

---

# On Gaussian Processes for Regression

---

Anonymous Author(s)

Affiliation

Address

email

## Abstract

1 Gaussian processes emerged in machine learning as a powerful tool for regression  
2 and classification that provides interpretability through kernel choice and uncer-  
3 tainty quantification. By leveraging properties of multivariate normal distributions  
4 and Bayes’s rule, we may infer a probability distribution over possible functions  
5 when fitting a dataset. This Bayesian framework allows flexibility through choosing  
6 a covariance function as a prior belief about the dataset, which can provide further  
7 insight into the trends of the training data. We implement a multi-dimensional Gaus-  
8 sian process regressor and evaluate its performance on the Boston Housing dataset,  
9 which is comparable to those in the top 25 of the Kaggle competition. Furthermore,  
10 we perform optimization on the hyperparameters through maximum likelihood  
11 estimation, to remove the need for manual tuning of the hyperparameters.

## 12 1 Gaussian Random Variables

13 A random variable is a function that maps from an event space to a measurable space. The event  
14 space represents a set of all possible outcomes that the random variable may take, and the measurable  
15 space is a probability measure between 0 and 1 (inclusive). We say that a random variable  $X$  is  
16 normally distributed if the event space has a Gaussian probability distribution, fully characterized by  
17 two parameters: a mean  $\mu$  and variance  $\sigma^2$ :

$$X \sim \mathcal{N}(\mu, \sigma^2).$$

18 For a one-dimensional Gaussian random variable, we refer to its distribution as a univariate Gaussian  
19 distribution. A set of Gaussian random variables may be characterized jointly as a multivariate  
20 Gaussian distribution, with joint probability distribution fully characterized by a mean vector and a  
21 covariance matrix:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \sim \mathcal{N}(\mu, \Sigma).$$

22 where  $\mu$  is the mean vector, and  $\Sigma$  is the covariance matrix whose entries describe the covariance  
23 between each pair of random variables.

## 24 2 Gaussian Process

25 A random process is essentially a collection of random variables jointly characterized as a set or  
26 vector of random variables with a multivariate joint probability distribution. A Gaussian process  
27  $f(x)$  is defined as a random process where each set of random variable in the random process is

28 has a multivariate Gaussian distribution.  $f(\mathbf{x})$  is fully characterized by a mean function  $m(\mathbf{x})$  and  
 29 covariance function,  $K(\mathbf{x}, \mathbf{x}')$ :

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

30 Typically, the mean function is zero. The kernel for the covariance is chosen based on some prior  
 31 belief about the dataset; more on kernels is discussed in 3.1.

## 32 **3 Regression**

33 Suppose we observe training data  $\mathbf{t}$  and choose kernel  $\kappa$ . Then the mean and covariance functions are  
 34 given by

$$m(\mathbf{x}) = C_{\mathbf{x}\mathbf{t}}^\top C_{\mathbf{t}}^{-1} \mathbf{t}$$

$$K(\mathbf{x}, \mathbf{x}') = C_{\mathbf{x}\mathbf{x}'} - C_{\mathbf{x}\mathbf{t}}^\top C_{\mathbf{t}}^{-1} C_{\mathbf{x}\mathbf{t}}$$

35  
 36 where  $C_{\mathbf{x}\mathbf{t}} = \kappa(\mathbf{x}, \mathbf{t})$ ,  $C_{\mathbf{t}} = \kappa(\mathbf{t}, \mathbf{t})$ , and  $C_{\mathbf{x}\mathbf{x}'} = \kappa(\mathbf{x}, \mathbf{x}')$ . Those interested in the derivation of the  
 37 results are encouraged to consult section 6.4.2 of [1].

### 38 **3.1 Kernels**

## 39 **References**

40 References follow the acknowledgments. Use unnumbered first-level heading for the references. Any  
 41 choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font  
 42 size to small (9 point) when listing the references. **Note that the Reference section does not count**  
 43 **towards the eight pages of content that are allowed.**

44 [1] C. M. Bishop, Pattern recognition and machine learning. New York: Springer, 2006.