Bayesian Nonparametric Nonlinear System Identification

A quick overview

Roger Frigola

University of Cambridge, UK

Reglerteknik Monday Meeting, Linköpings Universitet, June 2013
Two Different Approaches to Modelling

Maximum Likelihood (or PEM) + Regularisation

- Optimisation to find point estimates of parameters given the data.

\[ \theta^* = \arg \min_{\theta} L(D, \theta) + J(\theta) \]

\[ x_{t+1} = f(x_t, \theta^*) \]

Bayesian

- Averaging over parameter distributions to find distributions over predictions

\[ p(x_{t+1} | x_t, D) = \int f(x_{t+1} | x_t, \theta) \ p(\theta | D) \ d\theta \]
Why Bayesian?

- No overfitting because there is no fitting!
- No need to artificially limit the complexity of the models.
- Even with a finite set of perfect observations we may still be uncertain about our model/parameters.
Why Nonparametric?

In a parametric model

\[ p(x_{t+1} \mid x_t, \theta, \mathcal{D}) = p(x_{t+1} \mid x_t, \theta) \]

In a nonparametric model, we perform inference on the space of functions, not parameters!

- Data is not condensed in a finite set of parameters.
- Flexibility not constrained by choice of parametric form.
- Nonparametric model can be very complex if the data supports this complexity.
Parallels between Bayesian Approach and Regularisation

- Prior $\sim$ Regulariser.
- Inductive bias. There is no inference without assumptions.
- One should not be afraid of priors: they are a very honest way to make assumptions that in other methods may be hidden inside algorithms.
Generative Models

A model that describes the data that can be observed from the system. It allows us to:

- Generate “fantasies” invented by the model.
- **Condition** on actual observations and infer the latent quantities.

\[
\begin{align*}
p(a) &= \text{Uniform}(-1, 1) \\
p(x_0) &= \mathcal{N}(0, 1) \\
p(x_{t+1}) &= \mathcal{N}(ax_t, 1) \\
p(y_t) &= \mathcal{N}(x_t, 1)
\end{align*}
\]
Bayesian Nonparametric NARX Models

Take a NARX model

\[ y_t = f(y_{t-1}, y_{t-2}, \ldots, u_{t-1}, \ldots) \]

and learn \( f \) in a nonparametric fashion, e.g. putting a Gaussian process prior over it.

Problem: this is an errors-in-variables regression since the inputs to \( f \) are noisy.
Treating the model in a probabilistically consistent way is hard. 
Pragmatic solution: pre-process signals to reduce noise and put it as a regression problem.

$\text{SNR [dB]}$ vs $\text{RMSE [V]}$

Silverbox benchmark

- GP–FNARX (SoD), 23 ± 2
- GP–FNARX (FITC), 23 ± 2
- GP–NARX (SoD), 14 ± 1
- GP–NARX (FITC), 14 ± 1
- wavenet nlax, 5 ± 1
- sigmoidnet nlax, 8 ± 9
- treepartition nlax, 7 ± 0
- wavenet nlax (filt), 6 ± 2
- sigmoidnet nlax (filt), 74 ± 11
- treepartition nlax (filt), 7 ± 0

Wiener–Hammerstein benchmark

- GP–FNARX (SoD), 25 ± 2
- GP–FNARX (FITC), 25 ± 2
- GP–NARX (SoD), 16 ± 1
- GP–NARX (FITC), 16 ± 1
- wavenet nlax, 7 ± 3
- sigmoidnet nlax, 85 ± 12
- treepartition nlax, 8 ± 0
- wavenet nlax (filt), 5 ± 1
- sigmoidnet nlax (filt), 85 ± 8
- treepartition nlax (filt), 8 ± 0

$1$ work with Carl E. Rasmussen.
Bayesian Nonparametric Nonlinear State-Space Models

Nonparametric model for the state transition. Problem: states are not observed so this is not straightforward regression.

Example:
(this is the true, but unknown to us, system description)

\[ x_{t+1} = ax_t + \frac{bx_t}{1 + x_t^2} + cu_t + v_t, \quad v_t \sim \mathcal{N}(0, q) \]
\[ y_t = dx_t^2 + e_t, \quad e_t \sim \mathcal{N}(0, r) \]
Sys. Id. with Bayesian Nonparametric State-Space Models using Gaussian Processes and Particle MCMC

2 work with Fredrik Lindsten, Thomas B. Schön and Carl E. Rasmussen.