Distributions over parameters and functions

Carl Edward Rasmussen

October 13th, 2016

Key concepts

- In a parametric model, the model is represented using parameters
- a distribution over parameters implies a distribution over functions
- In Bayesian inference, we marginalize over parameters to make predictions
- Question: could we work directly in the space of functions?

Priors on parameters induce priors on functions

A model $\mathcal{M}$ is the choice of a model structure and of parameter values.

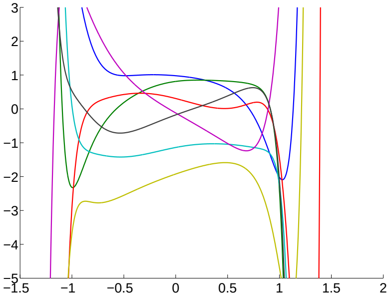
$$f_{\mathbf{w}}(x) = \sum_{m=0}^{M} w_m \, \phi_m(x)$$

The prior $p(\mathbf{w}|\mathcal{M})$ determines what functions this model can generate. Example:

- Imagine we choose $M = 17$, and $p(w_m) = \mathcal{N}(w_m;\ 0, \sigma_{\mathbf{w}}^2)$.
- We have actually defined a prior distribution over functions $p(\mathbf{f}|\mathcal{M})$.

This figure is generated as follows:

- Use polynomial basis functions, $\phi_m(x) = x^m$.
- Define a uniform grid of $n = 100$ values in x from $[-1.5, 2]$.
- Generate matrix $\mathbf{\Phi}$ for $M = 17$.
- Draw $w_m \sim \mathcal{N}(0, 1)$.
- Compute and plot $\mathbf{f} = \mathbf{\Phi}_{n\times 18}\, \mathbf{w}$.



Nuisance parameters and distributions over functions

We've seen that distributions over parameters induce distributions over functions.

We've set up a scheme where we

- first set up a model in terms a parameters
- then marginalize out the parameters

Typically, we're not really interested in parameters, we're interested in predictions.

The parameters are a nuisance.

Could we possibly work directly in the space of functions?

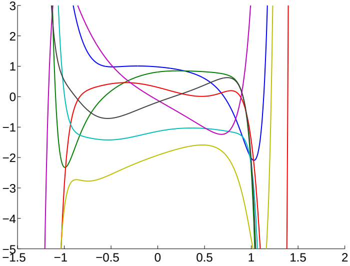
- simpler inference
- better understading of the distributions over functions

Posterior probability of a function

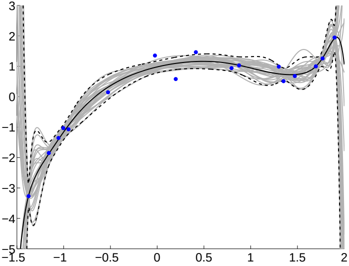
Given the prior functions $p(\mathbf{f})$ how can we make predictions?

- Of all functions generated from the prior, keep those that fit the data.
- The notion of closeness to the data is given by the likelihood $p(\mathbf{y}|\mathbf{f})$.
- We are really interested in the posterior distribution over functions:

$$p(\mathbf{f}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{f})\, p(\mathbf{f})}{p(\mathbf{y})} \qquad \text{Bayes Rule}$$

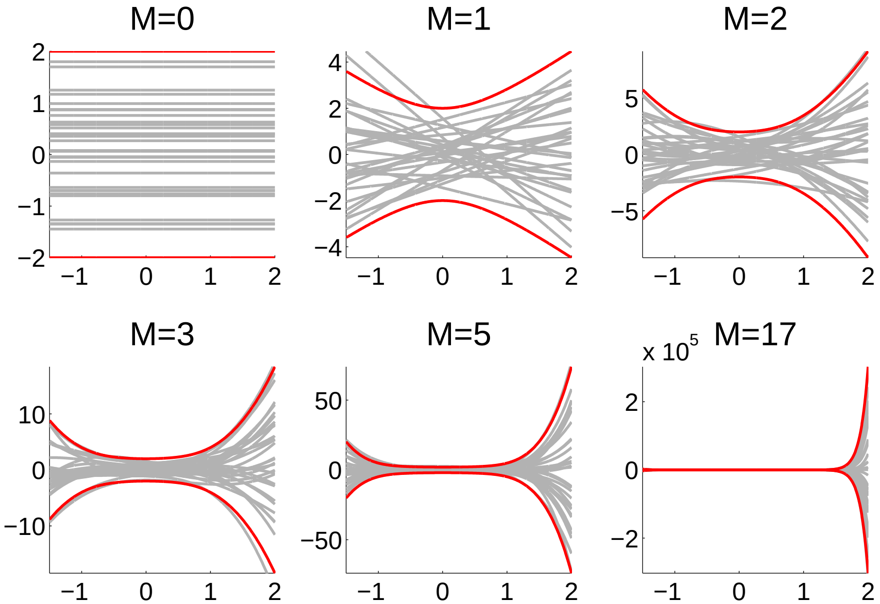


Some samples from the prior

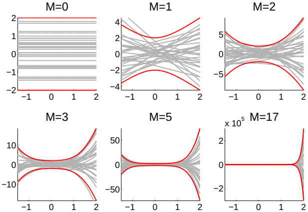


Samples from the posterior

Are polynomials a good prior over functions?

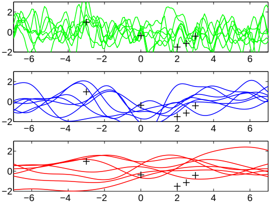


A prior over functions view



We have learnt that linear-in-the-parameter models with priors on the weights *indirectly* specify priors over functions.

True... but those priors over functions might not be good.



... why not try to specify priors over functions *directly*?

What? What does a probability density over functions even look like?