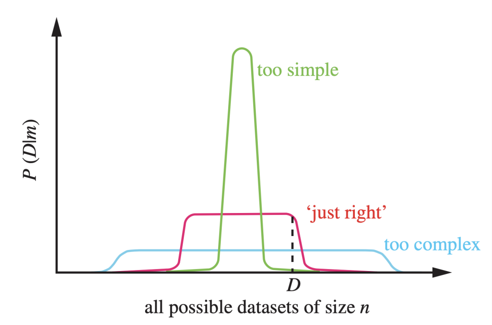
Bayesian Linear Regression

June 5, 2021

Bayesian Occam's Razor

Remember this slide?

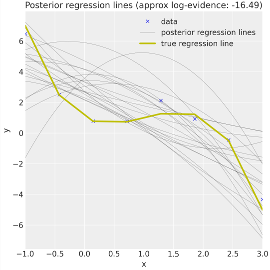


A *complex* model (shown in blue) spreads its mass over many more possible datasets, whereas a *simple* model (shown in green) concentrates its mass on a smaller fraction of possible data. Because probabilities have to sum to one, the complex model spreads its mass at the cost of not being able to model simple datasets as well as a simple model—this normalization is what results in an automatic Occam razor. Given any particular dataset, here indicated by the dotted line, we can use the marginal likelihood to reject both overly simple models, and overly complex models.

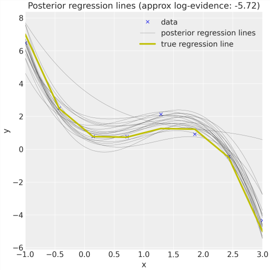
Ghahramani, Z. (2013). Bayesian non-parametrics and the probabilistic approach to modelling. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 371(1984), 20110553.

Bayesian Occam's Razor

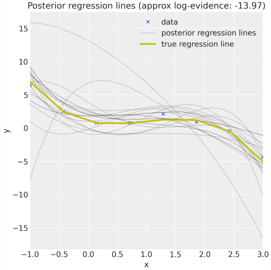
I generated $n = 8$ data points from a cubic distribution and used NUTS to fit Bayesian polynomials of various orders p.



Quadratic model ($p = 2$)



Cubic model ($p = 3$)

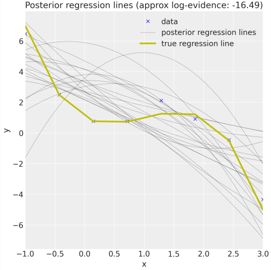


Quartic model ($p = 4$)

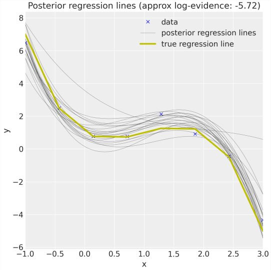
Observations

Bayesian Occam's Razor

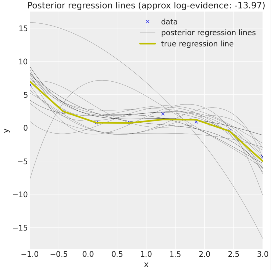
I generated $n = 8$ data points from a cubic distribution and used `NUTS` to fit Bayesian polynomials of various orders p.



Quadratic model ($p = 2$)



Cubic model ($p = 3$)



Quartic model ($p = 4$)

Observations

- Bayesian model selection works well here! The true (cubic) model has the highest evidence. The evidence is lower for models that are underfit (quadratic) or overfit (quartic).
- Correspondingly, typical draws from the posterior model best match the true data generating process.