ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Dr. Francis Colas
Institute of Robotics and Intelligent Systems
Autonomous Systems Lab

ETH Zürich
CLA E 26 fcolas@mavt.ethz.ch
Tannenstraße 3 www.asl.ethz.ch
8092 Zürich Switzerland

Information Processing in Robotics
Solution Sheet 5
Topic: Gaussian Process

Exercise 1: Implementation of a Gaussian Process for regression

(a) A Gaussian process is a probability distribution over functions. It is specified by a mean function and a kernel function. Additionally, for regression we need the observation noise.

(b) When a point is observed, we can update the $\boldsymbol{C_N}$ matrix, as well as store the new $\boldsymbol{x}, t$ pair.

(c) See code.

(d) We can get the mean and variance for any point $\boldsymbol{x}$. See code.

(e) See demo.

Exercise 2: Sampling from a Gaussian Process

(a) As $y(x)$ is a Gaussian Process, the distributions are Gaussians:

- $p(y(x_1)) = \mathcal{N}(y_1|\mu(x_1), k(x_1, x_1))$,
- $p(y(x_2)) = \mathcal{N}(y_2|\mu(x_2), k(x_2, x_2))$,
- $p(y(x_1), y(x_2)) = \mathcal{N}(\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} | \begin{pmatrix} \mu(x_1) \\ \mu(x_2) \end{pmatrix}, \begin{pmatrix} k(x_1, x_1) & k(x_1, x_2) \\ k(x_2, x_1) & k(x_2, x_2) \end{pmatrix})$.

(b) With the particular values for the Gaussian process $u(x)$ we have:

- $p(u(x_1)) = \mathcal{N}(u_1|0, 1)$,
- $p(u(x_2)) = \mathcal{N}(u_2|0, 1)$,
- $p(u(x_1), u(x_2)) = \mathcal{N}(\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} | \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix})$.

We can see that the covariance matrix is diagonal, that means that $u(x_1)$ and $u(x_2)$ are independent. It is the case for any set of index as the kernel function is diagonal. This process is also a Gaussian white noise.

(c) The Gram matrix relative to a kernel function is the application of the kernel function for each ordered pair of points. It is also the covariance matrix of the image of the points taken into account by the Gaussian process.

(d) We have $z(x_1) = \mu(x_1) + \boldsymbol{L}u(x_1)$, with μ a given function, and u a Gaussian process. We have therefore an affine transformation of a Gaussian distribution: $p(z(x_1))$ is a Gaussian distribution.

Moreover, $\boldsymbol{L}$ is the lower triangular matrix of the Cholesky factorisation of the Gram matrix $\boldsymbol{K} = (k(x_1, x_1))$ (remember that k refers to the Gaussian process y), that is $L = \sigma_1 = \sqrt{k(x_1, x_1)}$.

If we apply what we know of the affine transform of a (multivariate) Gaussian distribution, we have:

$$p(z(x_1)) = \mathcal{N}(z_1|\mu(x_1) + \boldsymbol{L}E[u(x_1)], \boldsymbol{L}\text{cov}[u(x_1)]\boldsymbol{L}^T)$$

However, $E[u(x_1)] = 0$ and cov$[u(x_1)] = 1$ Therefore

$$p(z(x_1)) = \mathcal{N}(z_1|\mu(x_1), \boldsymbol{L}\boldsymbol{L}^T) = \mathcal{N}(z_1|\mu(x_1), k(x_1, x_1))$$

Similarly:

- $p(z(x_2)) = \mathcal{N}(z_2|\mu(x_2), k(x_2, x_2))$
- $p(z(x_1), z(x_2)) = \mathcal{N}(\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} | \begin{pmatrix} \mu(x_1) \\ \mu(x_2) \end{pmatrix}, \begin{pmatrix} k(x_1, x_1) & k(x_1, x_2) \\ k(x_2, x_1) & k(x_2, x_2) \end{pmatrix})$.

That is, the distribution on $z(x_1)$ and $z(x_2)$ are the same as $y(x_1)$ and $y(x_2)$ with the same properties (mean and covariance).

(e) With the affine transform, we went from Gaussian white noise, to a given Gaussian process. If we want to sample from such a process, we can therefore do as follows:

- choose a set of indexes $(x_1, \cdots, x_N) = \boldsymbol{x}^T$,
- compute the Cholesky decomposition L of the Gram matrix defined by the kernel of the process,
- get N independent samples $(u_1, \cdots, u_N) = \boldsymbol{u}^T$ of the Normal distribution $\mathcal{N}(u|0, 1)$,

- and apply the affine transformation $\boldsymbol{y} = \boldsymbol{\mu}(x) + \boldsymbol{L}\boldsymbol{u}$.

The resulting vector $\boldsymbol{y}$ will be a valid sample of the Gaussian process.

Note that it is not an iterative process. To have an iterative process, you have to consider the conditional probability of the new point given what is already sampled. It is the normal prediction process of the regression. Therefore another sampling method is to first draw one sample then, for each new point we want to add, to compute the prediction of the regression with the same kernel of the points already sampled.