Heterogeneous Multi-output Gaussian Process Prediction

Pablo Moreno-Muñoz\textsuperscript{1} Antonio Artés-Rodríguez\textsuperscript{1} Mauricio A. Álvarez\textsuperscript{2}
\textsuperscript{1}Universidad Carlos III de Madrid, Spain \textsuperscript{2}University of Sheffield, UK
(pmoreno, antonio)@tsc.uc3m.es mauricio.alvarez@sheffield.ac.uk

Introduction
A novel extension of multi-output Gaussian processes (MOGPs) for handling heterogeneous outputs (binary, real, categorical, \ldots). Each output has its own likelihood distribution and we use a MOGP prior to jointly model the parameters in all likelihoods as latent functions. We are able to obtain tractable variational bounds amenable to stochastic variational inference (SVI).

Multi-output GPs
We will use a linear model of corregionalisation type of covariance function for expressing correlations between latent parameter functions \( f_{\eta_j}(x) \) (LPFs).

Each LPF is a linear combination of independent latent functions \( \mathcal{U} = \{ u_\ell(x) \}_{\ell=1}^Q \). Each \( u_\ell(x) \) is assumed to be drawn from a GP prior such that \( u_\ell(x) \sim \mathcal{GP}(0, k_\ell(\cdot, \cdot)) \), where \( k_\ell \) can be any valid covariance function.

\[
f_{\eta_j}(x) = \sum_{\ell=1}^Q \sum_{q=1}^K \eta_{qj} \phi_{qj}(x),
\]

where we assume that \( R_q = 1 \), meaning that the corregionalisation matrices are rank-one. In the literature such model is known as the semiparametric latent factor model.

Heterogeneous Likelihood Model
Consider a set of output functions \( Y = \{ y_\ell(x) \}_{\ell=1}^Q \) with \( x \in \mathbb{R}^p \), that we want to jointly model using GPs. Let \( y(x) = [y_1(x), y_2(x), \ldots, y_Q(x)]' \) be a vector-valued function. If outputs are conditionally independent given the vector of parameters \( \theta(x) = [\theta_1(x), \theta_2(x), \ldots, \theta_Q(x)]' \), we may define

\[
p(y(x)|\theta(x)) = p(y_1(x)|\theta_1(x)) \cdots p(y_Q(x)|\theta_Q(x)) = p(y_1(x)|\theta_1(x)) p(y_2(x)|\theta_2(x)) \cdots p(y_Q(x)|\theta_Q(x)).
\]

where \( \theta_\ell(x) = \{ f_{\eta_1}(x), \ldots, f_{\eta_K}(x) \} \subset \mathbb{R}^{K \times 1} \) are the set of LPFs that specify the parameters in \( \theta_\ell(x) \) for an arbitrary number \( D \) of likelihood functions.

Sparse Approximations in MOGPs: We define the set of \( M \) inducing variables per latent function \( u_\ell(x) \) as \( u_\ell = [u_{\ell1}(z_1), \ldots, u_{\ellM}(z_M)]' \), evaluated at a set of inducing inputs \( Z = \{ z_m \}_{m=1}^M \subset \mathbb{R}^{M \times p} \). We also define \( u = [u_1', \ldots, u_M'] \subset \mathbb{R}^{M \times q} \). We approximate the posterior \( p(f, u | Y, X) \) as follows:

\[
p(f, u | Y, X) \approx q(f, u) = p(f|u)q(u) = \prod_{d=1}^D \prod_{j=1}^{J_d} \prod_{q=1}^Q p(f_{\eta_dj}(u)) q(u).
\]

Variational Inference: Exact posterior inference is intractable in our model due to the presence of an arbitrary number of non-Gaussian likelihoods. We use variational inference to compute a lower bound \( \mathcal{L} \) for the marginal log-likelihood \( \log p(Y) \), and for approximating the posterior distribution \( p(f, u | D) \).

\[
\mathcal{L} = \sum_{d=1}^D \mathbb{E}_{q(f_{\eta_dj})} \left[ \log p(y_{\eta_dj}(x)|\theta_\ell) \right] - \sum_{q=1}^Q KL(q(u)|p(u))
\]

Results

London House Price: Complete register of properties sold in the Greater London County during 2017. All properties addresses were translated to latitude-longitude points. For each spatial input, we considered two observations, one binary (property type) and one real (sale price).

Conclusions
We present a MOGP model for handling heterogeneous observations that is able to work on large scale datasets. Experimental results show relevant improvements with respect to independent learning.

References
Y. W. Teh et al. Semiparametric latent factor models. \textit{AISTATS}, 2005
J. D. Hadfield, MCMC methods for multi-response GLMMs. \textit{JSS}, 2010
J. Hensman et al., Gaussian processes for big data. \textit{IJCAI}, 2013
A. Saul et al., Chained Gaussian processes. \textit{AISTATS}, 2016

Acknowledgments: PPM is supported by a doctoral FPI grant (BES2016-077526) under the project Macro-ADBIE (TEC2015-67719-P), MINECO, Spain. AAR acknowledges the projects ADVENTURE (TEC2015-69666-C3-1-R), AID (TEC2014-52914-XI) and CASM-CAM (2012/ICE-265). MAA has been partially financed by the Engineering and Physical Research Council (EPSRC) Research Projects EP/N014162/1 and EP/R034303/1.

\textbf{Likelihood} \hspace{2cm} \textbf{Linked Parameters}

\begin{tabular}{|c|c|c|}
\hline
\textbf{Gaussian} & \( \mu(x) = f, \sigma(x) \) & \\
\textbf{Het. Gaussian} & \( \mu(x) = f_1, \sigma(x) = \exp(f_2) \) & \\
\textbf{Bernoulli} & \( \rho(x) = \frac{\exp(f_3)}{1 + \exp(f_3)} \) & \\
\textbf{Categorical} & \( \rho(x) = \frac{\exp(f_4)}{\sum_{c=1}^C \exp(f_c)} \) & \\
\textbf{Poisson} & \( \lambda(x) = \exp(f) \) & \\
\textbf{Gamma} & \( a(x) = \exp(f_1), b(x) = \exp(f_2) \) & \\
\hline
\end{tabular}