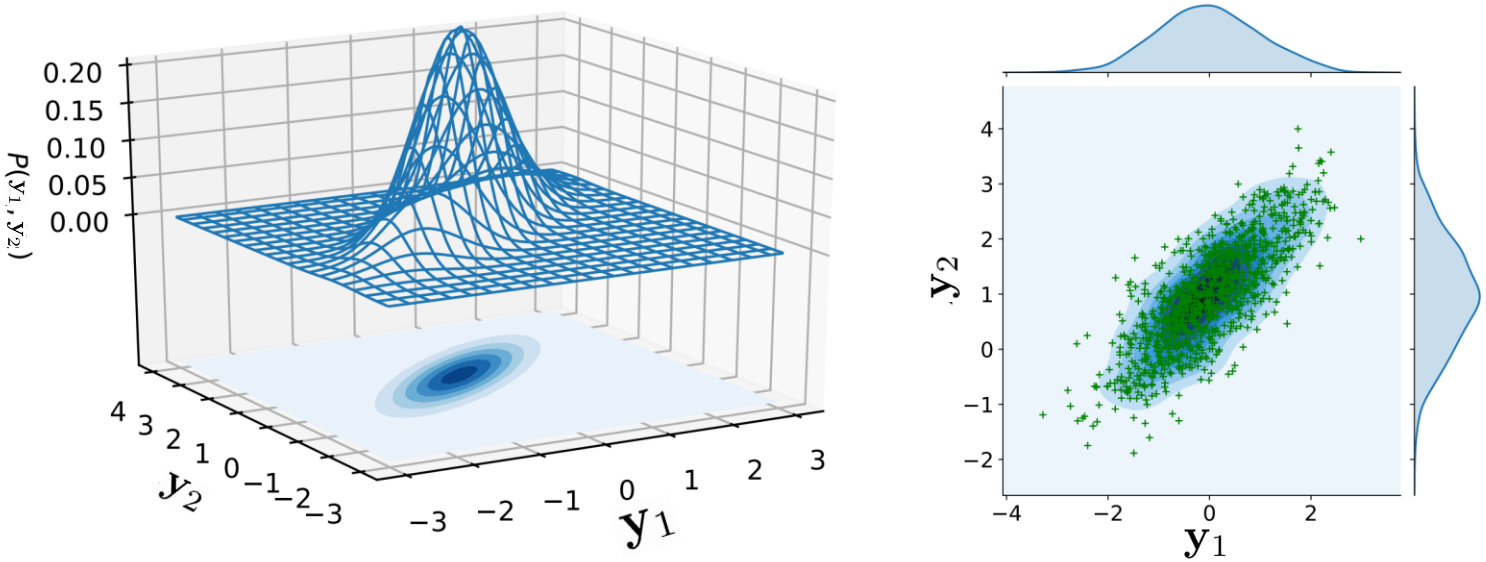
Gaussian Processes: An Introduction for Beginners

Juan José Giraldo Gutierrez

National Heart and Lung Institute
Imperial College London

Multivariate Gaussian Distribution

$$\mathcal{N}(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right)$$



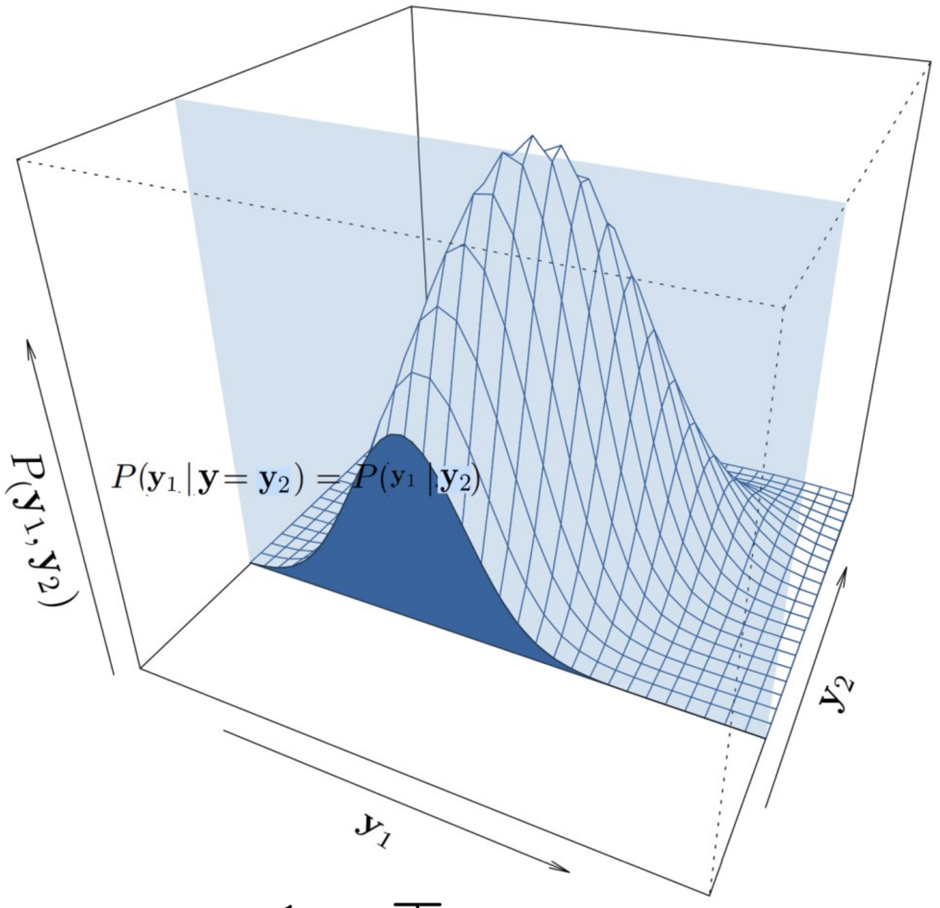
(a) 3-d bell curve

(b) 2-d ellipse contours

Figures taken from [2].

Conditioning Properties of a Gaussian Distribution

$$\mathcal{N}\left(\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \middle| \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}\right)$$



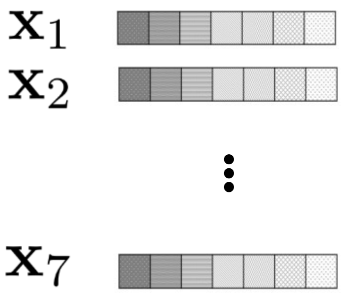
$$\mathcal{N}(\mathbf{y}_2 | \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{21}^{\top})$$

Figures taken from [2].

Gaussian Process

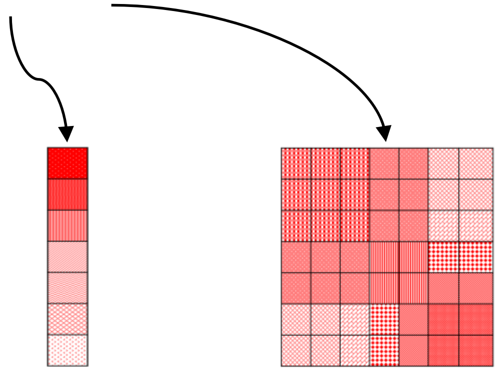
$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$$

If we evaluate **a finite number** input observations:



We would end up with a **Multivariate Gaussian Distribution**:

$$\mathbf{f} \sim \mathcal{N}(\mathbf{m}, \mathbf{K})$$



Details can be found in [1] and [2].

Bayesian Theorem for a Probabilistic Model

$$\underbrace{p(\mathbf{f}|\mathbf{y}, X)}_{\text{Posterior}} \propto \underbrace{p(\mathbf{y}|\mathbf{f}, X)}_{\text{Likelihood}} \underbrace{p(\mathbf{f}|X)}_{\text{Prior}}$$

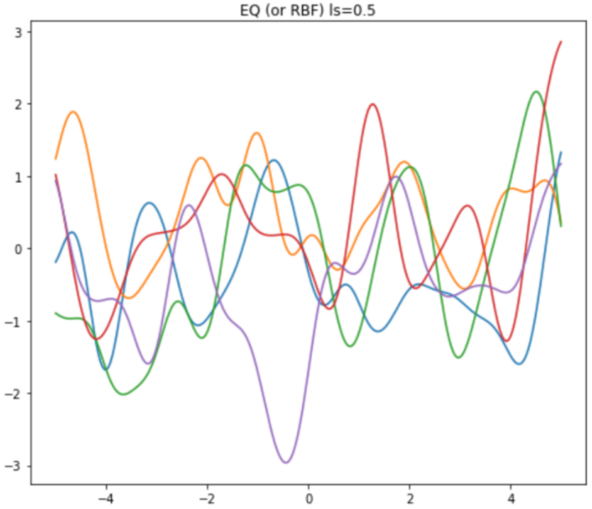
$$p(\mathbf{y}|\mathbf{f}, X) = \mathcal{N}(\mathbf{f}, \sigma_n^2 I) \qquad p(\mathbf{f}|X) = \mathcal{N}(\mathbf{0}, K)$$

The Covariance (or Kernel) Matrix

$$p(\mathbf{f}|X) = \mathcal{N}(\mathbf{0}, K) \quad \text{where} \quad K := K(X, X)$$

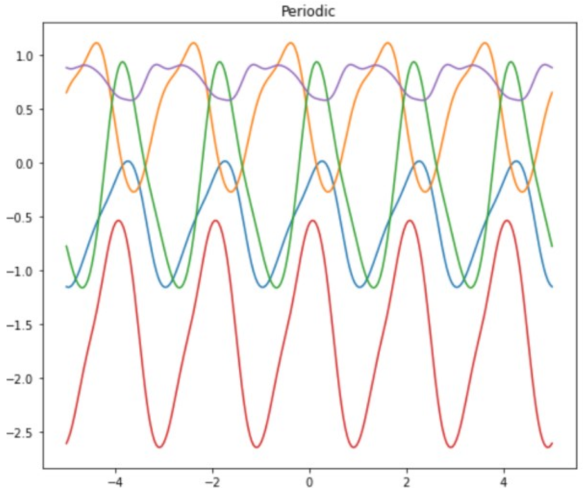
Exponentiated Quadratic

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\ell^2}\right)$$



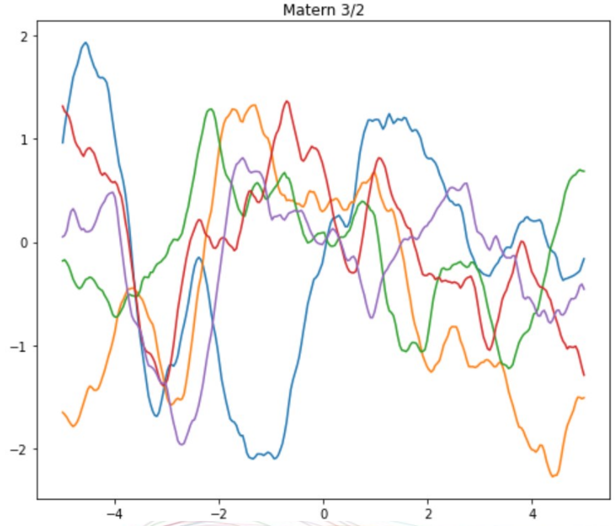
Periodic

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{2\sin^2\left(\frac{\|\mathbf{x}-\mathbf{x}'\|}{2}\right)}{l^2}\right)$$



Matérn

$$k(\mathbf{x}, \mathbf{x}') = \left(1 + \frac{\sqrt{3}\|\mathbf{x}-\mathbf{x}'\|}{\ell}\right)\exp\left(-\frac{\sqrt{3}\|\mathbf{x}-\mathbf{x}'\|}{\ell}\right)$$



Kernel cookbook in [3].

Fitting a Gaussian Process Model (with Gaussian Likelihood)

Marginal Likelihood:

$$p(\mathbf{y}|X) \;=\; \int p(\mathbf{y}|\mathbf{f}, X)p(\mathbf{f}|X)\, d\mathbf{f}$$

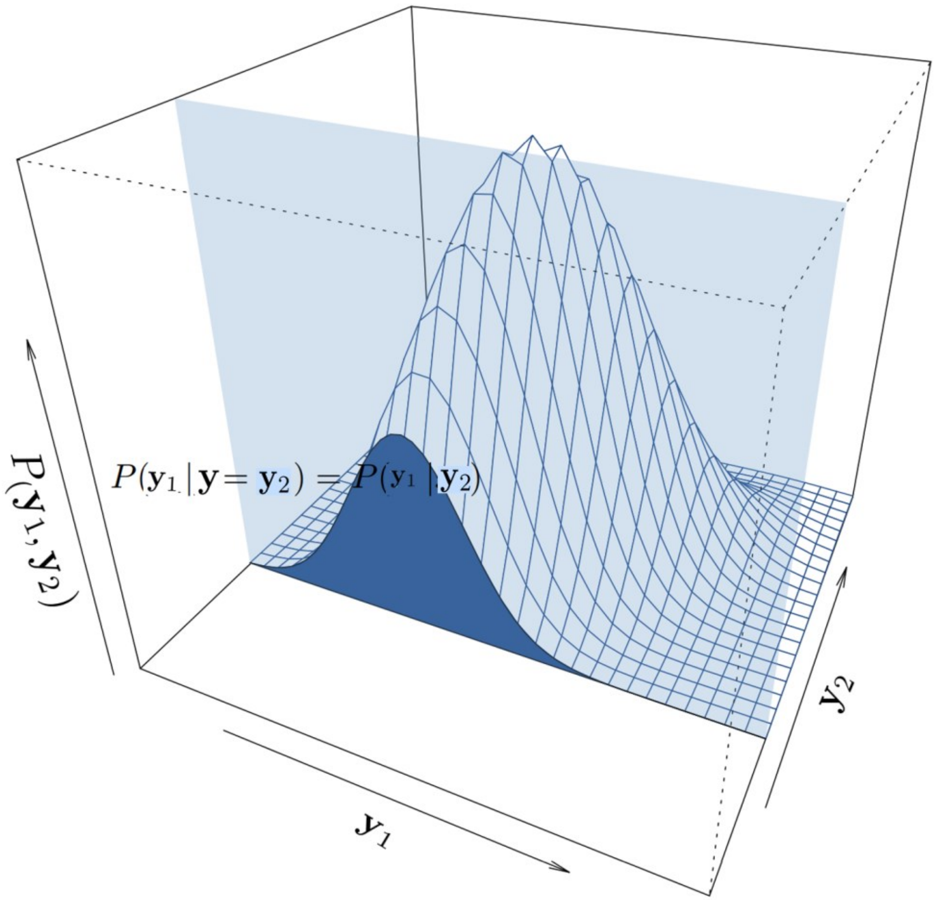
Maximize the Log Marginal Likelihood:

$$\log p(\mathbf{y}|X) \;=\; -\tfrac{1}{2}\mathbf{y}^\top(K+\sigma_n^2 I)^{-1}\mathbf{y} - \tfrac{1}{2}\log|K+\sigma_n^2 I| - \tfrac{n}{2}\log 2\pi$$

Detailed derivations of equations in [1].

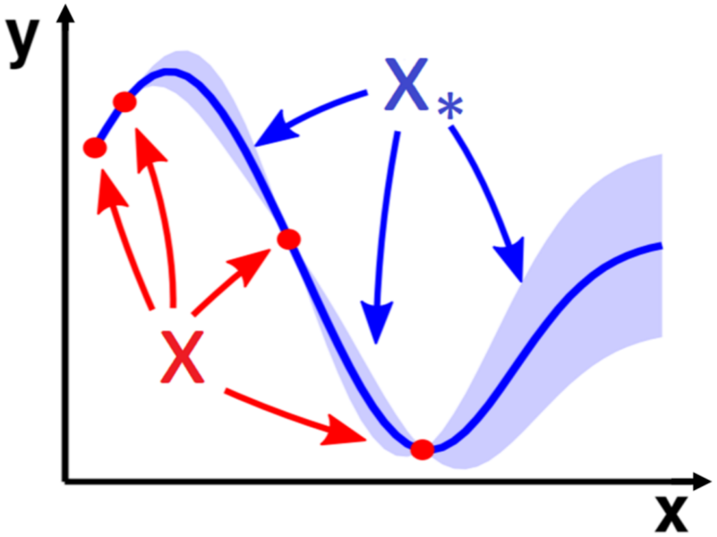
Recalling the Conditioning property (with zero mean)

$$\mathcal{N}\left(\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \middle| \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{bmatrix}\right)$$



$$\mathcal{N}(\mathbf{y}_2|\mathbf{K}_{21}\mathbf{K}_{11}^{-1}\mathbf{y}_1, \mathbf{K}_{22} - \mathbf{K}_{21}\mathbf{K}_{11}^{-1}\mathbf{K}_{21}^{\top})$$

Making Predictions



$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X)+\sigma_n^2 I & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right)$$

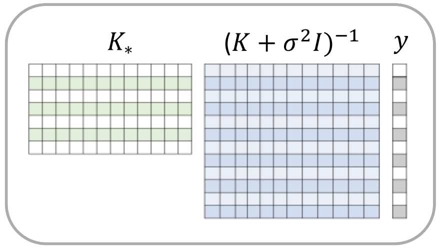
Image taken from [2]; detailed derivations of equations in [1].

Making Predictions

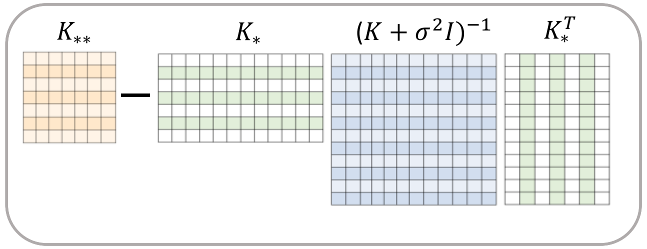
$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X)+\sigma_n^2 I & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right)$$

Apply **Conditional property** of a Multivariate Gaussian: $\mathcal{N}(\mathbf{y}_2|\mathbf{K}_{21}\mathbf{K}_{11}^{-1}\mathbf{y}_1, \mathbf{K}_{22}-\mathbf{K}_{21}\mathbf{K}_{11}^{-1}\mathbf{K}_{21}^{\top})$

$$\boldsymbol{\mu}_* = K(X_*,X)\left(K(X,X)+\sigma_n^2 I\right)^{-1}\mathbf{y}$$



$$\boldsymbol{\Sigma}_* = K(X_*,X_*) - K(X_*,X)\left(K(X,X)+\sigma_n^2 I\right)^{-1}K(X,X_*)$$



Important: The predictive **(co)variance encodes the uncertainty** of the prediction!!!!

Detailed derivations of equations in [1].

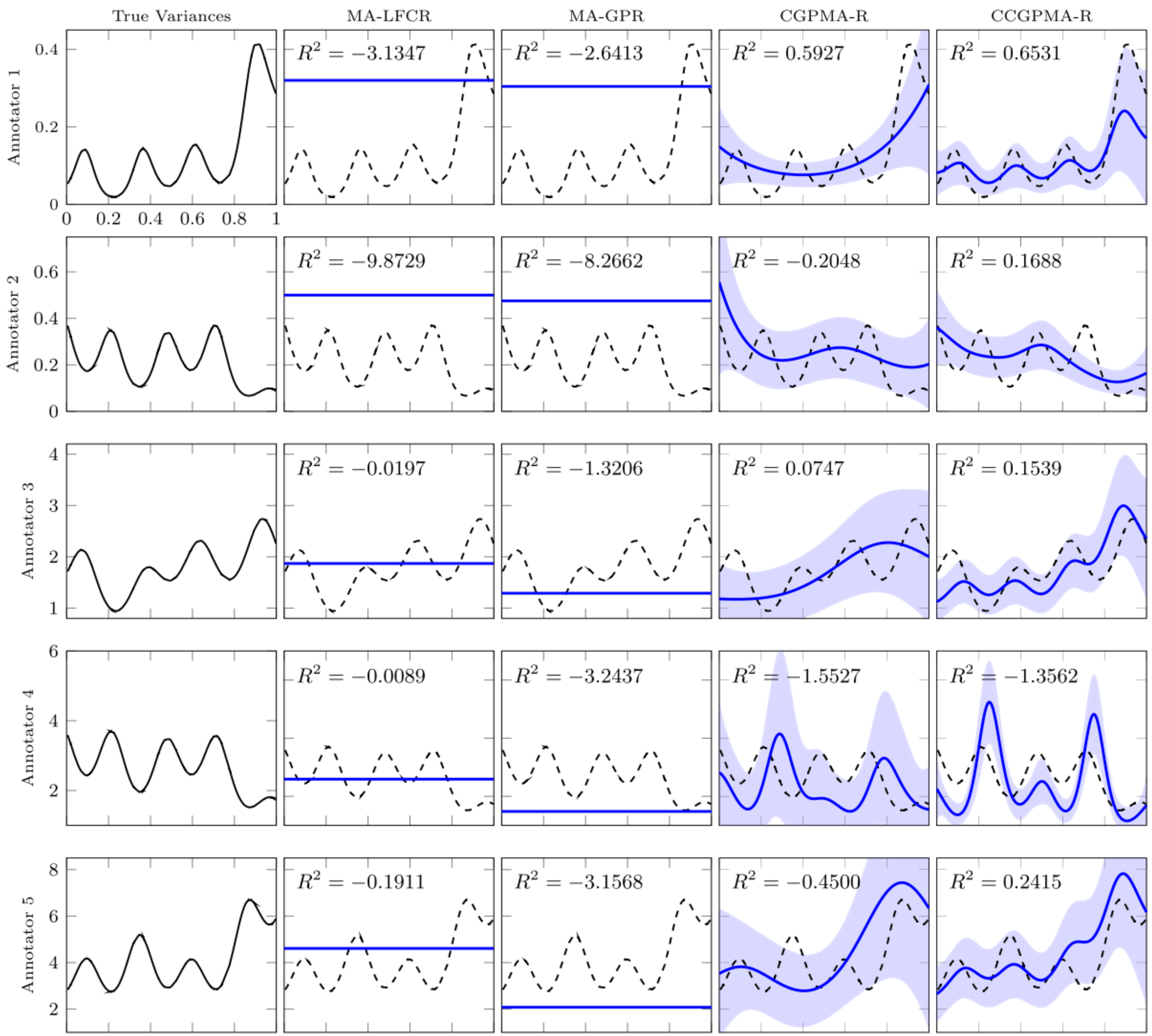
Applications of GP Models

4514 IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, VOL. 34, NO. 8, AUGUST 2023

Correlated Chained Gaussian Processes for Datasets With Multiple Annotators

J. Gil-González, Juan-José Giraldo, A. M. Álvarez-Meza, A. Orozco-Gutiérrez, and M. A. Álvarez

$$p(\boldsymbol{Y}|\boldsymbol{\theta}) = \prod_{n=1}^{N} \prod_{r \in R_n} \mathcal{N}(y_n^r | y_n, v_n^r)$$



Check ref [5].

Applications of GP Models

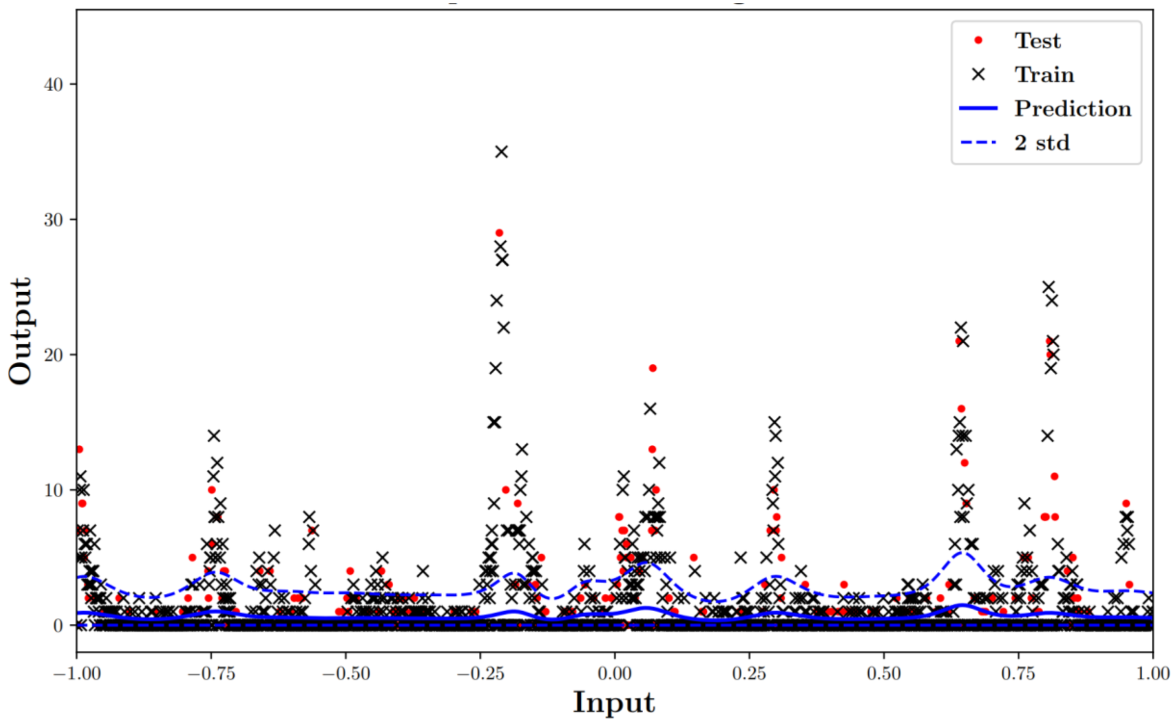
IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, VOL. XX, NO. X, MONTH 20XX

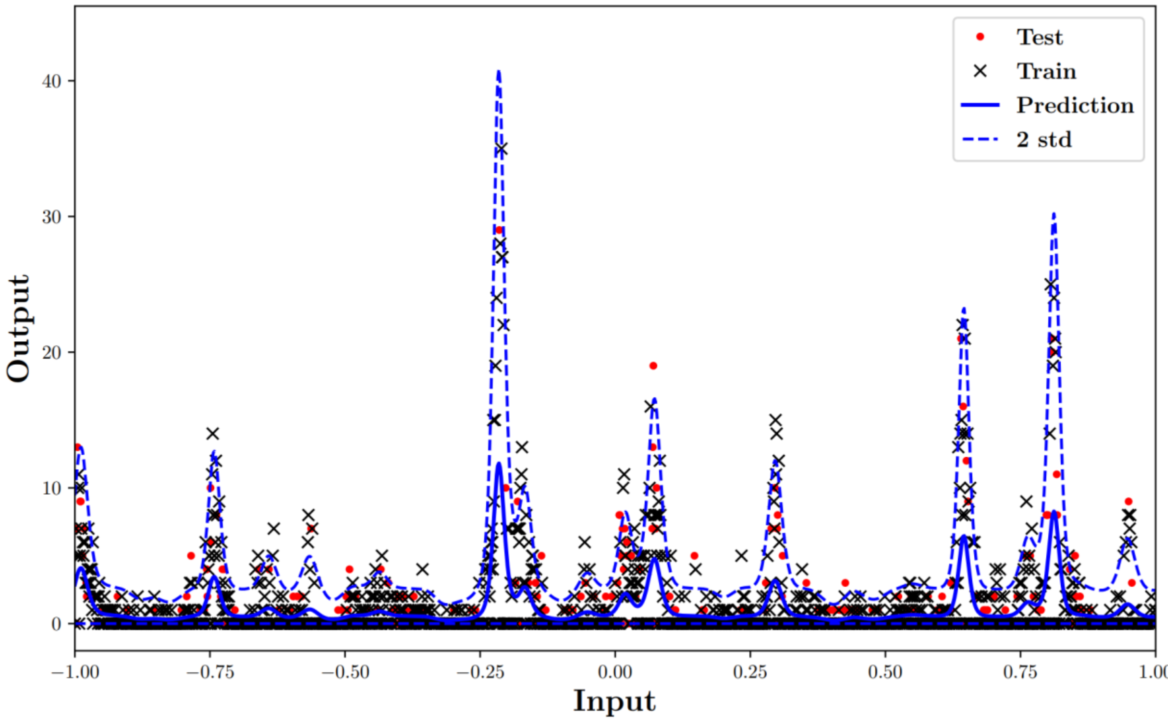
Correlated Chained Gaussian Processes for Modelling Citizens Mobility using a Zero-Inflated Poisson Likelihood

Juan-José Giraldo, Jie Zhang, and Mauricio A. Álvarez

$$p(\mathbf{Y}|\boldsymbol{\theta}) = \prod_{n=1}^{N} p(y_n|\phi_n, \rho_n)$$

$$p(y_n|\phi_n, \rho_n) = \mathbf{1}_{y_n}\phi_n + (1-\phi_n)\frac{\exp(-\rho_n)\rho_n^{y_n}}{y_n!}$$





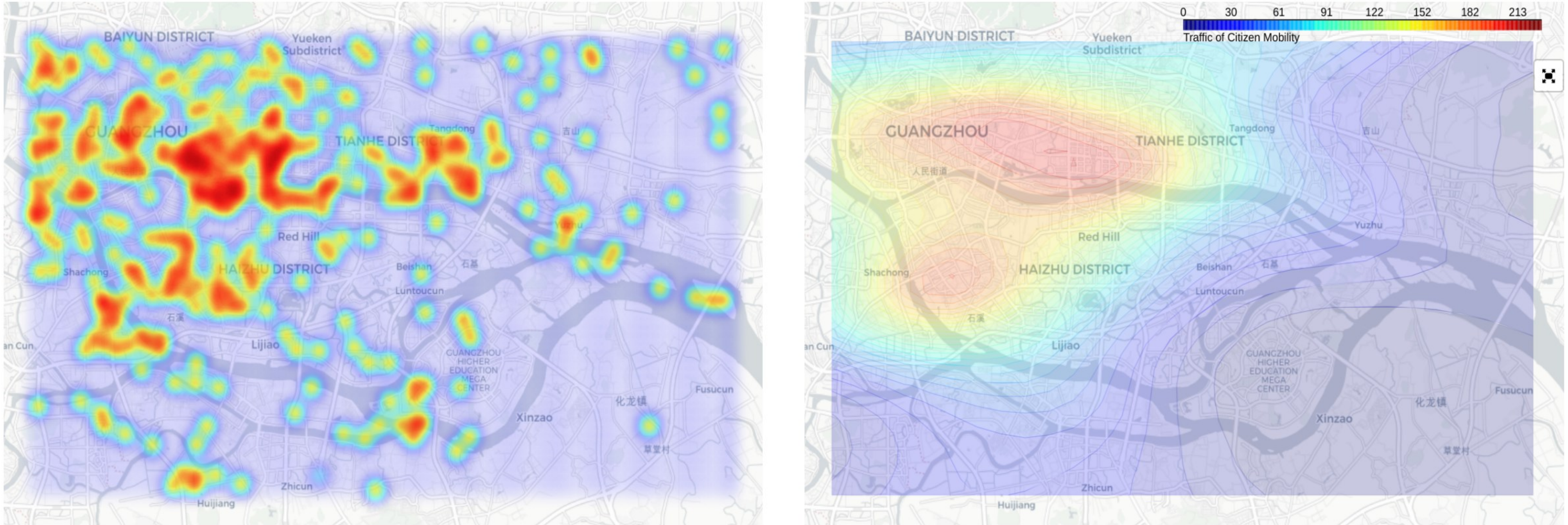
Check ref [4].

Applications of GP Models

IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, VOL. XX, NO. X, MONTH 20XX 1

Correlated Chained Gaussian Processes for Modelling Citizens Mobility using a Zero-Inflated Poisson Likelihood

Juan-José Giraldo, Jie Zhang, and Mauricio A. Álvarez

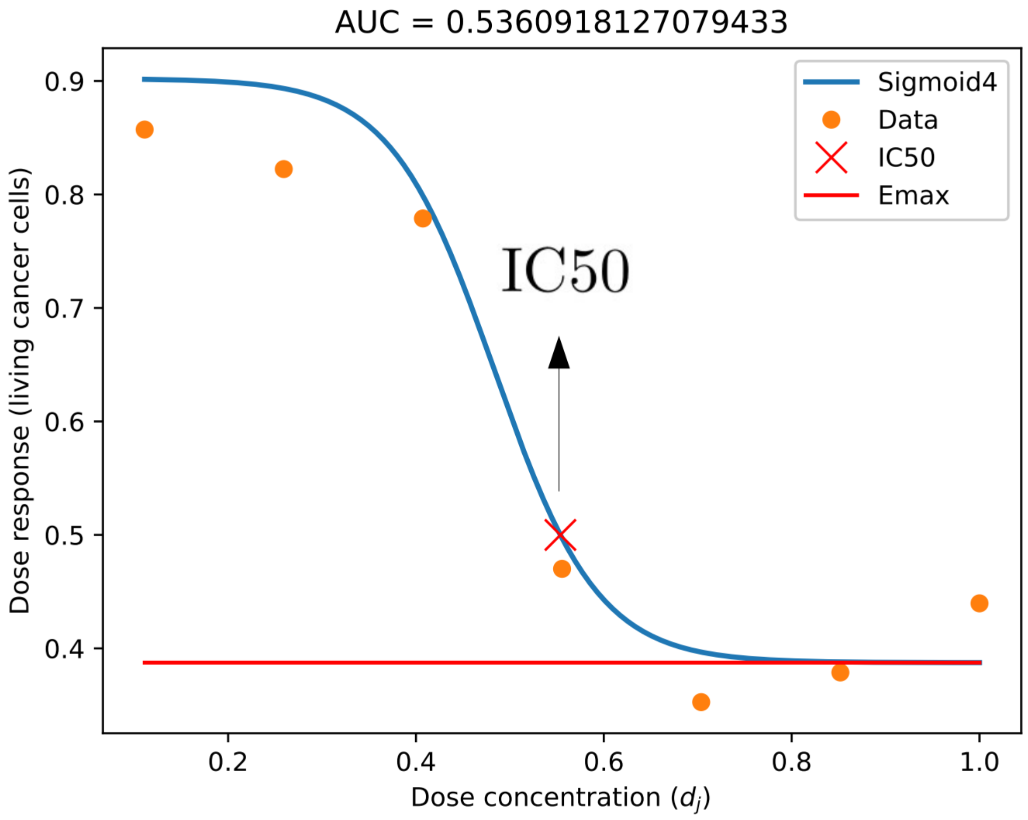


(c) Test Data Heatmap: Saturday, March 16 at 11:00 am

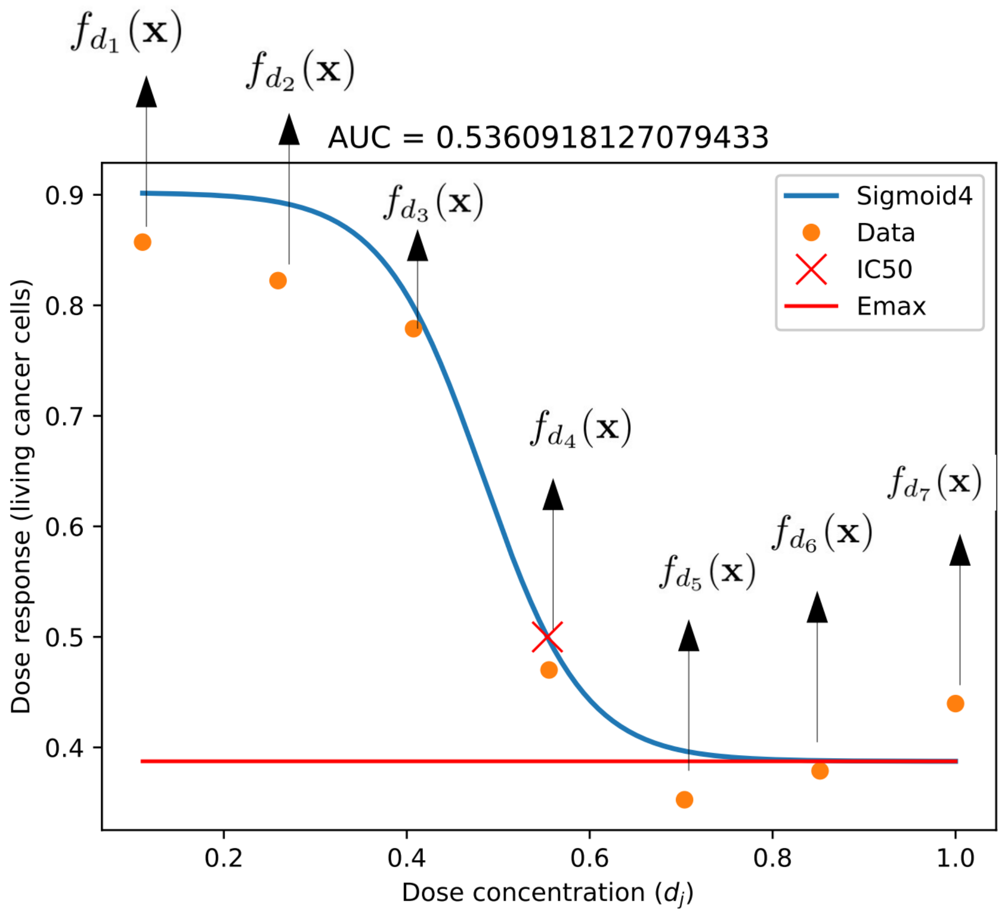
(d) Prediction at 11:00 am using CCGP-VIK with ZI-Poisson

Applications of GP Models: Cancer Research

-Generally summary statistics of observed dose-responses like the drug concentration to achieve 50% cell viability (IC50) have been the main target used to train ML models.



Applications of GP Models: Multi-task Gaussian processes for drug response curves



$$f_{d_j}(\mathbf{x}) = \sum_{q=1}^{Q} w_{d_j,q} u_q(\mathbf{x})$$

$$u_q(\cdot) \sim \mathcal{GP}(0, k_q(\cdot,\cdot))$$

Genomics + drug-compounds:

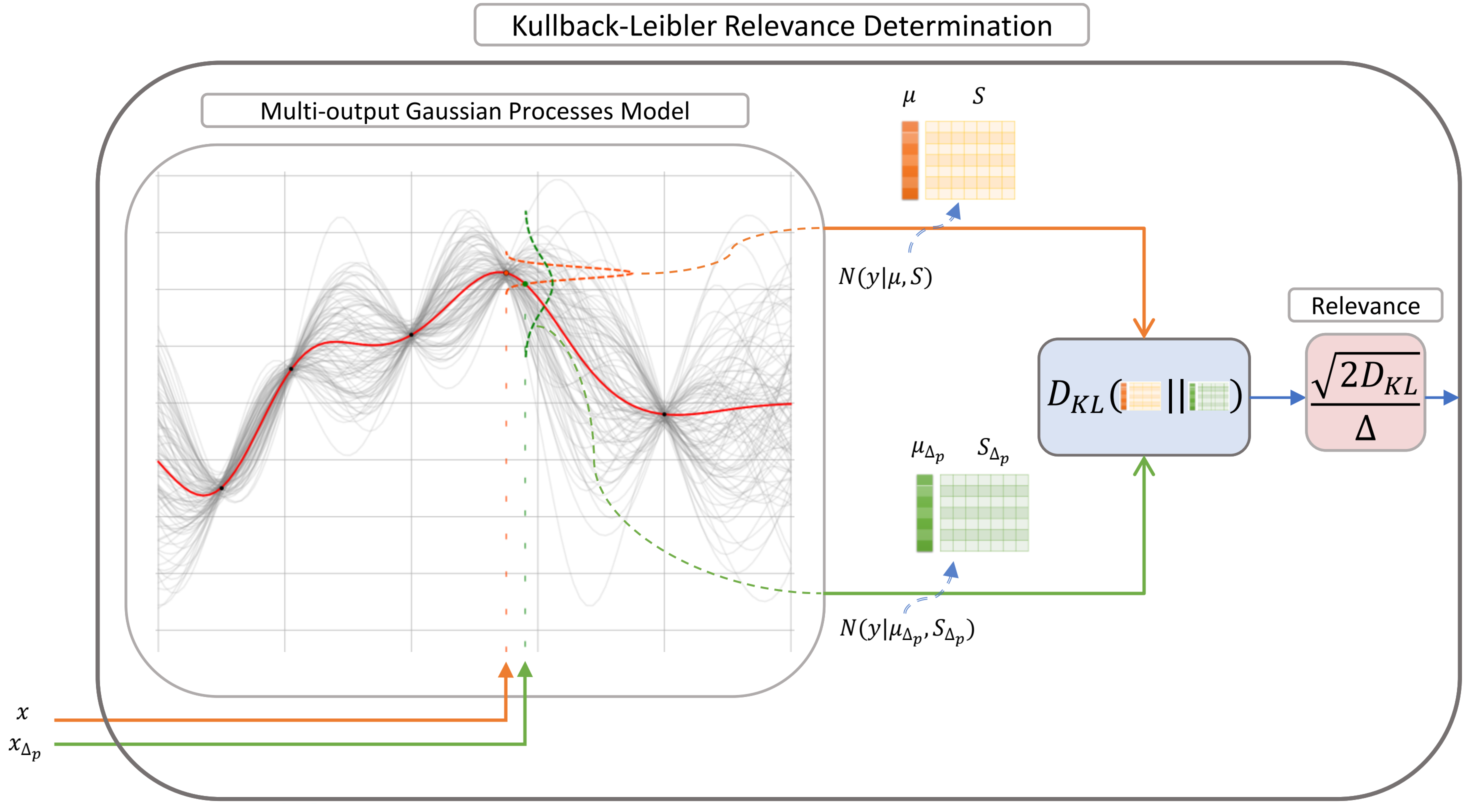
$$\mathbf{x} \in \mathbb{R}^P$$

Multi-task GPs can be studied in [6].

Applications of GP Models: Multi-task Gaussian processes for drug response curves

-In the interest of identifying genomic features as biomarkers that predict sensitivity in a disease like cancer, we would expect to find biomarkers that interact with the full DRC behavior.

-We propose to identify the biomarkers by means of a Kullback-Leibler relevance determination strategy.



Advantages of Gaussian Processes

- Gaussian Process models are **powerful tools** for: Regression and classification.

- GPs typically have **fewer hyperparameters** compared to other machine learning models.

- Powerful for **uncertainty quantification** tasks, particularly:
  In settings with limited data and **complex relationships**.

As research challenges, they may have limitations in terms of:

- Computational **scalability.**

- Particularly when working with **very large datasets or high-dimensional** feature spaces.

Bibliography

[1] C. E. Rasmussen, Gaussian processes for machine learning. MIT Press, 2006.

[2] J. Wang, “An Intuitive Tutorial to Gaussian Processes Regression”, arXiv:2009.10862v4, 2022.

[3] David Duvenaud, James Robert Lloyd, Roger Grosse, Joshua B. Tenenbaum, and Zoubin Ghahramani, “Structure Discovery in Nonparametric Regression through Compositional Kernel Search”, arXiv:1302.4922v4

[4] J-J. Giraldo, J. Zhang and M. Alvarez “Correlated Chained Gaussian Processes for Modelling Citizens Mobility Using a Zero-Inflated Poisson Likelihood”, IEEE Transactions on Intelligent Transportation Systems, 2022.

[5] J. Gil-González , J-J. Giraldo, A. M. Álvarez-Meza, A. Orozco-Gutiérrez, and M. A. Álvarez , Correlated Chained Gaussian Processes for Datasets With Multiple Annotators, IEEE Transactions on Neural Networks and Learning Systems, 20223.

[6] M. A. Alvarez L. Rosasco and N. D. Lawrence "Kernels for vector-valued functions: A review" Found. Trends Mach. Learn. vol. 4 no. 3 pp. 195-266 2012.

j.giraldo-gutierrez@imperial.ac.uk