted variables

$$\mathrm{P}(X = x) = \sum_t P(X = x, \theta = t)$$

These are general tools of probability and not specific to Bayesian stats/ML

This quantity is called the evidence

Worksheet 2