Some remarks about Bayesian Infinite Regression and Gaussian Processes

Carl Edward Rasmussen
Max Planck Institute for Biological Cybernetics
Spemann Straße 38, 72076 Tübingen, Germany
carl@tuebingen.mpg.de

Non-parametric Workshop, NIPS*03, Dec 13th

Linear in the parameters regression

Goal:

Given a data set $\{\mathbf{x}_c, y_c | c = 1, \ldots, n\}$, we want to predict the output y_* given a new input $\mathbf{x}_*$.

Model:

Consider a linear model with fixed basis functions $\phi_1(\mathbf{x}), \ldots, \phi_m(\mathbf{x})$:

$$\mathbf{y} \;=\; \Phi\mathbf{w} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \;\sim\; \mathcal{N}(0, \sigma_w^2 I), \tag{1}$$

where all the ϕ's for each case form rows in Φ.

Prior:

Independent Gaussian prior on the weights:

$$\mathbf{w} \;\sim\; \mathcal{N}(0, \sigma_w^2 I). \tag{2}$$

Making predictions

Under the prior, the function values are jointly Gaussian:

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, Q), \qquad \begin{aligned} \boldsymbol{\mu} &= \langle \Phi\mathbf{w} + \boldsymbol{\varepsilon} \rangle = \mathbf{0}, \\ Q &= \langle \mathbf{y}\mathbf{y}^\top \rangle = \langle \Phi\mathbf{w}\mathbf{w}^\top\Phi^\top + \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top \rangle = \sigma_w^2 \Phi\Phi^\top + \sigma_n^2 I. \end{aligned} \tag{3}$$

Augmenting with the test case:

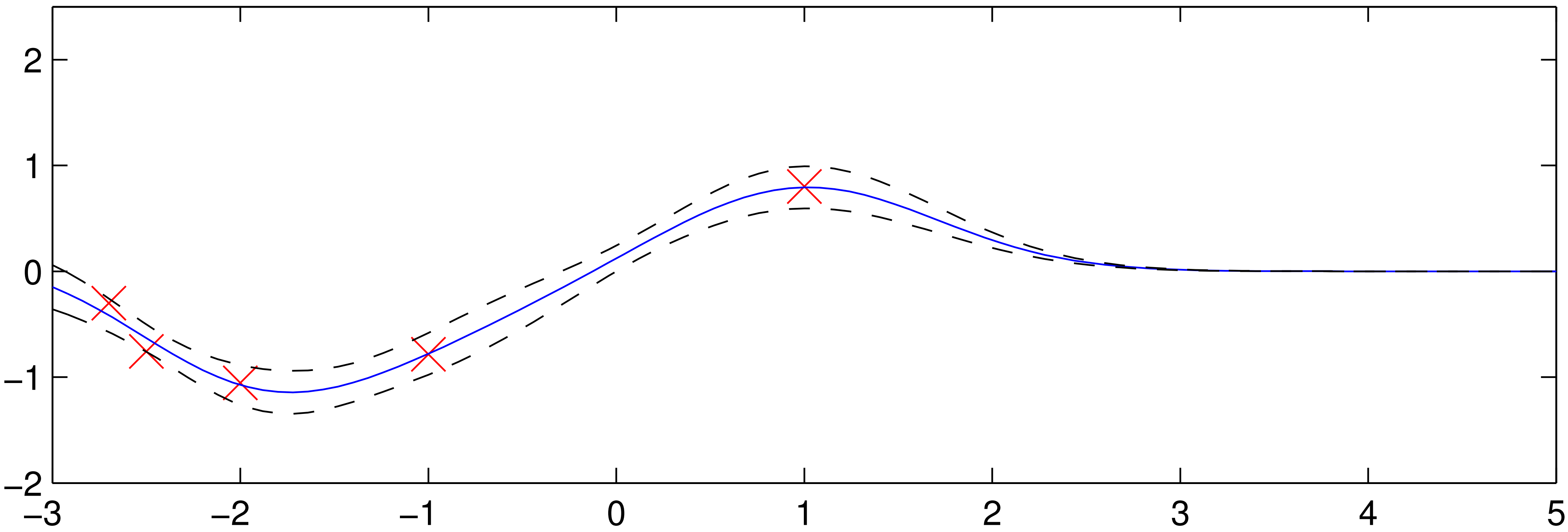
$$\mathbf{y}_{\mathrm{aug}} = (y_1, \ldots, y_n, y_*)^\top, \qquad \mathbf{y}_{\mathrm{aug}} \sim \mathcal{N}(\mathbf{0}, Q_{\mathrm{aug}}). \tag{4}$$

Conditioning on the observed targets:

$$y_* | y_1, \ldots, y_n \sim \mathcal{N}(\mu, \sigma^2), \qquad \begin{aligned} \mu &= Q_{*,1:n} Q^{-1} \mathbf{y} \\ \sigma^2 &= Q_{*,*} - Q_{*,1:n} Q^{-1} Q_{1:n,*}. \end{aligned} \tag{5}$$

Gaussian bumps on the training data

A common choice for the basis functions ϕ_m is "Gaussian bumps", centered on the training data points:



Noise-free 95% posterior confidence region. 5 data points, known noise magnitude $\sigma_n^2 = 0.01$ and basis functions $\phi_j(x) = \exp(-(x - x_j)^2)$.

The infinite limit

Key idea: Let's put (scaled down) bumps everywhere!

The sum over basis functions becomes and integral:

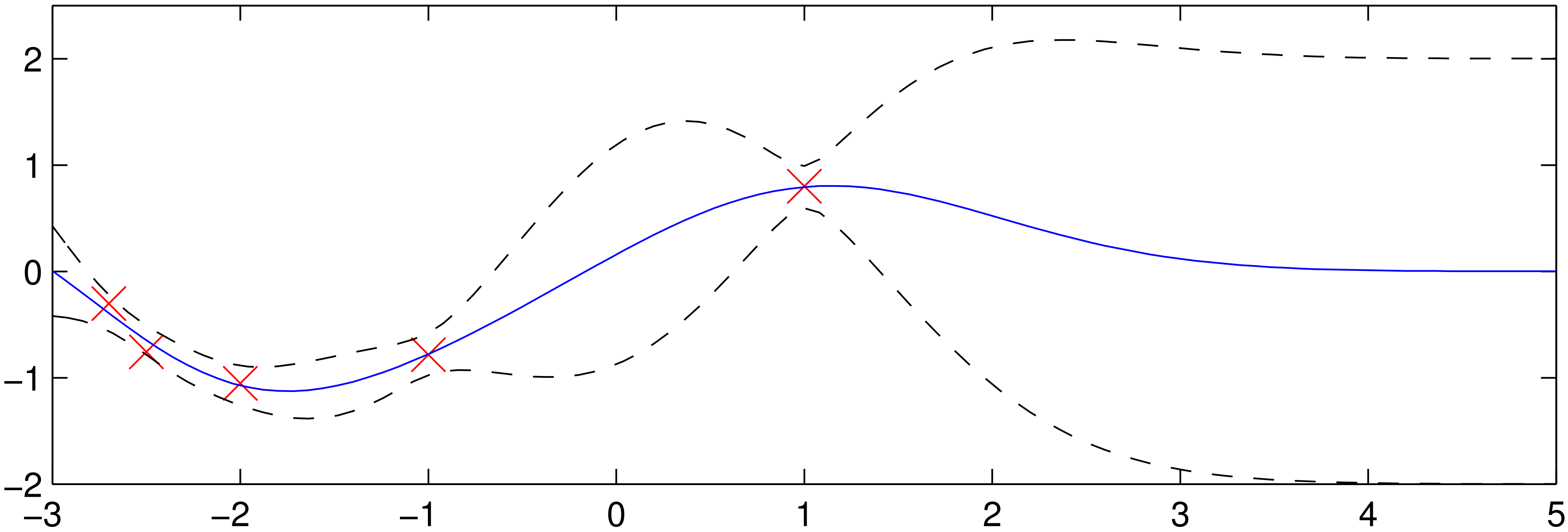
$$\begin{aligned} Q_{ij} &\propto \int_{z_{\min}}^{z_{\max}} \phi_z(x_i)\phi_z(x_j)dz + \delta_{ij}\sigma_n^2 \\ &= \int_{z_{\min}}^{z_{\max}} \exp(-(x_i - z)^2)\exp(-(x_j - z)^2))dz + \delta_{ij}\sigma_n^2. \end{aligned} \tag{6}$$

Extending the limits to infinity, we can solve the integral:

$$Q_{ij} \propto \exp(-(x_i - x_j)^2/2) + \delta_{ij}\sigma_n^2. \tag{7}$$

See eg. MacKay [2003] for this construction.

The same example in the infinite case



Computational note

The computation still "only" requires inversion of an $n \times n$ matrix.

Conclusions

- Even if your model manages to explain the training points, it may still be underfitting.
- Uncertainties may be underestimated in too simple (finite, parametric) models.
- Trivial to implement for Gaussian process regression models with Gaussian noise.
- Important distinction: degenerate vs non-degenerate Gaussian process models.