Gaussian process priors for functions with non-Euclidean domain

Viacheslav Borovitskiy (Slava)

May 12, 2021

Mathematics and Computer Science Department, St. Petersburg University
St. Petersburg Department of Steklov Mathematical Institute

Talk structure

Gaussian Process Regression

Gaussian Process Priors

Non-Euclidean Case

Applications

Conclusion

Gaussian Process Regression

Gaussian Processes

Gaussian processes are function-valued Gaussian random variables.

Defn. Gaussian process (GP) is a family of random variables $\{f(x)\}_{x \in X}$ such that every finite collection $f(x_1), .., f(x_n)$ is jointly Gaussian.

The distribution of a Gaussian process is determined by

- a mean function $m(x) = \mathbb{E}(f(x))$,
- a covariance function $k(x, x') = \mathrm{Cov}(f(x), f(x'))$.
  Note: k is also called a kernel.

Notation: $f \sim \mathrm{GP}(m, k)$.

k must be positive definite, i.e. for $\boldsymbol{x} = x_1, .., x_n$ with $x_i \in X$ the matrix

$$K_{\boldsymbol{xx}} := \{k(x_i, x_j)\}_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}}$$

must always be positive definite.

Gaussian Process Regression

Gaussian Process Regression (GPR) is a regression method that takes in

- a prior Gaussian process distribution $\mathrm{GP}(m, k)$ for $X \to \mathbb{R}$ functions,
- and data $(x_1, y_1), .., (x_n, y_n) \in X \times \mathbb{R}$.

It then yields a posterior (conditional) GP distribution $\mathrm{GP}(\hat{m}, \hat{k})$ that respects the data.

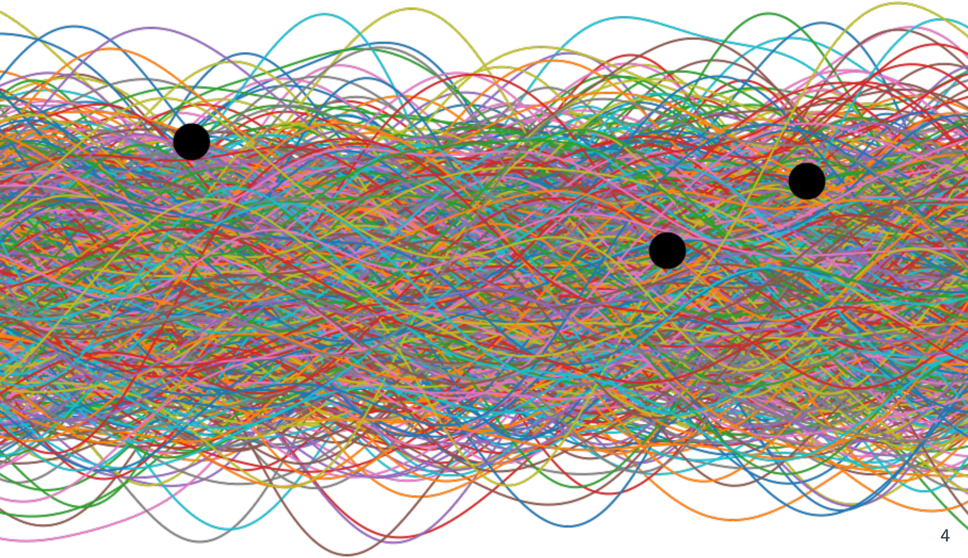
Instead of looking at the formulas, let us explore this visually!

Visual guide to Gaussian process regression

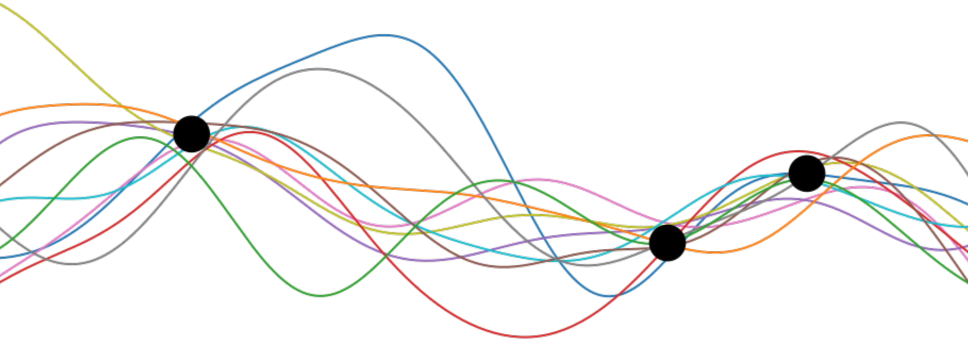
Visual guide to Gaussian process regression

Visual guide to Gaussian process regression

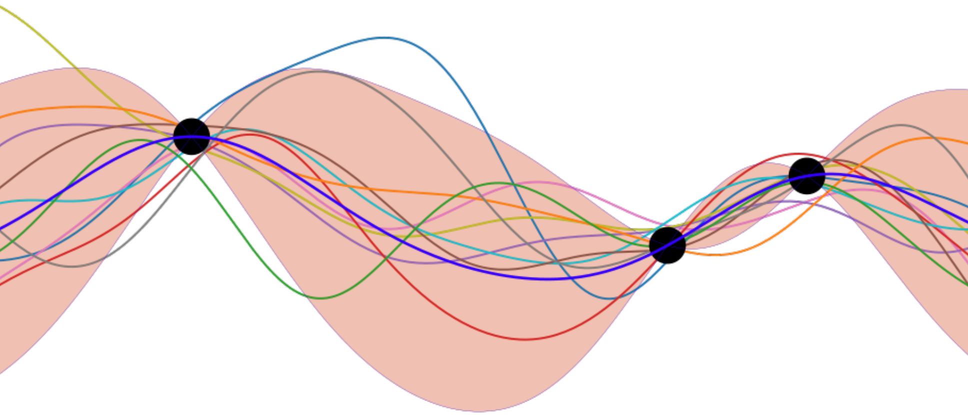
Visual guide to Gaussian process regression



Visual guide to Gaussian process regression



Visual guide to Gaussian process regression



Applications

Bayesian Optimization in AlphaGo

Yutian Chen, Aja Huang, Ziyu Wang, Ioannis Antonoglou, Julian Schrittwieser, David Silver & Nando de Freitas

DeepMind, London, UK

Also

- geostatistcs,
- robotics (dynamical systems modeling),
- more...

Gaussian Process Priors

General Purpose Gaussian Process Priors

Assume we want to model a function $\mathbb{R}^d \to \mathbb{R}$, which prior to take?

When we don't know much about the function, we take $\mathrm{GP}(m, k)$ with

- $m(x) = \mu$ a constant mean function,
- $k(x, x') = k_{\nu,\kappa,\sigma^2}(x, x')$ the Matérn kernel.

Matérn kernels are defined by

$$k_{\nu,\kappa,\sigma^2}(x, x') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{\|x - x'\|}{\kappa} \right)^{\nu} K_\nu \left(\sqrt{2\nu} \frac{\|x - x'\|}{\kappa} \right)$$

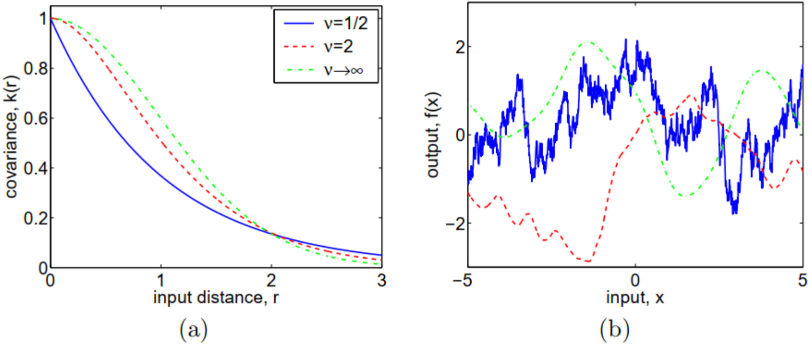
σ^2: variance $\quad$ κ: length scale $\quad$ ν: smoothness

K_ν — modified Bessel function of the 2nd kind.

In the limit $\nu \to \infty$ we also get the famous Gaussian kernel:

$$k_{\infty,\kappa,\sigma^2}(x, x') = \sigma^2 \exp\left(-\frac{\|x - x'\|^2}{2\kappa^2} \right)$$

Visual guide to Matérn kernels



(a) Matérn kernels as functions of $\|x - x'\|$; (b) GP sample paths

Non-Euclidean Case

Generalizing these priors to a non-Euclidean setting

Consider substituting Riemannian geodesic distance $d_M(x, x')$ on some manifold M instead of $\|x - x'\|$, e.g. the following kernel:

$$k_{\infty,\kappa,\sigma^2}(x, x') = \sigma^2 \exp\left(-\frac{d_M(x, x')^2}{2\kappa^2}\right)$$

This doesn't work, the kernel may fail to be positive-definite:

Geodesic Exponential Kernels: When Curvature and Linearity Conflict

Aasa Feragen
DIKU, University of Copenhagen
Denmark
aasa@diku.dk

François Lauze
DIKU, University of Copenhagen
Denmark
francois@diku.dk

Søren Hauberg
DTU Compute
Denmark
sohau@dtu.dk

SPDE Definition of Matérn Gaussian Processes

By Whittle 63, if $f \sim \mathrm{GP}(0, k_{\nu,\kappa,1})$, then it satisfies this particular SPDE:

$$\left(\frac{2\nu}{\kappa^2} - \Delta\right)^{\frac{\nu}{2}+\frac{d}{4}} f = \mathcal{W}.$$

Here $\mathcal{W}$ is the Gaussian white noise, d is the dimension.

The fractional differential operator on the left is made precise through

$$\left(\frac{2\nu}{\kappa^2} - \Delta\right)^{\frac{\nu}{2}+\frac{d}{4}} f = \mathcal{F}^{-1} \left(\frac{2\nu}{\kappa^2} + |\zeta|^2\right)^{\frac{\nu}{2}+\frac{d}{4}} \cdot (\mathcal{F}f)(\zeta)$$

where $\mathcal{F}$ denotes the Fourier transform operator.

For compact Riemannian manifolds

1. Δ — Laplace–Beltrami,
2. $\mathcal{W}$ — white noise controlled by the Riemannian volume.

For weighted undirected graphs

1. Δ — Laplacian matrix,
2. $\mathcal{W}$ — collection of i.i.d. Gaussians, i.e. $\mathrm{N}(0, I)$.

3. $\left(\frac{2\nu}{\kappa^2} - \Delta\right)^{\frac{\nu}{2}+\frac{d}{4}} f = \sum_{n \geq 0} \left(\frac{2\nu}{\kappa^2} + \lambda_n\right)^{\frac{\nu}{2}+\frac{d}{4}} \langle f, f_n \rangle f_n$

Solution of the SPDE

Represent the Gaussian white noise by

$$\mathcal{W} = \sum_{n \geq 0} w_n f_n, \qquad w_n \sim N(0,1) \text{ (i.i.d.)}$$

Then the SPDE can be rewritten in form

$$\sum_{n \geq 0} \left(\frac{2\nu}{\kappa^2} + \lambda_n \right)^{\frac{\nu}{2} + \frac{d}{4}} \langle f, f_n \rangle f_n = \sum_{n \geq 0} w_n f_n.$$

Hence

$$f = \sum_{n \geq 0} \langle f, f_n \rangle f_n = \sum_{n \geq 0} \left(\frac{2\nu}{\kappa^2} + \lambda_n \right)^{-\frac{\nu}{2} - \frac{d}{4}} w_n f_n.$$

Finally

$$k(x, x') = \mathrm{Cov}(f(x), f(x')) = \sum_{n \geq 0} \left(\frac{2\nu}{\kappa^2} + \lambda_n \right)^{-\nu - d/2} f_n(x) f_n(x').$$

Matérn kernels on compact Riemannian manifolds

The kernel: $$k_{\nu,\kappa,\sigma^2}(x,x') = \frac{\sigma^2}{C_\nu}\sum_{n=0}^{\infty}\left(\frac{2\nu}{\kappa^2}+\lambda_{\mathrm{n}}\right)^{-\nu-\frac{d}{2}}\mathrm{f}_{\mathrm{n}}(x)\mathrm{f}_{\mathrm{n}}(x')$$

The process: $$f(x) = \sqrt{\frac{\sigma^2}{C_\nu}}\sum_{n=0}^{\infty}\left(\frac{2\nu}{\kappa^2}+\lambda_n\right)^{-\frac{\nu}{2}-\frac{d}{4}} w_n f_n(x).$$

$\lambda_{\mathrm{n}}, \mathrm{f}_{\mathrm{n}}$ are Laplace–Beltrami eigenpairs.

In some cases (e.g. tori, spheres) — known analytically, in others (e.g. the dragon manifold) may be approximated numerically.

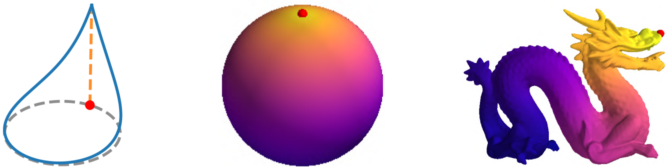


Figure: values of Matérn kernel $k_{1/2}(x,\cdot)$. x is marked with a red dot.

Matérn kernels on finite weighted undirected graphs

The kernel: $$k_{\nu,\kappa,\sigma^2}(i,j) = \frac{\sigma^2}{C_\nu} \sum_{n=0}^{|V|-1} \left(\frac{2\nu}{\kappa^2} + \lambda_\mathrm{n}\right)^{-\nu} \mathrm{f_n}(i)\mathrm{f_n}(j)$$

The process: $$f(i) = \sqrt{\frac{\sigma^2}{C_\nu}} \sum_{n=0}^{|V|-1} \left(\frac{2\nu}{\kappa^2} + \lambda_n\right)^{-\frac{\nu}{2}} w_n f_n(i).$$

$\lambda_\mathrm{n}, \mathrm{f_n}$ are eigenpairs of the Laplacian matrix.

May be evaluated by SVD or an iterative (Lanczos type) algorithm.

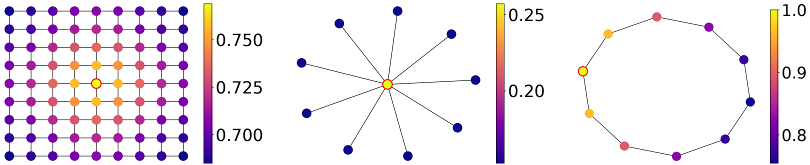
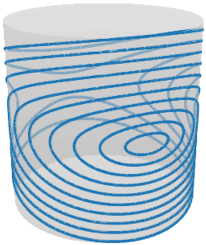


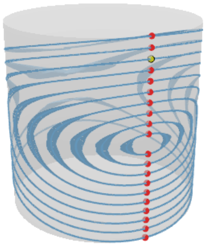
Figure: values of Matérn kernel $k_{5/2}(x, \cdot)$. x node has red outline.

Applications

Pendulum Dynamics — GP on a Cylinder

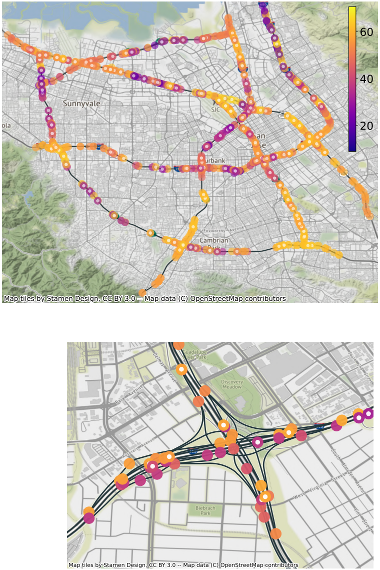


(a) Ground truth

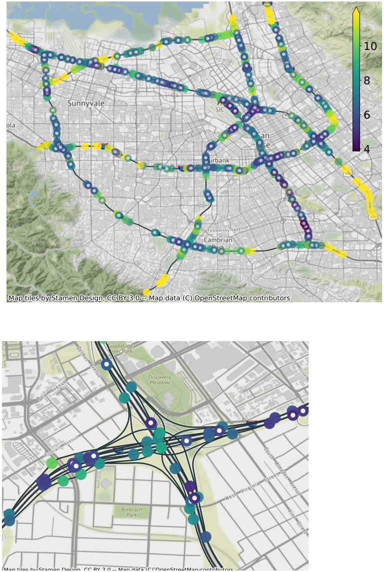


(b) 95%-confidence

Traffic speed interpolation — GP on a Road Network



(a) Mean



(b) Standard deviation

Conclusion

Summary

1. Euclidean general purpose GP priors may be generalized via the SPDE definition, while the distance based approach does not work.
2. In compact Riemannian or finite graph settings, the SPDE may be solved yielding tractable (approximate) kernels.
3. This enables GP based methods to be used in new applications.

Matérn Gaussian processes on Riemannian manifolds

NEURAL INFORMATION PROCESSING SYSTEMS

Viacheslav* Borovitskiy

Alexander* Terenin

Peter* Mostowsky

Marc Deisenroth

St Petersburg University

Imperial College London

UCL

* Equal contribution

NeurIPS 2020

Matérn Gaussian processes on graphs

AISTATS


Viacheslav[*] Borovitskiy


Iskander[*] Azangulov


Alexander[*] Terenin


Peter Mostowsky


Marc Deisenroth


Nicolas Durrande

St Petersburg University

Imperial College London

UCL

Secondmind

[*] Equal contribution

Best Student Paper Award at AISTATS 2021

Thank you for your attention!

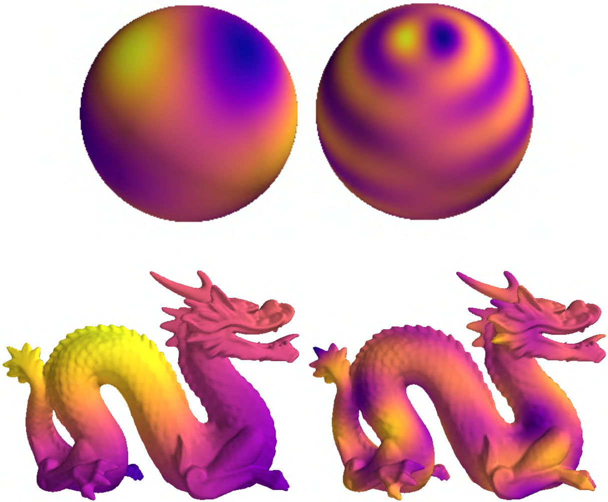
viacheslav.borovitskiy@gmail.com https://vab.im

St Petersburg University

Department of Mathematics and Computer Science

Some figures were taken from: http://inverseprobability.com/talks/.

Visual guide to Laplace–Beltrami eigenfunctions



Sample eigenfunction on the spheredragon

A Gaussian process regression problem on the dragon

(a) Ground truth

(b) Posterior mean

(c) Standard deviation

(d) A sample path