Faculty of Science

Bayesian Interpolation

Helene Ausar, Morten Hels and Jens Rix

Bayesian Interpolation

David J. C. MacKay*
Computation and Neural Systems, California Institute of Technology 139–74, Pasadena, CA 91125 USA

Although Bayesian analysis has been in use since Laplace, the Bayesian method of *model-comparison* has only recently been developed in depth. In this paper, the Bayesian approach to regularization and model-comparison is demonstrated by studying the inference problem of interpolating noisy data. The concepts and methods described are quite general and can be applied to many other data modeling problems. Regularizing constants are set by examining their posterior probability distribution. Alternative regularizers (priors) and alternative basis sets are objectively compared by evaluating the *evidence* for them. "Occam's razor" is automatically embodied by this process. The way in which Bayes infers the values of regularizing constants and noise levels has an elegant interpretation in terms of the effective number of parameters determined by the data set. This framework is due to Gull and Skilling.

Outline

- Introduction to Bayesian Interpolation
  - Evidence
  - Occam's razor
- Interpolation of noisy data
  - Model comparison
- Conclusion

Introduction to Bayesian Interpolation

- How can we find the best possible interpolant?
- 1st level of inference

$$P(\mathbf{w} \mid D, \mathcal{H}_i) = \frac{P(D \mid \mathbf{w}, \mathcal{H}_i) P(\mathbf{w} \mid \mathcal{H}_i)}{P(D \mid \mathcal{H}_i)}$$

- 2nd level of inference

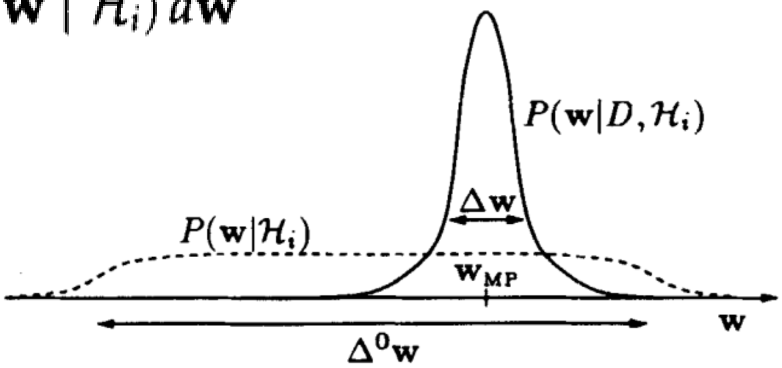
$$P(\mathcal{H}_i \mid D) \propto P(D \mid \mathcal{H}_i) P(\mathcal{H}_i)$$

Evidence for H_i

Occam factor

- Occam's razor: "*Among competing hypotheses, the one with the fewest assumptions should be selected.*"
- The evidence for H_i:

$$P(D \mid \mathcal{H}_i) = \int P(D \mid \mathbf{w}, \mathcal{H}_i) P(\mathbf{w} \mid \mathcal{H}_i)\, d\mathbf{w}$$



$$\underbrace{P(D \mid \mathcal{H}_i)}_{\text{Evidence}} \simeq \underbrace{P(D \mid \mathbf{w}_{MP}, \mathcal{H}_i)}_{\text{Best fit likelihood}} \; \underbrace{P(\mathbf{w}_{MP} \mid \mathcal{H}_i)\, \Delta\mathbf{w}}_{\text{Occam factor}}$$

$$\text{Occam factor} = \frac{\Delta\mathbf{w}}{\Delta^0\mathbf{w}}$$

Interpolation of noisy data

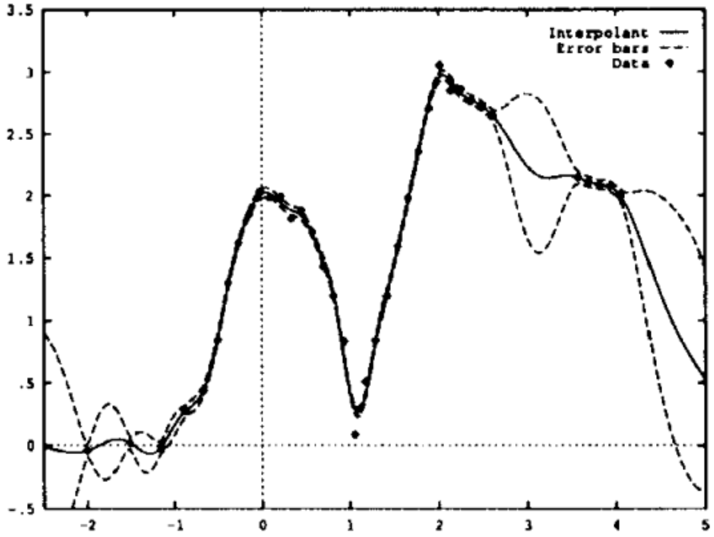
Interpolated function:

$$y(x) = \sum_{h=1}^{k} w_h \phi_h(x)$$

Regularizer (prior):

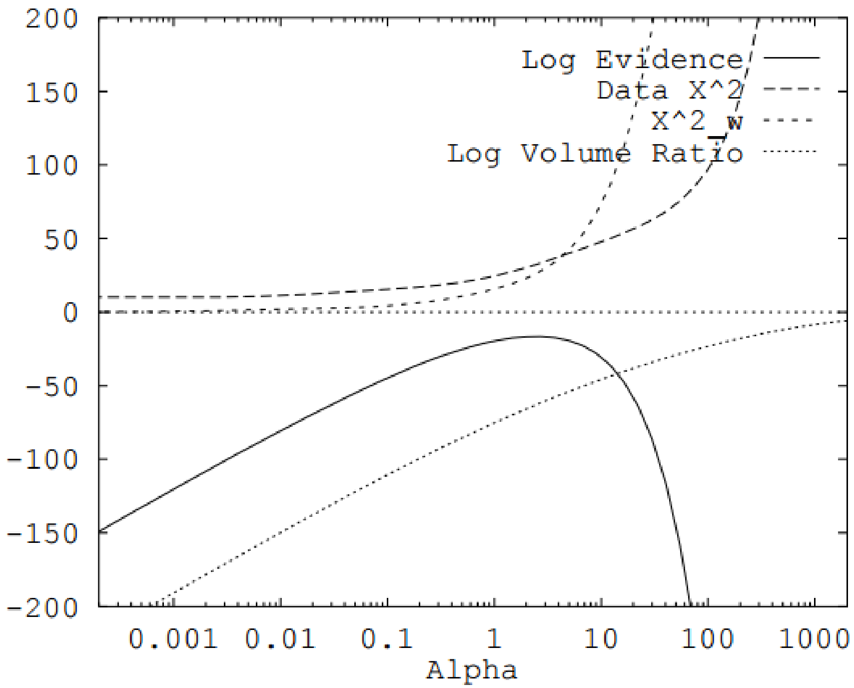
$$P(y \mid \mathcal{R}, \alpha) = \frac{\exp[-\alpha E_y(y \mid \mathcal{R})]}{Z_y(\alpha)}$$

$$E_y = \int y''(x)^2 dx$$



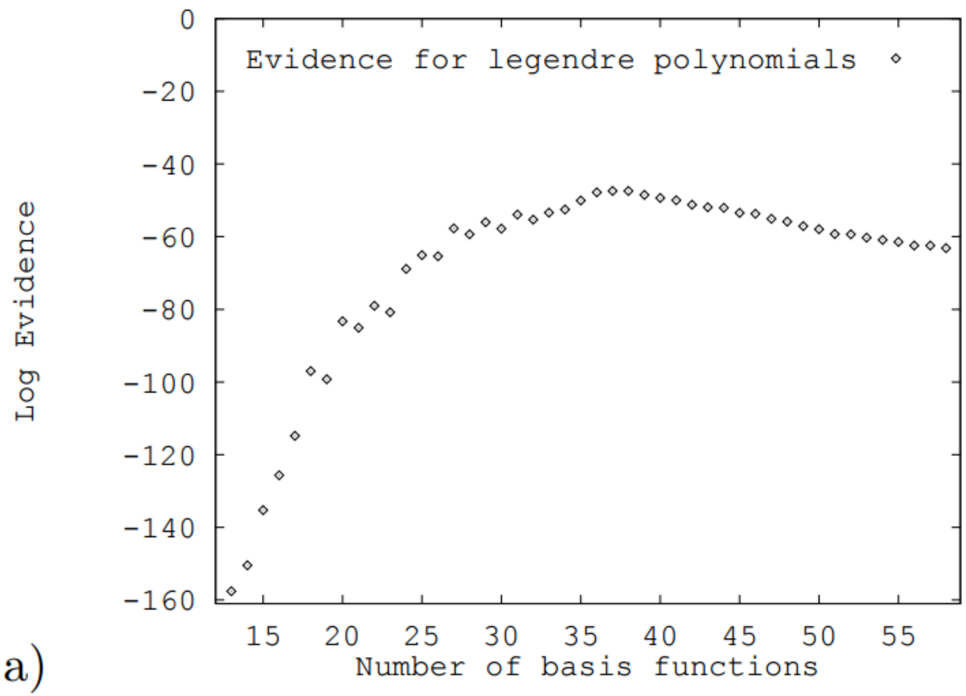
Evidence for the smoothening parameter α

$$P(\alpha, \beta \mid D, \mathcal{A}, \mathcal{R}) = \frac{P(D \mid \alpha, \beta, \mathcal{A}, \mathcal{R}) P(\alpha, \beta)}{P(D \mid \mathcal{A}, \mathcal{R})}$$



Evidence for the basis functions

$$P(\mathcal{A}, \mathcal{R} \mid D) \propto P(D \mid \mathcal{A}, \mathcal{R}) P(\mathcal{A}, \mathcal{R})$$



a)

Model comparison

- The highest value for the evidence gives the best model

Model	Data Set X: Best parameter values	Data Set X: Log evidence
Legendre polynomials	$k = 38$	-47
Gaussian radial basis functions	$k > 40$, $r = .25$	-28.8 ± 1.0
Cauchy radial basis functions	$k > 50$, $r = .27$	-18.9 ± 1.0
Splines, $p = 2$	$k > 80$	-9.5
Splines, $p = 3$	$k > 80$	-5.6
Splines, $p = 4$	$k > 80$	-13.2
Splines, $p = 5$	$k > 80$	-24.9
Splines, $p = 6$	$k > 80$	-35.8
Hermite functions	$k = 18$	-66

Conclusion

- The evidence is a ’solely’ data-dependent measure
- Different models can be ranked by their evidences
- The best model is the one that both fits the data well and is not too complex