Heterogeneous Multi-output Gaussian Process Prediction

Pablo Moreno-Muñoz[1] Antonio Artés-Rodríguez[1] Mauricio A. Álvarez[2]

[1]Universidad Carlos III de Madrid, Spain [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, . . .). 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 $f_{d,j}(\mathbf{x})$ (LPFs).

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

$$f_{d,j}(\mathbf{x}) = \sum_{q=1}^{Q}\sum_{i=1}^{R_q} a^i_{d,j,q} u^i_q(\mathbf{x}),$$

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 $\mathcal{Y} = \{y_d(\mathbf{x})\}_{d=1}^{D}$, with $\mathbf{x} \in \mathbb{R}^p$, that we want to jointly model using GPs. Let $\mathbf{y}(\mathbf{x}) = [y_1(\mathbf{x}), y_2(\mathbf{x}), \cdots, y_D(\mathbf{x})]^\top$ be a vector-valued function. If outputs are conditionally independent given the vector of parameters $\boldsymbol{\theta}(\mathbf{x}) = [\theta_1(\mathbf{x}), \theta_2(\mathbf{x}), \cdots, \theta_D(\mathbf{x})]^\top$, we may define

$$p(\mathbf{y}(\mathbf{x})|\boldsymbol{\theta}(\mathbf{x})) = p(\mathbf{y}(\mathbf{x})|\mathbf{f}(\mathbf{x})) = \prod_{d=1}^{D} p(y_d(\mathbf{x})|\boldsymbol{\theta}_d(\mathbf{x})) = \prod_{d=1}^{D} p(y_d(\mathbf{x})|\widetilde{\mathbf{f}}_d(\mathbf{x})),$$

where $\widetilde{\mathbf{f}}_d(\mathbf{x}) = [f_{d,1}(\mathbf{x}), \cdots, f_{d,J_d}(\mathbf{x})]^\top \in \mathbb{R}^{J_d \times 1}$ are the set of LPFs that specify the parameters in $\boldsymbol{\theta}_d(\mathbf{x})$ for an arbitrary number D of likelihood functions.

Variational Bounds

Sparse Approximations in MOGPs: We define the set of M inducing variables per latent function $u_q(\mathbf{x})$ as $\mathbf{u}_q = [u_q(\mathbf{z}_1), \cdots, u_q(\mathbf{z}_M)]^\top$, evaluated at a set of *inducing inputs* $\mathbf{Z} = \{\mathbf{z}_m\}_{m=1}^{M} \in \mathbb{R}^{M\times p}$. We also define $\mathbf{u} = [\mathbf{u}_1^\top, \cdots, \mathbf{u}_Q^\top]^\top \in \mathbb{R}^{QM\times 1}$. We approximate the posterior $p(\mathbf{f}, \mathbf{u}|\mathbf{y}, \mathbf{X})$ as follows:

$$p(\mathbf{f}, \mathbf{u}|\mathbf{y}, \mathbf{X}) \approx q(\mathbf{f}, \mathbf{u}) = p(\mathbf{f}|\mathbf{u})q(\mathbf{u}) = \prod_{d=1}^{D}\prod_{j=1}^{J_d} p(\mathbf{f}_{d,j}|\mathbf{u}) \prod_{q=1}^{Q} q(\mathbf{u}_q),$$

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 $\log p(\mathbf{y})$, and for approximating the posterior distribution $p(\mathbf{f}, \mathbf{u}|\mathcal{D})$.

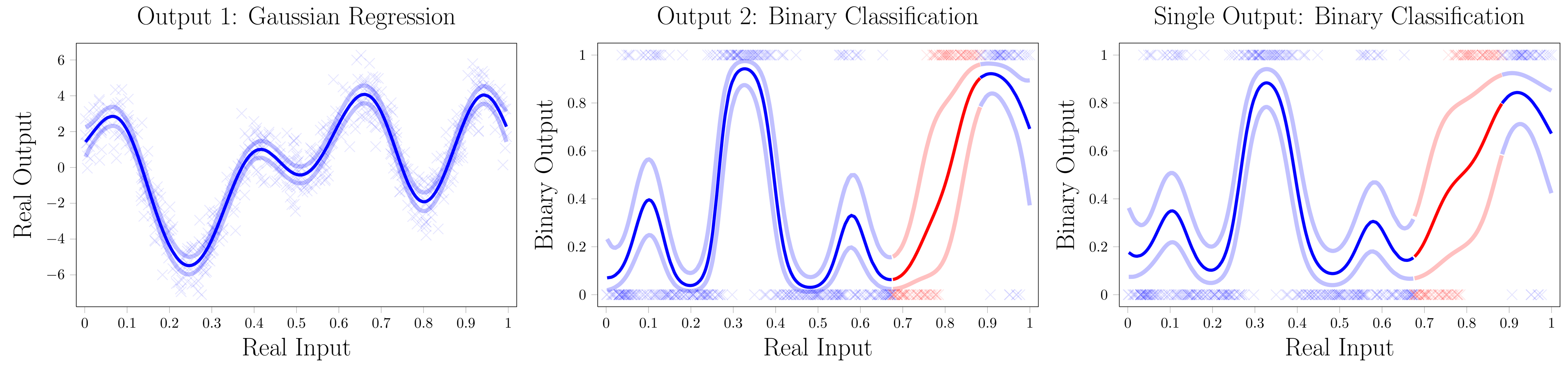
$$\mathcal{L} = \sum_{d=1}^{D} \mathbb{E}_{q(\widetilde{\mathbf{f}}_d)}\left[\log p(y_d(\mathbf{x}_n)|\widetilde{\mathbf{f}}_d)\right] - \sum_{q=1}^{Q} \mathrm{KL}\big(q(\mathbf{u}_q)||p(\mathbf{u}_q)\big)$$

Acknowledgements: PMM is supported by a doctoral FPI grant (BES2016-077626) under the project Macro-ADOBE (TEC2015-67719-P), MINECO, Spain. AAR acknowledges the projects ADVENTURE (TEC2015-69868-C2-1-R), AID (TEC2014-62194-EXP) and CASI-CAM-CM (S2013/ICE-2845). MAA has been partially financed by the Engineering and Physical Research Council (EPSRC) Research Projects EP/N014162/1 and EP/R034303/1.

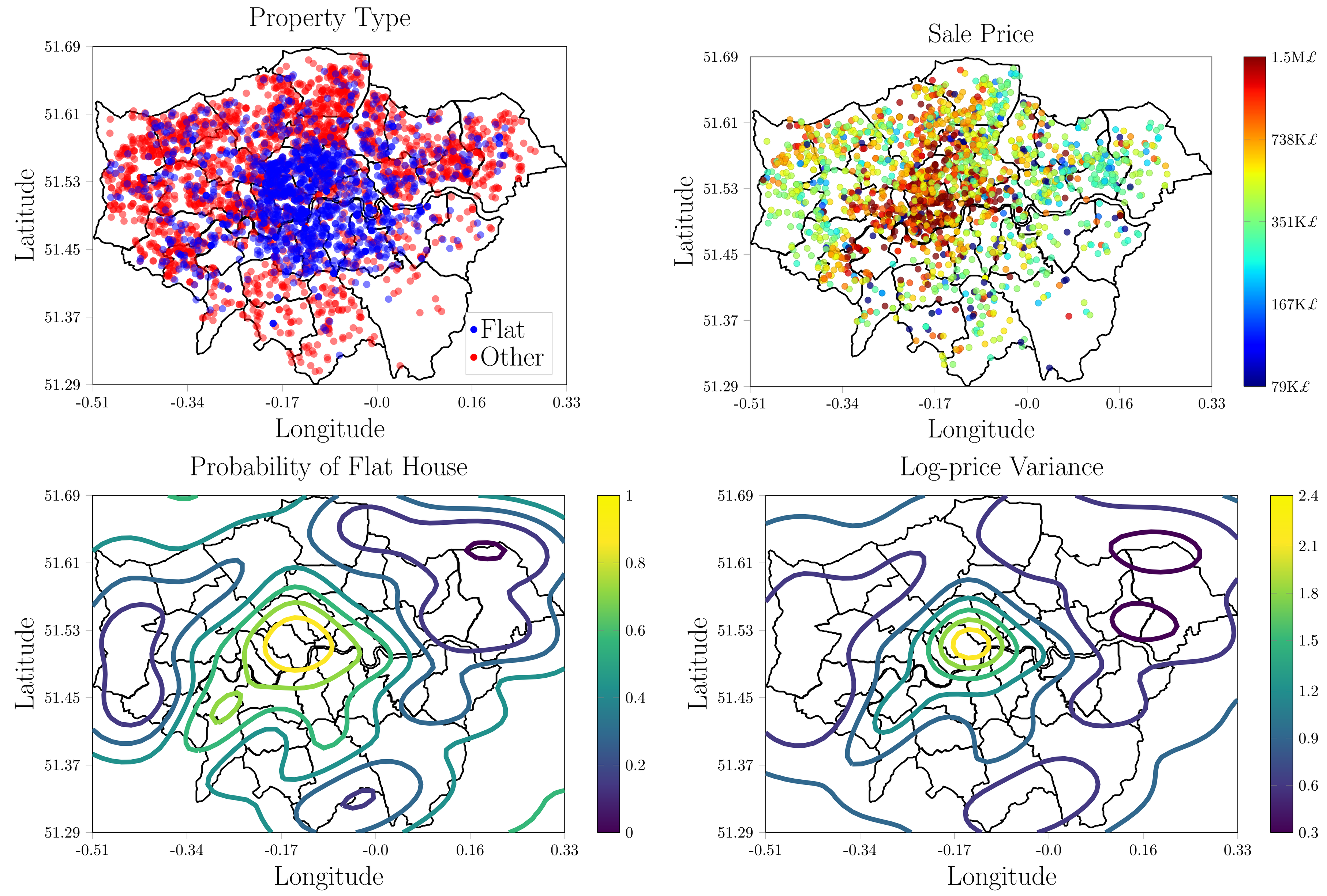
Results

Code → **github.com/pmorenoz/HetMOGP**

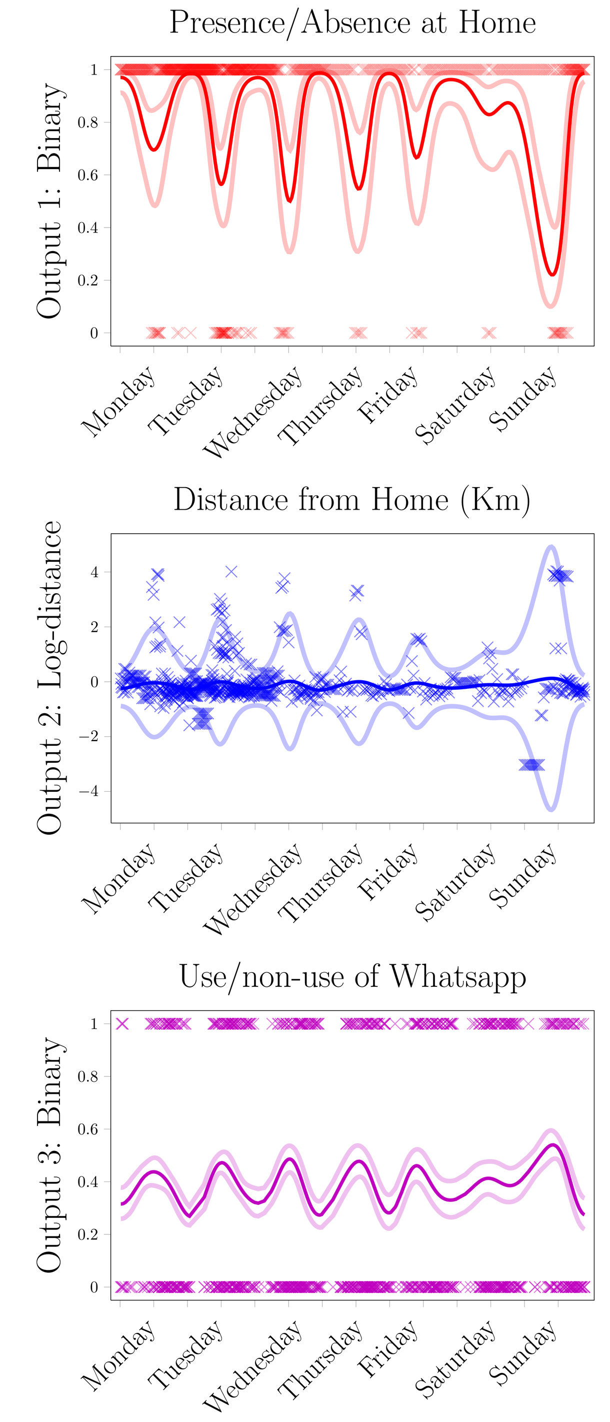
Missing Gap Prediction: We predict observations in one output (binary classification) using training information from another one (Gaussian regression). Multi-output test-NLPD value: $32.5 \pm 0.2 \times 10^{-2}$ / Single-output test-NLPD value: $40.51 \pm 0.08 \times 10^{-2}$.



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).



Human Behavior Data: We model human behavior in psychiatric patients. Our data comes from a medical study that uses the monitoring *app* (eB2).



TEST-NLPD

[London]	Bernoulli	Heteroscedastic	Global
HetMOGP	$\mathbf{6.38 \pm 0.46}$	$\mathbf{10.05 \pm 0.64}$	$\mathbf{16.44 \pm 0.01}$
ChainedGP	6.75 ± 0.25	10.56 ± 1.03	17.31 ± 1.06

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.

Likelihood	Linked Parameters
Gaussian	$\mu(\mathbf{x}) = \mathbf{f}, \sigma(\mathbf{x})$
Het. Gaussian	$\mu(\mathbf{x}) = \mathbf{f}_1, \sigma(\mathbf{x}) = \exp(\mathbf{f}_2)$
Bernoulli	$\rho(\mathbf{x}) = \frac{\exp(\mathbf{f})}{1+\exp(\mathbf{f})}$
Categorical	$\rho_k(\mathbf{x}) = \frac{\exp(\mathbf{f}_k)}{1+\sum_{k'=1}^{K-1}\exp(\mathbf{f}_{k'})}$
Poisson	$\lambda(\mathbf{x}) = \exp(\mathbf{f})$
Gamma	$a(\mathbf{x}) = \exp(\mathbf{f}_1), b(\mathbf{x}) = \exp(\mathbf{f}_2)$

References

Y. W. Teh et al. Semiparametric latent factor models. *AISTATS*, 2005
M. A. Álvarez et al., Sparse convolved Gaussian processes for multi-output regression. *NIPS*, 2008
J. D. Hadfield, MCMC methods for multi-response GLMMs. *JSS*, 2010
J. Hensman et al., Gaussian processes for big data. *UAI*, 2013
A. Saul et al., Chained Gaussian processes. *AISTATS*, 2016