Infinite Mixture of Global Gaussian Processes

Melanie F. Pradier[1], and Fernando Perez-Cruz[1,2]
1. University Carlos III in Madrid, Spain and 2. Bell Labs, Alcatel-Lucent, New Jersey, USA
{melanie,fernando}@tsc.uc3m.es

Introduction

- Gaussian Processes (GP) are useful to solve non-linear regression problems, but they are limited to unimodal Gaussian output distributions, stationary functions and i.i.d noise scenarios.

Objective

- Build a probabilistic model for general non-linear regression problems to deal with:
  - arbitrary output distribution (including multimodality)
  - non-stationary functions
  - heteroscedastic noise

Paper Contributions

- A general model for non-linear regression: Infinite Mixture of Global GPs (IMoGGP). Novel interpretation as a single-p Dependent Dirichlet Process.
- An easy-to-implement MCMC sampling algorithm.
- Comparative results against Infinite Mixture of Experts (IMoE).

Model

We want to estimate $y \in \mathbb{R}$ given an input $\boldsymbol{x} \in \mathbb{R}^D$ and a database $\mathcal{D}_n = \{\boldsymbol{x}_i, y_i\}_{i=1}^n$, that is

$$p(y|\boldsymbol{x}, \mathcal{D}_n). \tag{1}$$

Our model is based on the stick-breaking construction of a Dirichlet Process (DP):

$$\boldsymbol{\pi}|\alpha \sim \mathrm{GEM}(\alpha) \tag{2}$$
$$z_i|\boldsymbol{\pi} \sim \mathrm{Multinomial}(\boldsymbol{\pi}) \tag{3}$$
$$\theta_m|H \sim H \tag{4}$$
$$y_i|z_i, \{\theta_m\} \sim F(\theta_{z_i}), \tag{5}$$

where GEM stands for the stick-breaking prior by Griffiths, Engen and McCloskey, α is the concentration parameter of the DP, z_i indicates the cluster assignment, θ_m designates the cluster parameters, H is a base measure, and $F(\cdot)$ is the likelihood function, typically Gaussian.

In the regression setting, each y_i is associated with an input $\boldsymbol{x}_i$ and we can directly modify (5) as

$$y_i|z_i, \{\theta_m\}, \boldsymbol{x}_i \sim F\big(\theta_{z_i}(\boldsymbol{x}_i)\big), \tag{6}$$
$$\theta_m|H, \phi_m \sim H_{\phi_m}, \tag{7}$$

where we assume that $F\big(\theta_{z_i}(\boldsymbol{x}_i)\big)$ is Gaussian-distributed with mean $\mu_{z_i}(\boldsymbol{x}_i)$ and variance $\sigma^2_{z_i}(\boldsymbol{x}_i)$, and H_{ϕ_m} is a Gaussian process prior with hyperparameters ϕ_m.

- Now each cluster parameter θ_m corresponds to a latent function over the input space.
- We can interpret the model as a Single-p Dependent Dirichlet Process whose atoms are GP functions.

Conceptual Comparison of Methods

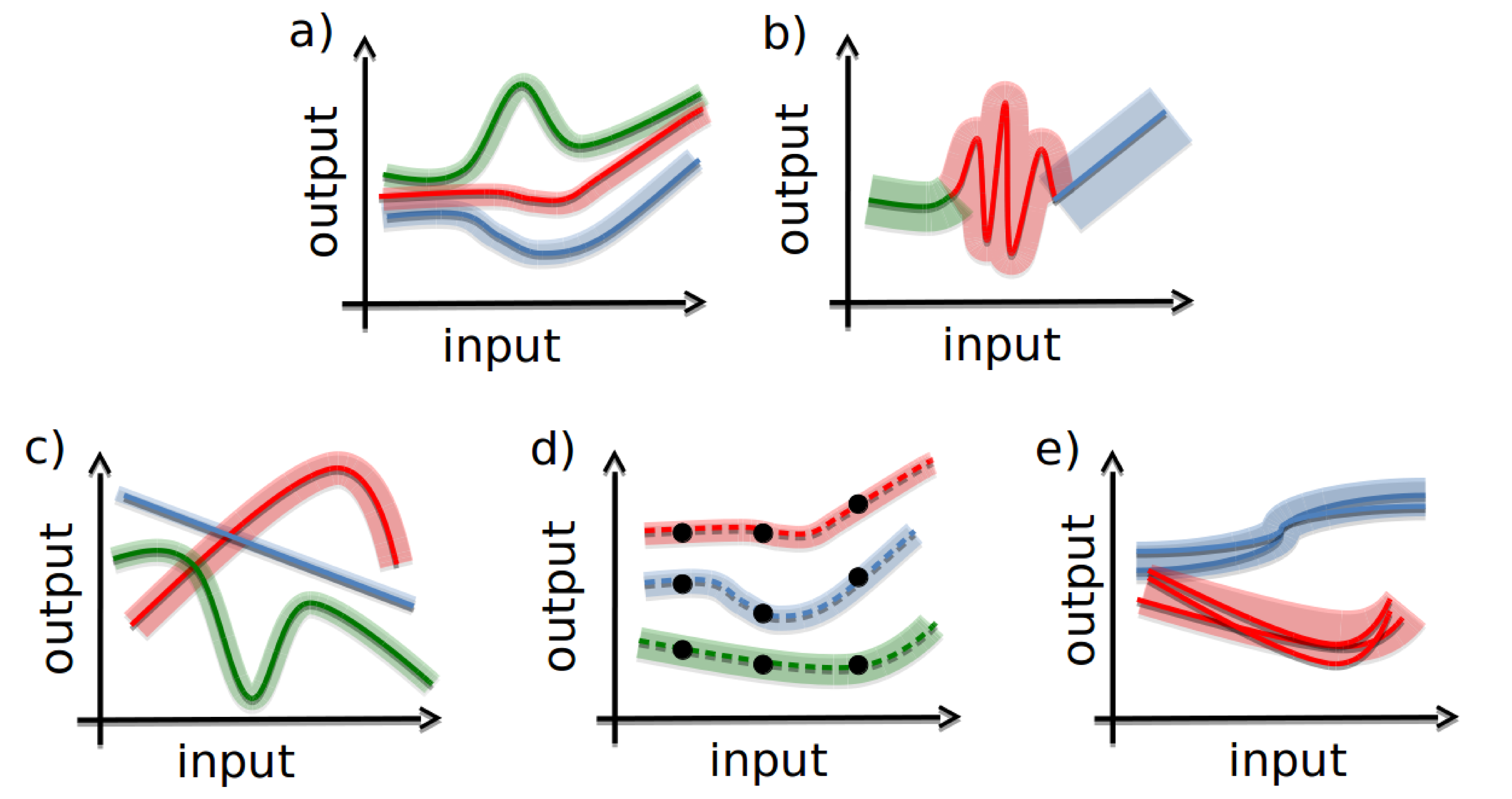
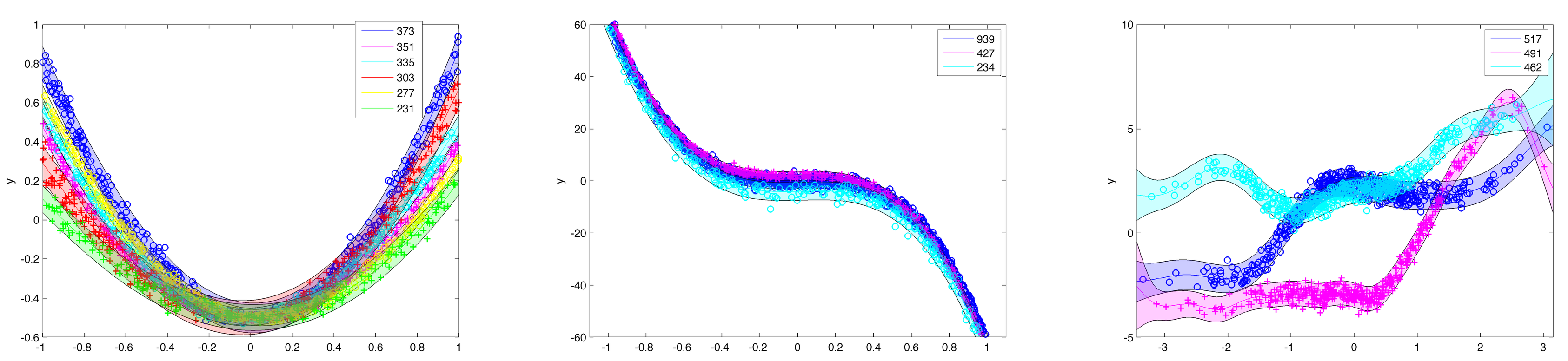


Figure 1: **Conceptual comparison of different approaches.** Sketch comparing a) Infinite Mixture of Global GPs (proposed approach), b) Infinite Mixture of Experts, c) Overlapping GPs for multi-tracking, d) Spatial Dirichlet Process, and e) time series clustering. Each color represents a different GP.

Properties



(a) Heteroscedasticity. (b) Non-Gaussianity. (c) Multimodality.

Figure 2: **Properties that can be captured by the IMoGGP model**: (a) non-stationary, heteroscedastic noise; (b) non-Gaussian likelihoods, specifically a Student's t with Gamma distributed noise; and, (c) multimodal predictive distributions.

Inference for the IMoGGP

Algorithm 1 For each Gibbs sampling iteration:

1: Sample extended vector of mixture proportions:

$$\boldsymbol{\pi}|\boldsymbol{z}, \alpha \sim \mathrm{Dirichlet}\big(n_1, \ldots n_K, \underbrace{\alpha/T \ldots \alpha/T}_{T \text{ times}}\big) \tag{8}$$

2: Sample latent functions, i.e., cluster parameters θ_m, $m = 1, \ldots, M^+$:

$$p(\theta_m|\boldsymbol{\pi}, \boldsymbol{y}, \mathbf{X}, \boldsymbol{z}) \propto p(\theta_m|H_{\phi_m})p(\boldsymbol{y}|\mathbf{X}, \boldsymbol{z}, \theta_m) \tag{9}$$

3: Sample cluster assignments:

$$p(z_i|\boldsymbol{\pi}, y_i, \boldsymbol{x}_i, \boldsymbol{z}_{-i}, \{\theta_m\}) \propto p(z_i|\boldsymbol{\pi})\, p(y_i|\boldsymbol{x}_i, \boldsymbol{z}, \{\theta_m\}) \tag{10}$$

4: Sample hyperparameters ϕ_m, $m = 1, \ldots, M^+$:

$$p(\phi_m|\boldsymbol{\pi}, \boldsymbol{y}, \mathbf{X}, \boldsymbol{z}) \propto p(\phi_m|H)p(\boldsymbol{y}|\mathbf{X}, \boldsymbol{z}, \phi_m) \tag{11}$$

5: Sample concentration parameter α for the mixture model

- This algorithm is simple and computationally efficient, as it divides data into mutiple GPs (we have smaller matrices to invert).

Results

		(a)	(b)	(c)	(d)	(e)	(f)
PLLH	sGP	-0.0217	-3.4920	-3.3030	-0.5855	-1.6373	0.2033
	IMoE	0.7017	-2.1248	-2.1604	1.9452	-1.6308	**0.9943**
	IMoGGP	**0.9008**	**-2.1237**	**-1.2575**	**2.3587**	**-1.5723**	0.9846
MSE	sGP	0.0288	**4.8115**	4.2815	93.4640	0.7877	86.6815
	IMoE	0.0331	4.8394	5.2263	93.4640	0.7780	82.8929
	IMoGGP	**0.0287**	4.8500	**4.2703**	**43.6710**	**0.7754**	**82.4264**

Table 1: **Comparison of the single GP (sGP), the Infinite Mixture of Experts (IMoE) and the proposed Infinite Mixture of Global Gaussian Processes (IMoGGP).** The three first columns correspond to synthetic toy examples showed above: (a) Heteroscedasticity, (b) Non-Gaussianity, (c) Multimodality. The last three columns correspond to real databases available online: (d) Concrete, (e) Marathon, (f) RSSI.

Future work

- Study sensibility to hyperparameters.
- Relax constant weights assumption (similar to Kernel-based Stick breaking process).
- Extend to higher dimensional problems (selection mechanism of input-dimension).

Funding

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 316861.