Distributions over parameters and functions

Carl Edward Rasmussen

July 1st, 2016

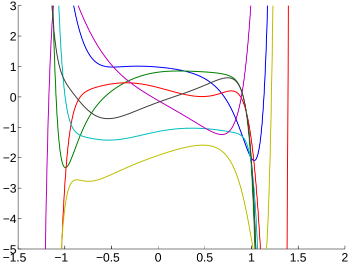
Key concepts

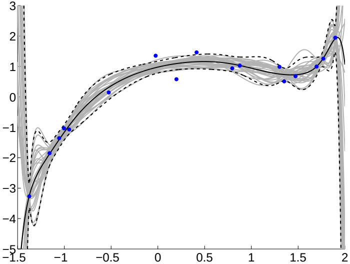
Multiple explanations of the data

- We do not believe all models are equally probable to explain the data.
- We may believe a simpler model is more probable than a complex one.

Model complexity and uncertainty:

- We do not know what particular function generated the data.
- More than one of our models can perfectly fit the data.
- We believe more than one of our models could have generated the data.
- We want to reason in terms of a set of possible explanations, not just one.



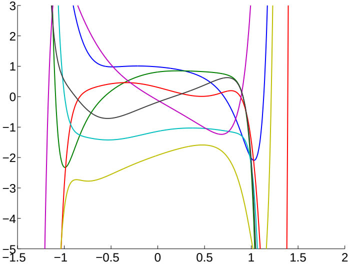


Posterior probability of a function

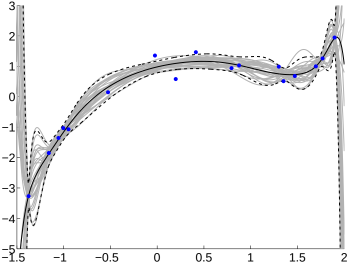
Given the prior functions $p(\mathbf{f})$ how can we make predictions?

- 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

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 $\mathbf{f} = \mathbf{\Phi}_{n \times 18}\, \mathbf{w}$.