Model Comparison Introduction

Bayesian Data Analysis

Steve Buyske

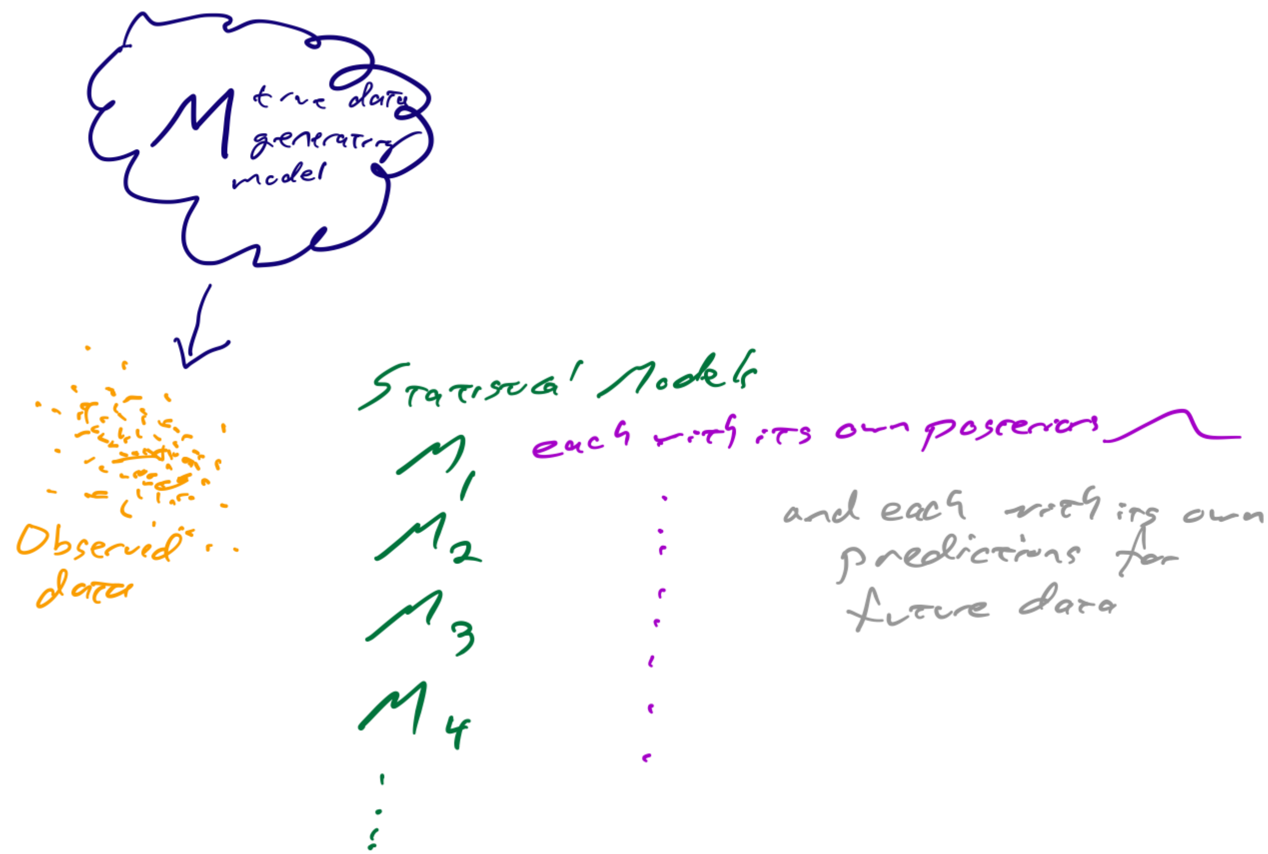
Two goals

- When we have several possible models for the same data, how should we compare them?
  - Emphasis on *comparison*, not *selection*.
  - We may want to include non-statistical criteria in selecting a model, but it would be useful to have statistical criteria to judge if one model is much better than another.
- Why do we fit a statistical model to data?
- Generally, one of two reasons:
  - For scientific understanding
  - For prediction

Two goals cont.

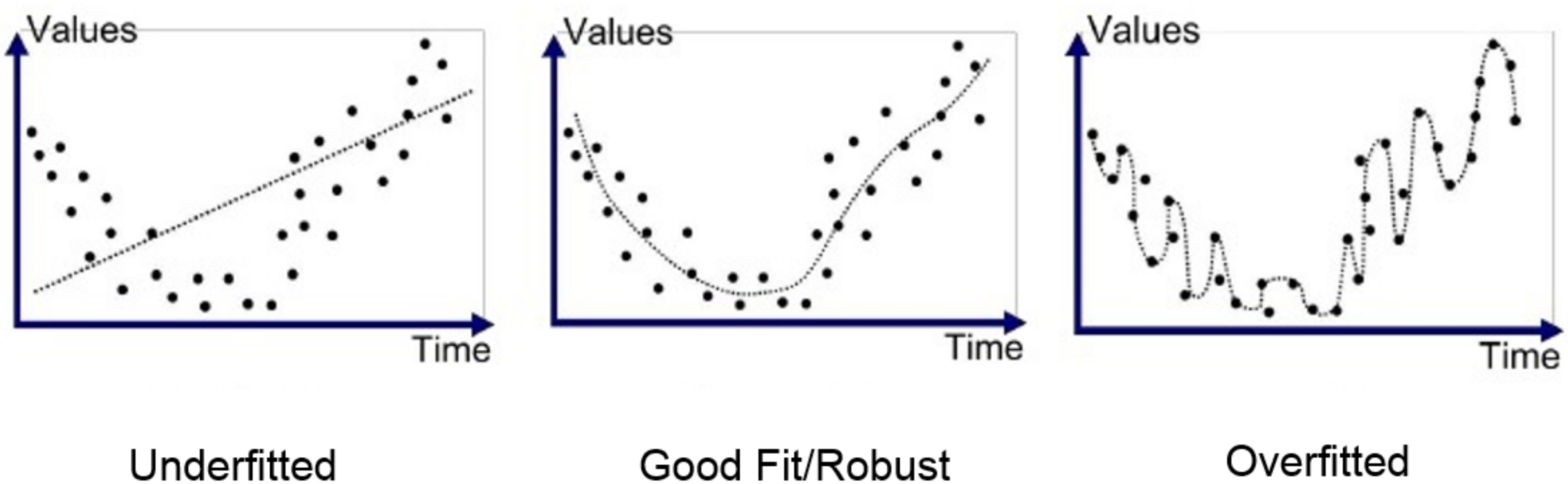
- These two goals may lead to the same result (if x_1, and only x_1, actually determines y, then both goals are met with the model `y ~ x1`).
- But the goals can also be in conflict
  - For scientific understanding, we often want the simplest model that adequately "explains" the data
  - For prediction, more complicated—but not too complicated—models can often do better at prediction than simpler models.
- Worth remembering that without designed experiments the kinds of models we are talking about do not actually establish causality.

A picture to keep in mind



Underfitting and overfitting

- When we fit a statistical model, we want to avoid both underfitting and overfitting.
  - In underfitting our model is too simple and in overfitting our model is too complex.



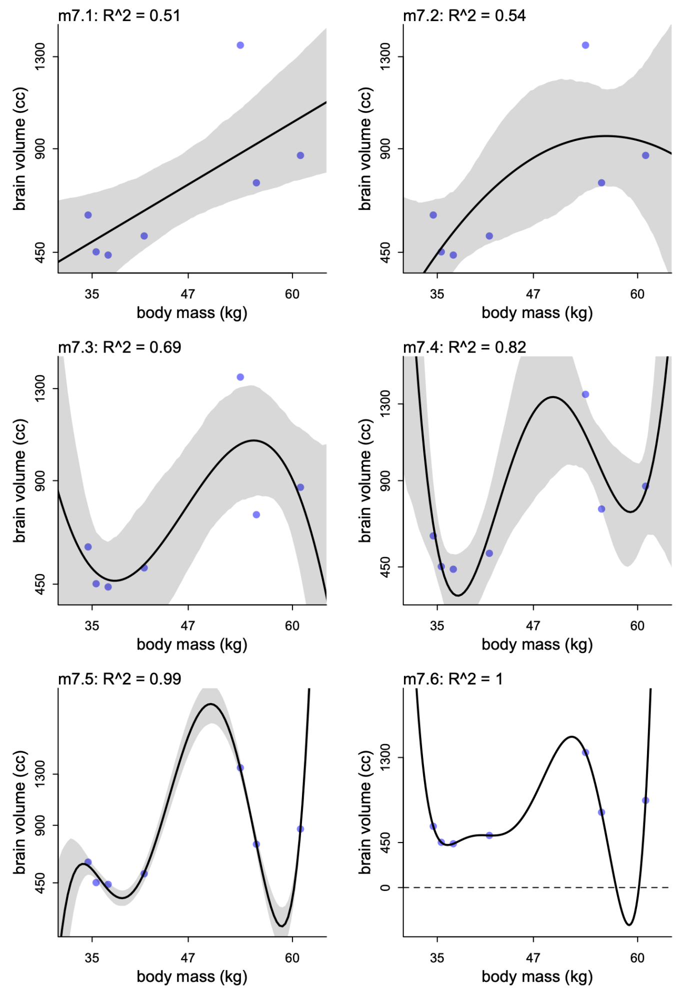


FIGURE 7.3. Polynomial linear models of increasing degree for the hominin data. Each plot shows the posterior mean in black, with 89% interval of the mean shaded. R^2, is displayed above each plot. In order from top-left: First-degree polynomial, second-degree, third-degree, fourth-degree, fifth-degree, and sixth-degree.

Underfitting and overfitting cont.

- If you underfit, you are leaving out variables that carry useful information.
- If you overfit, also known as "fitting the noise", you will get better prediction in the current sample but will predict poorly on future samples.
- In general, the Bayesian framework is less susceptible to overfitting than a frequentist framework.
    - Priors "regularize" the fit, by giving lower probability to extreme values of the parameters.
    - In the Bayesian world, if the model is too big, then the posteriors will be centered near zero.
    - The more variables there are in the model, the smaller the prior probability will be for meaningfully non-zero values of any particular parameter, and so—without strong evidence—the smaller the posterior for meaningfully non-zero values of any particular parameter.

How can you tell?

- We would like a measure that would indicate that a model fits well while avoiding overfitting.
- One way to evaluate a model is through the accuracy of its predictions.
- Sometimes we care about this accuracy for its own sake
    - e.g., evaluating a forecast
- Often predictive accuracy is valued as a way of comparing different models, and we will use it in that way.

- It is a mistake to think a single number is sufficient to make a decision about a model's value with respect to a dataset.
- Nonetheless, we would like a unbiased, accurate estimate of the prediction error of a *new sample*.
    - Other things being equal, a model that predicts better on new data is considered better.
- We would also like it to require minimal additional computation.
- And we would also like it to be valid over a large class of models.
- **There is not an ideal measure**, but there are some pretty good ones.

- We will proceed in two steps:
  - Settle on a measure of model accuracy
  - Adjust that model to be a measure of *future* accuracy