Multi-Mean Gaussian Processes: A novel probabilistic framework for multi-correlated longitudinal data

Arthur Leroy[1]

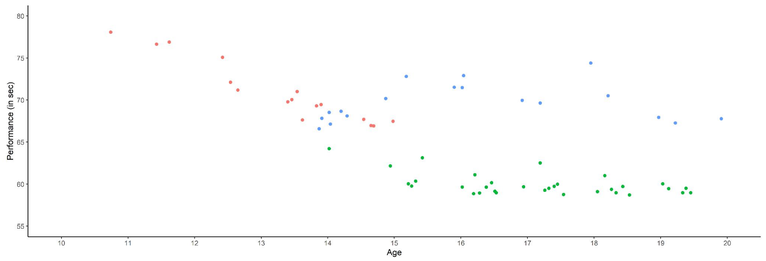
Mauricio A Alvarez Lopez[1]

[1]Department of Computer Science, The University of Manchester

NEURAL INFORMATION PROCESSING SYSTEMS

Context

Suppose you observe irregular longitudinal data coming from multiple sources, collected at different locations (e.g. on the figure below, the performance on 100m freestyle events for 3 different swimmers have been observed at different ages). Multi-task (or multi-output) Gaussian processes (GPs) approaches have been successfully applied despite scaling and flexibility limitations [1,2,3].



We recently proposed a novel paradigm for defining multi-task GPs by **transferring knowledge through a common latent mean process** [4]. This framework has first been extended to perform a **simultaneous clustering** of the tasks [5], and then to handle **multiple sources of correlations** at large scales (e.g. data from hundreds of individuals, each of them gathering thousands of gene-related measurements evolving over time).

Sharing information through a mean process

If longitudinal data are collected from M correlated tasks, let us assume that each observation $y_i(t), i = 1, \dots, M$ can be decomposed as the sum of a common mean process μ_0, a task-specific process f_i and a noise term:

$$y_i(t) = \mu_0(t) + f_i(t) + \varepsilon_i(t),$$

with

$$\mu_0(\cdot) \sim \mathcal{GP}(m_0(\cdot), K_0(\cdot,\cdot)), \quad f_i(\cdot) \sim \mathcal{GP}(0, \Sigma_i(\cdot,\cdot)), \quad \varepsilon_i(\cdot) \sim \mathcal{GP}(0, \sigma_i^2 I)$$

where we do not impose any assumption on the covariance structures. Conditionally to the mean process, all tasks are independent and the conditional likelihood of the model factorises and remains Gaussian. Thanks to an **EM algorithm**, we can still learn kernel hyper-parameters through gradient-descent optimisation (M step) and compute the hyper-posterior distribution of the latent mean process in closed form (E step):

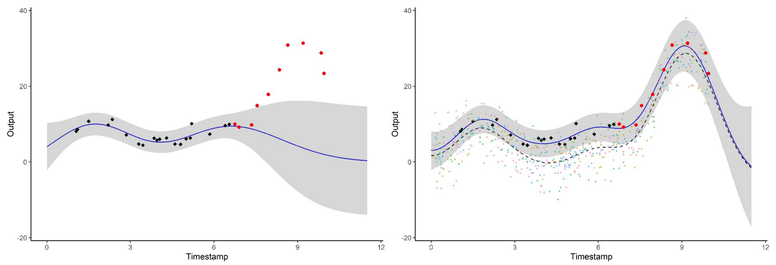
$$p(\boldsymbol{\mu}_0|\{\mathbf{y}_i\}_{i=1,\dots,M}) \propto p(\boldsymbol{\mu}_0)\prod_{i=1}^{M} p(\mathbf{y}_i|\boldsymbol{\mu}_0) = \mathcal{N}(\boldsymbol{\mu}_0;\ \hat{m}, \hat{K})$$

The overall learning procedure has a **linear complexity** in the number of tasks.

After learning, the key idea for prediction is to integrate out the mean process to define a **multi-task prior** distribution for any task $\mathbf{y}_*$ we want to predict.

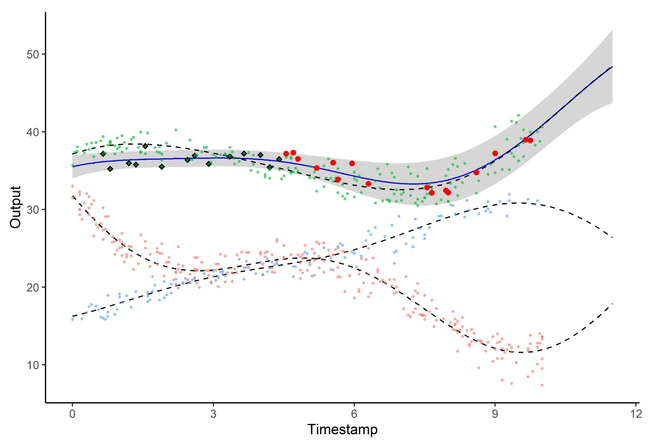
$$\begin{aligned} p(\mathbf{y}_* \mid \{\mathbf{y}_i\}_{i=1,\dots,M}) &= \int p(\mathbf{y}_*, \boldsymbol{\mu}_0 \mid \{\mathbf{y}_i\}_{i=1,\dots,M}) \mathrm{d}\boldsymbol{\mu}_0 \\ &= \int p(\mathbf{y}_* \mid \boldsymbol{\mu}_0) p(\boldsymbol{\mu}_0 \mid \{\mathbf{y}_i\}_{i=1,\dots,M}) \mathrm{d}\boldsymbol{\mu}_0 \\ &= \mathcal{N}(\mathbf{y}_*;\ \hat{m}, \hat{K} + \boldsymbol{\Sigma}_* + \sigma_*^2 I). \end{aligned}$$

In the figure below, we display a comparison between **standard GP** (left) and this approach, called **Magma** (right), to perform prediction on (red) testing points from (black) training points and information from other tasks (backward points).



Simultaneously allocate tasks into clusters

A valuable extension of the previous framework can be defined using a mixture of multi-task GPs where all tasks are allocated into clusters during learning. Each cluster is associated with a dedicated latent mean process leading to more accurate **cluster-specific predictions** for group-structured datasets, as presented in the following figure.



The R package ***MagmaClustR*** (https://arthurleroy.github.io/MagmaClustR) offers all tools to perform learning, prediction and visualisation using these frameworks.

Multi-Mean Gaussian Processes

Suppose that you now observe longitudinal data with two **different sources of correlation,** as in figures below, with methylation values for different genes of several individuals. We can write a generative model similar as before:

$$y_{i,j}(t) = \mu_0(t) + f_i(t) + g_j(t) + \varepsilon_{i,j}(t),$$

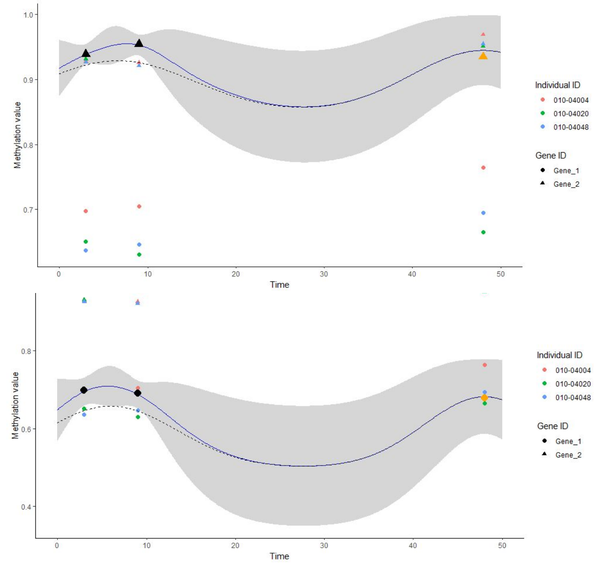
where i and j index the M individuals and N genes, respectively. All quantities remain GPs, and we can still derive an EM algorithm with closed-form E step:

$$p(\boldsymbol{\mu}_0|\{\mathbf{y}_{i,j}\}_{i=1,\dots,M}^{j=1,\dots,N}) = \mathcal{N}(\boldsymbol{\mu}_0;\ \hat{m}, \hat{K})$$

However, this quantity is rarely relevant in practice and we compute instead **multiple hyper-posterior** distributions using adequate subsets of training data.

$$p(\boldsymbol{\mu}_0|\{\mathbf{y}_{i,j}\}^{j=1,\dots,N}) = \mathcal{N}(\boldsymbol{\mu}_0;\ \hat{m}_i, \hat{K}_i),\ \forall i = 1, \dots, M$$
$$p(\boldsymbol{\mu}_0|\{\mathbf{y}_{i,j}\}_{i=1,\dots,M}) = \mathcal{N}(\boldsymbol{\mu}_0;\ \hat{m}_j, \hat{K}_j),\ \forall j = 1, \dots, N$$

This idea leads to adaptive and really efficient predictions, as displayed below:



References

[1] P. Goovaerts (1997) Geostatistics for Natural Resources Evaluation. In: Oxford University Press

[2] E. V. Bonilla et al. (2008) Multi-task Gaussian Process Prediction. In: Advances in Neural Information Processing Systems

[3] M. A. Álvarez et al. (2012) Kernels for Vector-Valued Functions: A Review. In: Foundations and Trends in Machine Learning

[4] A. Leroy et al. (2022) MAGMA: Inference and prediction using multi-task Gaussian processes with common mean. In: Machine Learning

[5] A. Leroy et al. (2022) Cluster-Specific Predictions with Multi-Task Gaussian Processes. In: Journal of Machine Learning and Research