### GENERALIZATION

I have data, and I've fitted my model to it. What can I say about the future?

To be able to say *anything*, let's assume there's a "law of nature", i.e. a true probability model that governs both.

What can we usefully say about the future (or equivalently about the true probability model)?

Parameter uncertainty

Prediction uncertainty

Choosing between models (e.g. models of varying complexity)

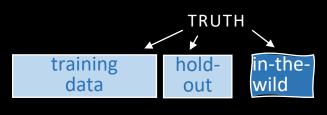
Model adequacy

## TRUE PROBABILITY MODEL

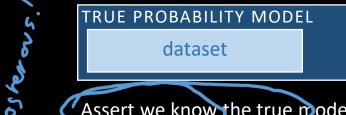
wild

FITTED MODEL

dataset



Approximate in-the-wild data by a holdout dataset



Assert we know the true model (apart for its parameter values, which we learn by Bayes's rule)



Simulate parallel-universe datasets, and thus produce an ensemble of fitted models

#### **EMPIRICISM**

#### **BAYESIANISM**

#### **FREQUENTISM**



posterior distribution for parameters
LECTURE 10

ensemble of parameter estimates LECTURE 12



pangerous!
(my regearch)

... derived from parameter uncertainty
EX2 Q5b

an ensemble of predictions, one from each fitted model EX3 Q7

choose the model that has the highest holdout likelihood LECTURE 11

which is just another parameter LECTURE 11

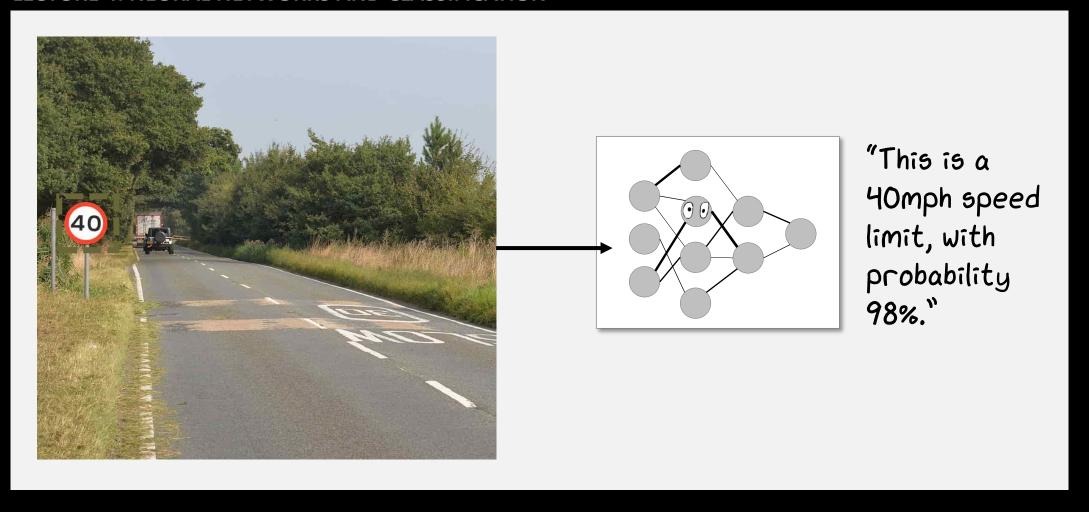
X





hypothesis testing

#### **LECTURE 4: NEURAL NETWORKS AND CLASSIFICATION**

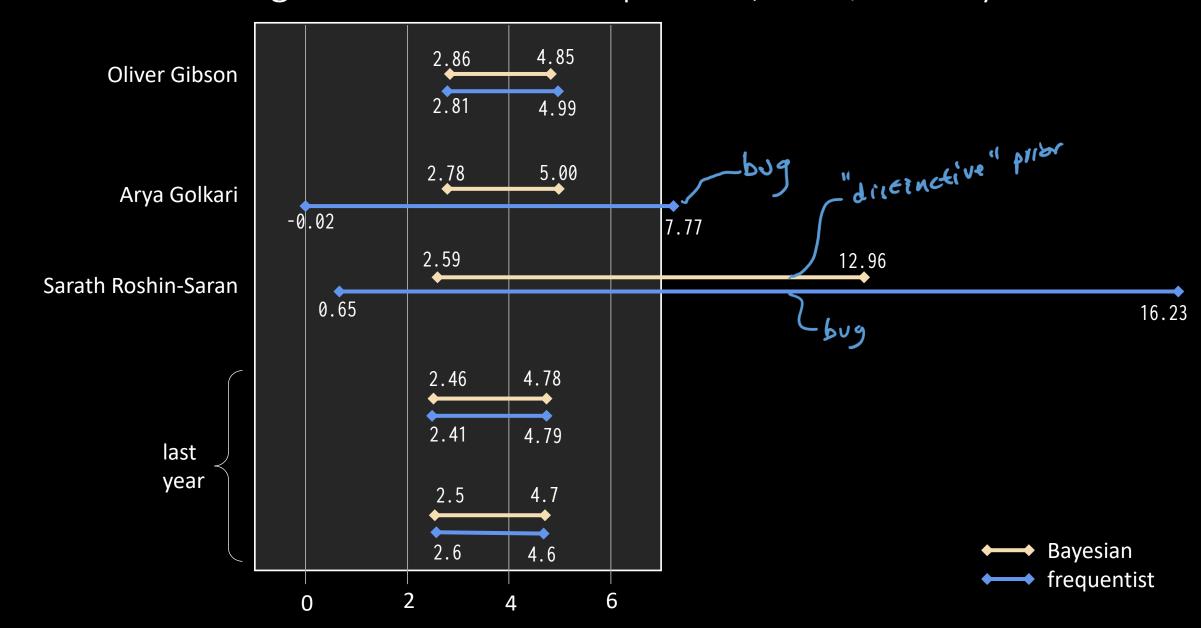




This "98%" is the model's estimated probability. It's not the same thing as uncertainty — is it  $98\pm1\%$  or  $98\pm40\%$ ?

No one has worked out how to extract prediction uncertainty from a neural network.

# Challenge: find a 95% confidence interval for the rate of temperate increase in Cambridge from 1985 to the present, in °C/century



## Is ChatGPT empiricist, Bayesianist, or frequentist?

I have a dataset of student marks on an exam (numbers in the range 0 to 20). I also have a record of the gender of the student (male / female / other). I want to learn if there is gender bias. Give me code to do this.

Sure — here's a clean, reproducible way to test for **gender bias in exam marks** using Python. I'll give you code that both:

- 1. Explores the data visually and statistically, and
- **2. Formally tests** for significant differences in the mark distributions across genders.

•••

To test if gender groups differ in their distributions:

- a. 2-group parametric test (if marks are roughly normal and variances similar)
- b. 2-group nonparametric test (if not normal)
- c. multi-group parametric and nonparametric tests