CMM Center for Mathematical Modeling

Box-Cox Gaussian Processes

Gonzalo Rios and Felipe Tobar

Centro de Modelamiento Matemático, Universidad de Chile

GAMES

Abstract

Gaussian processes (GP) are Bayesian nonparametric regression models that have gained popularity within Machine Learning due perform training and inference in closed form, but assumes that all observations are jointly Gaussian. We proposed Box-Cox Gaussian processes, a generalization to model non-Gaussian observations in closed form. We validate our model through simulations using a real-world sunspot time series.

Gaussian Processes

A stochastic process $\{x_t\}_{t\in\mathcal{T}}$ is a Gaussian process [1] with mean $m(\cdot)$ and covariance $k(\cdot,\cdot)$, detonated

$$x(t) \sim \mathcal{GP}\left(m(t), k\left(t,\bar{t}\right)\right)$$

if any finite collection values of the process in $\mathbf{t} \in \mathbb{R}^N$ are distributed as a multivariate Normal with mean $\vec{\mu} = m(\mathbf{t})$ and covariance $\Sigma = k(\mathbf{t},\mathbf{t})$

$$\mathcal{N}\left(\vec{\mu},\Sigma\right) = \frac{1}{(2\pi)^{\frac{n}{2}}\left|\Sigma\right|^{\frac{1}{2}}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^{\top}\Sigma^{-1}(\vec{x}-\vec{\mu})}$$

The Gaussian process is then a prior distribution over functions, samples of this prior with a square-exponential kernel are shown in Fig. 1.

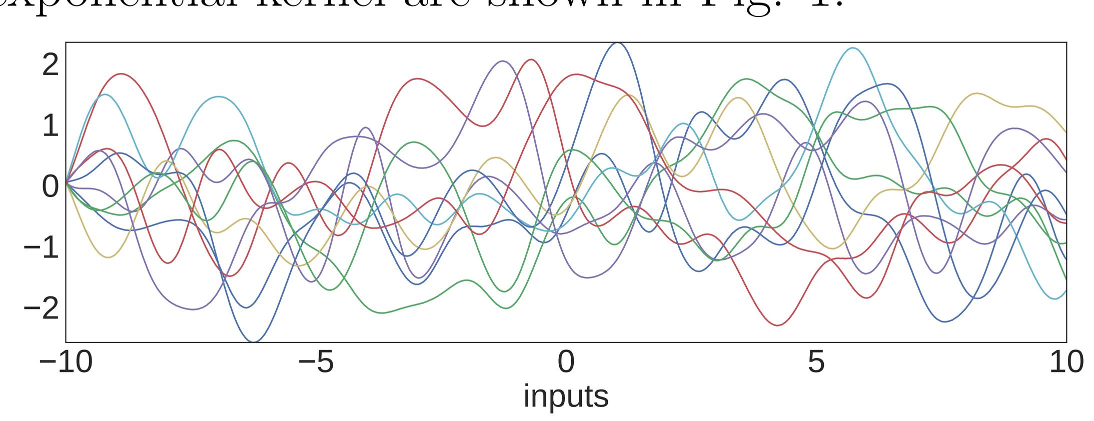

Figure 1: Draws from a Gaussian process prior

The Change of Variables Theorem

Let $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^n$ a random vector with a probability density function given by $p_x(\mathbf{x})$, and let $\mathbf{y} \in \mathcal{Y} \subseteq \mathbb{R}^n$ be a random vector such that $\varphi(\mathbf{y}) = \mathbf{x}$ for a function $\varphi : \mathcal{Y} \to \mathcal{X}$ bijective of class $\mathcal{C}^1$, such that $|\nabla\varphi(\mathbf{y})| > 0$ for all $\mathbf{y} \in \mathcal{Y}$. Then, the probability density function induced in $\mathcal{Y}$ is [2]

$$p_y(\mathbf{y}) = p_x(\varphi(\mathbf{y}))\left|\nabla\varphi(\mathbf{y})\right|.$$

Transformed Gaussian Processes

If the map φ is applied coordinate-wise, the Jacobian of φ is diagonal. If x is a GP, the posterior distribution of y is

$$p\left(y|\bar{y}\right) = \left|\nabla\varphi\left(y\right)\right|\mathcal{N}\left(\varphi\left(y\right)|m_{x|\bar{x}},\Sigma_{x|\bar{x}}\right)$$
$$m_{x|\bar{x}} = m_x + \Sigma_{x\bar{x}}\Sigma_{\bar{x}\bar{x}}^{-1}\left(\varphi\left(\bar{y}\right) - m_{\bar{x}}\right)$$
$$\Sigma_{x|\bar{x}} = \Sigma_{xx} - \Sigma_{x\bar{x}}\Sigma_{\bar{x}\bar{x}}^{-1}\Sigma_{\bar{x}x}$$

where $\Sigma_{\bar{x}x}$ is the joint covariance of $\bar{x}$ and x, and $\Sigma_{x|\bar{x}}$ is the posterior covariance of x given $\bar{x}x$.

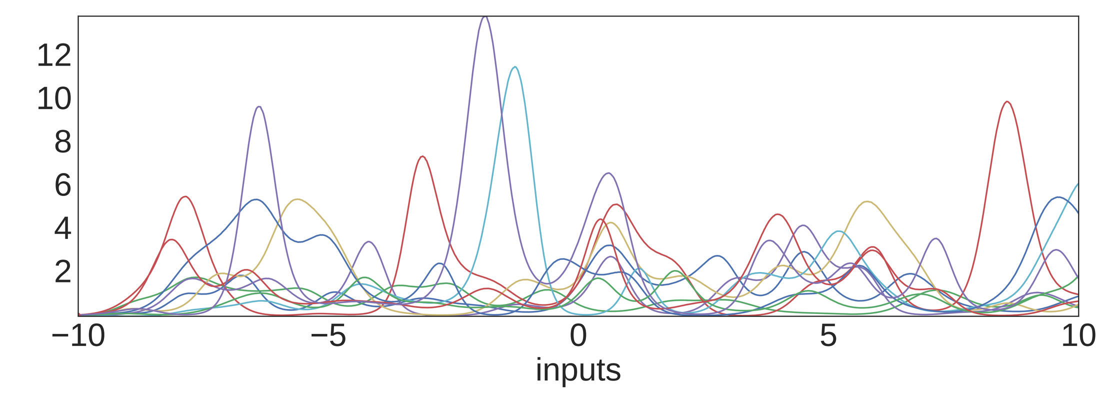

Figure 2: Draws from a log-Gaussian process

An example is to apply the logarithmic transformation $\varphi(y) = \log(y)$, that induces a stochastic process of different nature than x, since we emphasise that y is positive and the mean and covariance of y are no longer a transformation of the statistics m and k. Samples from such process are shown in Fig. 2.

Box-Cox Gaussian Processes

The Box-Cox transformations [3] is a family of power functions that generalizes the logarithm, these depend on a parameter $\lambda \in \mathbb{R}_0^+$ and are given by

$$\varphi_\lambda(y) = \frac{sgn(y)\left|y\right|^\lambda - 1}{\lambda}$$

where $\lim_{\lambda\to\infty}\varphi_\lambda(y) = \log(y)$. This functions transform Gaussian distributions on non-Gaussian and non-symmetric distributions as show in Fig. 3.

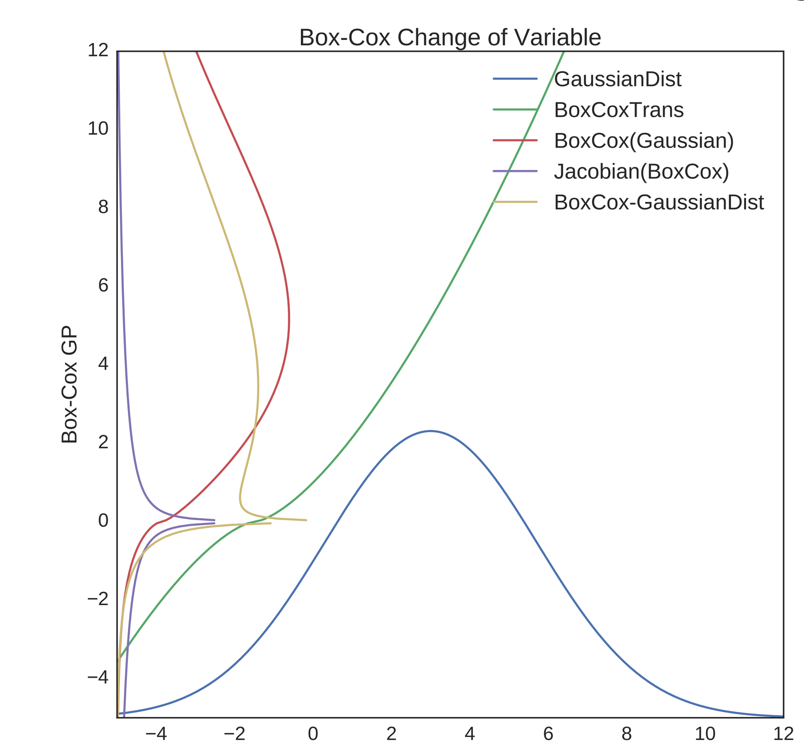

Figure 3: Change of Variable with Box-Cox Transformation

Experiment: Sunspot time series

We validate the proposed Box-Cox transformation against the standard GP formulation in the reconstruction of a sunspot time series (available from the `statsmodels` toolbox in Python) using a 25% of the data for training. This time series is always positive and skewed and thus not Gaussian. In Fig. 4 we show the posterior distribution with a standard GP, and in Fig. 5 we show the posterior distribution with a Box-Cox GP. In both cases we use an squared exponential kernel with a constant mean, and we add an shift parameter for the transformation on y.

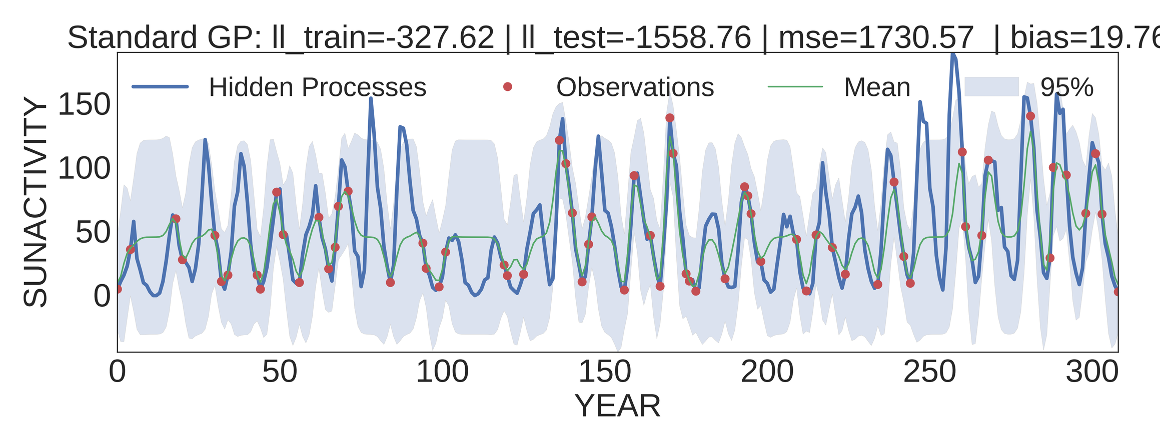

Figure 4: Inference on the time series using a standard GP

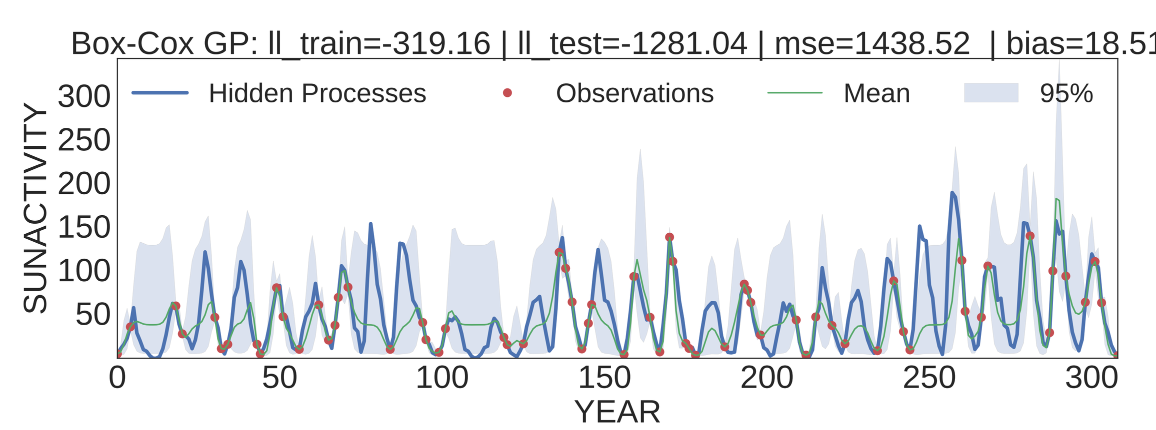

Figure 5: Inference on the time series using a Box-Cox GP

Discussion

The standard GP implementation exhibited unnecessarily large posterior variance and a reconstruction that is symmetric w.r.t. its mean, conversely, the proposed Box-Cox GP captured the skewness of the process and reported a smaller variance. The models can be compared in terms of the log-likelihood in train and test datasets, where the Box-Cox GP had better results on both datasets. Additionally, the models can be compared in terms of the mean square error (MSE) and the mean absolute bias (MAB), where the standard GP had 1017 and 19.76, while the Box-Cox GP had 1438 and 18.51, respectively. It is straightforward to check the goodness of the proposed model.

Conclusion

We have presented a principled procedure of transforming a Gaussian process to model non-Gaussian processes, like previous works [5] y [6] but we proposed a class of transformations based on compositions of functions that allow for expressive models at low computational cost relying on minimal numerical approximations. Through a case study that consisted in learning a sunspot time series, we validated our proposed model against the standard GP approach were the non-Gaussianity of the model. Further research in transformed GPs will be devoted to constructing even more expressive transformations, multiple-coordinate transformation such as the Gaussian process mixture of measurements [4], relating the transformation with the resulting moments of the observed process and finally devise a way of incorporating additional prior knowledge about the domain of the processes - such as nonnegativity or boundedness.

Acknowledgements

This work was partially funded by the projects CONICYT-PAI 82140061 and CONICYT-PCHA/DocNac/2016-21161789.

References

[1] C. E. Rasmussen and C. K. I. Williams.
Gaussian Processes for Machine Learning.
MIT, 2006.

[2] Terence Tao.
An introduction to measure theory, volume 126.
American Mathematical Soc., 2011.

[3] Peter J. Bickel and Kjell A. Doksum.
An analysis of transformations revisited.
Journal of the american statistical association, 76:296–311, 1981.

[4] F. Tobar, G. Rios, P. Guerrero, and T. Valdivia.
Recovering latent signals from a mixture of measurements using a gaussian process prior (in press).
IEEE Signal Processing Letters, 2016.

[5] E. Snelson, C. E. Rasmussen, and Z. Ghahramani.
Warped Gaussian processes.
NIPS, 16:337–344, 2004.

[6] M. Lázaro-Gredilla.
Bayesian warped Gaussian processes.
NIPS, 25:1619–1627, 2012.