Department of Teleinformatics Engineering , Federal University of Ceará

Introduction to Gaussian Processes

Filipe P. de Farias, IC
filipepfarias@fisica.ufc.br

September 17, 2019

Outline

Gaussian Processes

Gaussian processes, the definition...

- An equivalent way is considering the inference directly in **function-space**. We use a **Gaussian process** (GP) to describe a distribution over functions.

Definition

A **Gaussian process** is a collection of random variables, any finite number of which have a joint Gaussian distribution.

- A Gaussian process is completely specified by its **mean function** and **covariance function** of a real process $f(x)$, defined as

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})], \quad k\left(\mathbf{x}, \mathbf{x}'\right) = \mathbb{E}\left[(f(\mathbf{x}) - m(\mathbf{x}))\left(f\left(\mathbf{x}'\right) - m\left(\mathbf{x}'\right)\right)\right]$$

- Finally we obtain

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

Definition

A function $k : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ is a **Mercer kernel**, if for any finite collection $X = [x_1, \ldots, x_N]$, the matrix $k_{XX} \in \mathbb{R}^{N \times N}$ with elements $k_{XX,(i,j)} = k\left(x_i, x_j\right)$ is positive semidefinite.

Lemma

Any kernel that can be written as

$$\left(T_k f\right)(\mathbf{x}) = \int_{\mathbb{X}} k\left(\mathbf{x}, \mathbf{x}'\right) f\left(\mathbf{x}'\right) d\mu\left(\mathbf{x}'\right)$$

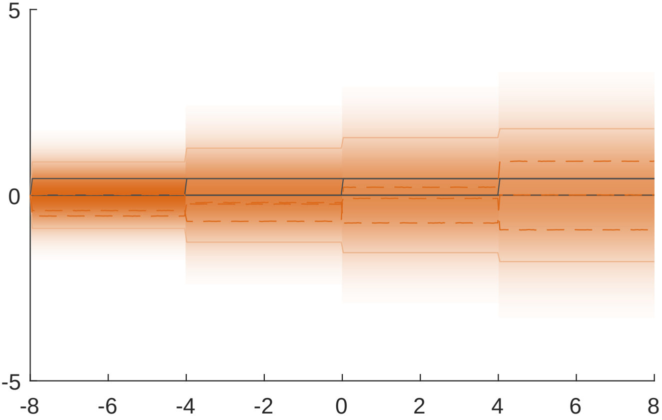
is a Mercer kernel.

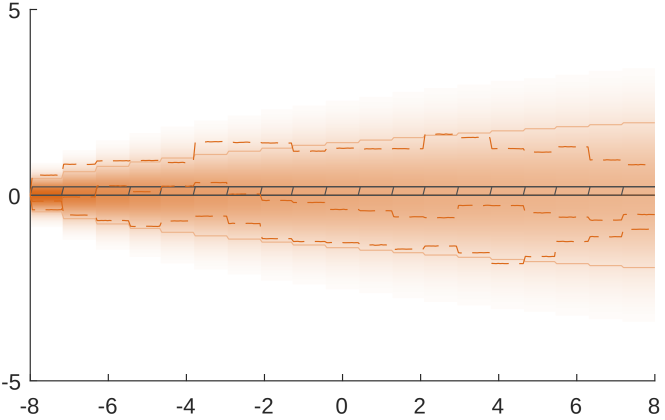
Definition

Let $\mu : \mathbb{X} \to \mathbb{R}$ be any function, $k : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ be a Mercer kernel. A **Gaussian process** $p(f) = \mathcal{GP}(f; \mu, k)$ is a probability over the function $f : \mathbb{X} \to \mathbb{R}$, such that every finite restriction to function values $f_X := \left[f_{x_1}, \ldots, f_{x_N}\right]$ is $p\left(f_X\right) = \mathcal{N}\left(f_X; \mu_X, k_{XX}\right)$.

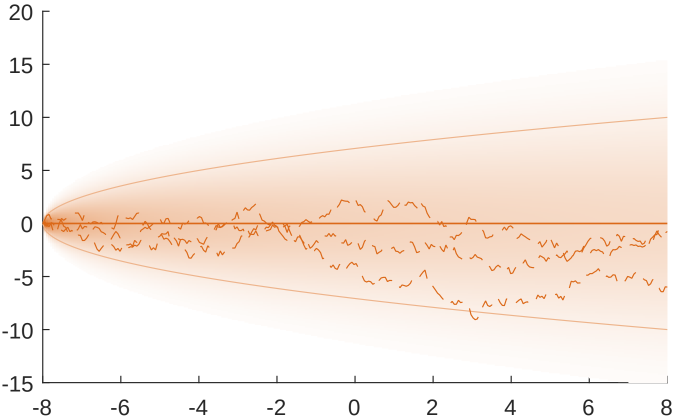
Gaussian Processes

Those step functions



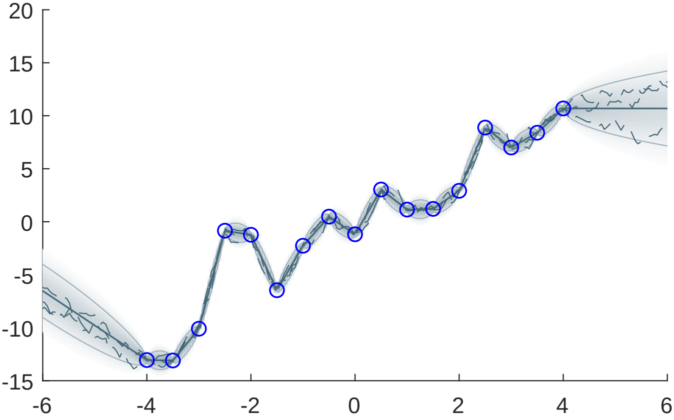

5
0
-5
-8
-6
-4
-2
0
2
4
6
8

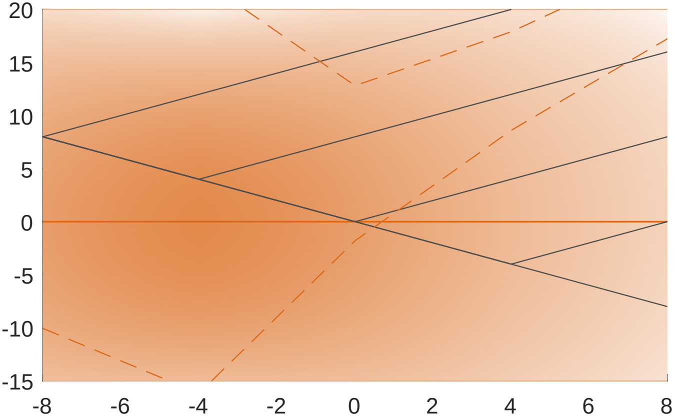
$$\operatorname{cov}\left(f_{x_i}, f_{x_j}\right) = \int_{c_{\min}}^{\infty} \theta\left(x_i - c\right)\theta\left(x_j - c\right)\mathrm{d}c = \min\left(x_i, x_j\right) - c_{\min}$$

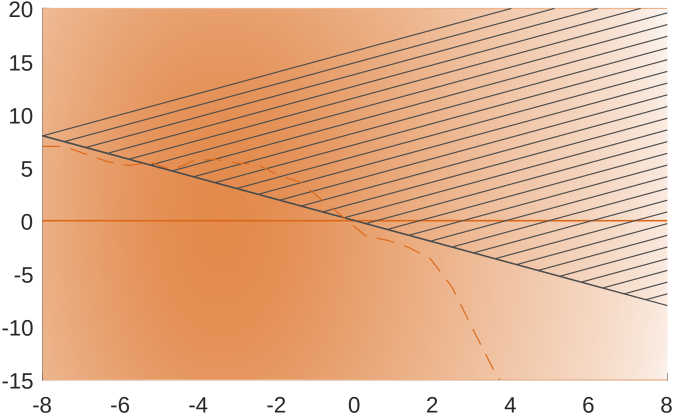


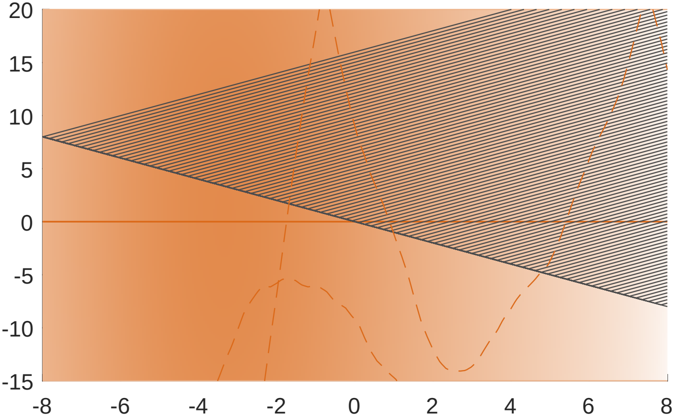
- Posterior

$$\operatorname{cov}\left(f_{x_i}, f_{x_j}\right) = \int_{c_{\min}}^{\infty} \theta\left(x_i - c\right) \theta\left(x_j - c\right) \mathrm{d}c = \min\left(x_i, x_j\right) - c_{\min}$$

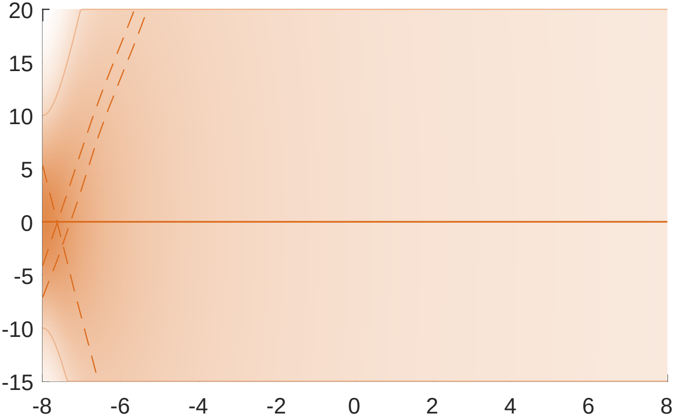



20
15
10
5
0
-5
-10
-15
-8
-6
-4
-2
0
2
4
6
8

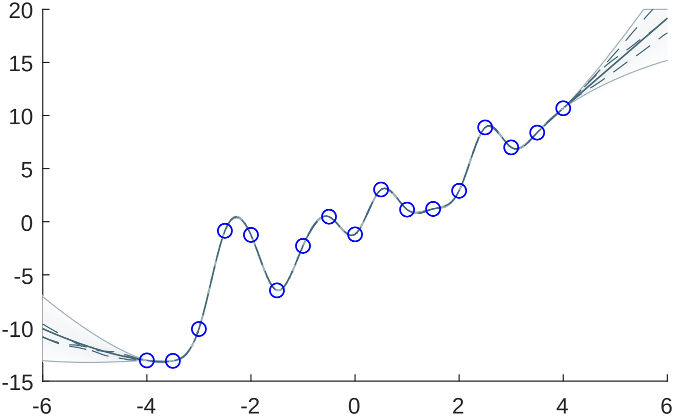

20
15
10
5
0
-5
-10
-15
-8
-6
-4
-2
0
2
4
6
8


20
15
10
5
0
-5
-10
-15
-8
-6
-4
-2
0
2
4
6
8

$$\mathrm{cov}\left(f_{x_i}, f_{x_j}\right) = 1 + (1+b)x_i x_j + \frac{b}{3}\left(\left|x_i - x_j\right|^3 - x^3 - y^3\right)$$



$$\operatorname{cov}\left(f_{x_i}, f_{x_j}\right) = 1 + (1+b)x_i x_j + \frac{b}{3}\left(\left|x_i - x_j\right|^3 - x^3 - y^3\right)$$



Bibliography I

Anil Aksu (Jan. 2018). *Decision Theory and Bayesian Analysis*.

C.M. Bishop (2016). *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer New York. ISBN: 9781493938438. URL: https://books.google.com.br/books?id=kOXDtAEACAAJ.

P.J. Dhrymes (2013). *Mathematics for Econometrics*. SpringerLink : Bücher. Springer New York. ISBN: 9781461481454. URL: https://books.google.com.br/books?id=HIK8BAAAQBAJ.

J. L. Doob (Sept. 1944). "The Elementary Gaussian Processes". In: *Ann. Math. Statist.* 15.3, pp. 229–282. DOI: 10.1214/aoms/1177731234. URL: https://doi.org/10.1214/aoms/1177731234.

F.A. Graybill (2001). *Matrices with Applications in Statistics*. Duxbury Classic Series. Brooks/Cole. ISBN: 9780534401313. URL: https://books.google.com.br/books?id=BV3CAAAACAAJ.

Philipp Hennig (Sept. 2013). *Animating Samples from Gaussian Distributions*. Technical Report 8. Spemannstraße, 72076 Tübingen, Germany: Max Planck Institute for Intelligent Systems.

T.B. Schön (2011). *Manipulating the Multivariate Gaussian Density*. URL: http://user.it.uu.se/~thosc112/pubpdf/schonl2011.pdf.