Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 12, No. 1, 2025

Application of Composite Kernel Function in Gaussian Process Prediction

Yulin Zhang *

School of Science, Tianjin University of Commerce, Tianjin, China

*** Corresponding author Email:** cherry888yu@163.com

Abstract: Gaussian Process is a non-parametric Bayesian method based on kernel functions, widely used in regression tasks and Bayesian optimization. Kernel functions play a crucial role in Gaussian processes as they directly impact the model's ability to fit the data distribution. Different kernel functions can be selected based on the characteristics of the data to improve model performance. In practical applications, combining multiple basic kernel functions can construct more complex kernel functions, enhancing the flexibility and adaptability of Gaussian processes, making them better suited to complex data modeling challenges. This paper uses California housing price data as an example to analyze the impact of different kernel function combinations on model prediction performance. The experimental results validate the importance of kernel selection in Gaussian process models and demonstrate the effectiveness of composite kernels in handling complex data distributions.

Keywords: Gaussian Process; Composite; Kernel Function.

1. Introduction

Gaussian Process [1] is a non-parametric Bayesian method widely used in regression tasks and Bayesian optimization [2,3]. The kernel function (or covariance function) in a Gaussian Process is one of its core components, defining the similarity between two input data points and determining the covariance structure of the Gaussian Process, thereby influencing the predicted mean and variance [4]. Different kernel functions have different smoothness and periodicity properties, affecting the Gaussian Process's ability to fit data. Thus, the choice of kernel function impacts the model's generalization ability—incorrect kernels may lead to overfitting or underfitting [5]. Kernel functions are interchangeable, and appropriate kernel functions can be selected or designed based on data characteristics and task requirements. In many cases, composite kernel functions [6] are considered. Composite kernels combine multiple base kernels to create a new kernel function, enhancing the performance of Gaussian Process models and enabling them to fit more complex data patterns [7]. Common ways to construct composite kernels include additive and multiplicative combinations. Organization of the Text

2. Gaussian Process Regression

2.1. Model

Gaussian Process is a non-parametric Bayesian method used to model stochastic processes. Essentially, it is an infinite-dimensional probability distribution operating in function space, where the function values at any set of input points follow a joint Gaussian distribution. This property allows Gaussian Processes to probabilistically describe function values at different points for prediction and analysis.

Given an input space $\mathcal{X}$, a Gaussian Process is defined as:

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \tag{1}$$

where $m(\mathbf{x})$ is the mean function (typically assumed to be zero), and $k(\mathbf{x}, \mathbf{x}')$ is the kernel function (or covariance function), which describes the similarity between different input points $\mathbf{x}$ and $\mathbf{x}'$.

2.2. Prediction

Given a set of training data $D = \{(\mathbf{x}_i, y_i)\}$,, where $\mathbf{x}_i \in \mathcal{X}$ is the input and the corresponding target value $y_i \in \mathbb{R}$ consists of the true function value generated by the Gaussian Process and observation noise:

$$\begin{cases} \epsilon_i \sim \mathcal{N}(0, \sigma_n^2) \\ y_i = f(\mathbf{x}_i) + \epsilon_i \end{cases} \tag{2}$$

For a new test point $\mathbf{x}_*$, in order to predict $y_* = f(\mathbf{x}_*) + \epsilon_*$, we can calculate the conditional probability distribution of $f(\mathbf{x}_*)$ based on the training data D. Among them, the joint distribution of training data and the test point is given by:

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{y}_* \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} K(X,X) + \sigma_n^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) + \sigma_n^2 I \end{pmatrix}\right) \tag{3}$$

By applying the conditional distribution property of Gaussian Processes, we obtain the predicted mean and covariance matrix:

$$\begin{cases} \mu_* = K(X_*, X)(K(X,X) + \sigma_n^2 I)^{-1}\mathbf{y} \\ \Sigma_* = K(X_*, X_*) - K(X_*, X)(K(X,X) + \sigma_n^2 I)^{-1} K(X, X_*) \end{cases} \tag{4}$$

2.3. Kernel Functions

Kernel functions define the similarity between input data points and determine the covariance structure of Gaussian Processes, directly affecting model generalization. Incorrect kernels can lead to overfitting or underfitting. Common kernel functions include: Radial Basis Function, Matern Kernel.

(1) Radial Basis Function (RBF): $k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{||\mathbf{x} - \mathbf{x}'||^2}{2\ell^2}\right)$

(2) Matern Kernel: $k(\mathbf{x},\mathbf{x}') = \frac{1}{\Gamma(\nu)}(\frac{\sqrt{2\nu}||\mathbf{x}-\mathbf{x}'||}{\ell})^{\nu}K_{\nu}(\frac{\sqrt{2\nu}||\mathbf{x}-\mathbf{x}'||}{\ell}) = (1+\frac{||\mathbf{x}-\mathbf{x}'||^2}{2\alpha\ell^2})^{-\alpha}$

Selecting and optimizing kernel functions is one of the most critical steps in Gaussian Process regression as it directly impacts prediction performance and generalization ability.

3. Empirical Study

3.1. Kernel Selection

In this article, we used three different combination kernel functions for prediction in order to compare and evaluate their differences in model performance. Through this approach, we can observe the impact of different combinations of kernel functions on the prediction results.

(1) ConstantKernel * RBF

ConstantKernel defines a constant kernel that determines the overall scaling between data. Radial Basis Functio defines the RBF kernel, which represents how the similarity between data points varies with distance. The product of these two kernels represents the similarity between data points, taking into account both the constant scaling factor and the variation of similarity in the input space with distance.

(2) Matern + WhiteKernel:

The Matern kernel can control the smoothness of the function. WhiteKernel defines the white noise kernel and simulates the noise in observed data. The addition of these two kernels is suitable for data with certain noise, while the Matern kernel can fit incomplete smooth data changes.

(3) Rational Quadratic + WhiteKernel:

The Rational Quadratic kernel can handle multi-scale variations and is suitable for situations where data changes are uneven. WhiteKernel also defines a white noise kernel to represent noise in the data. These two kernels are applicable to data with multi-scale variations and take into account the noise in the data.

3.2. Evaluation Metrics

This article chooses RMSE to evaluate the effectiveness of the model. RMSE is a very useful regression model evaluation tool, mainly used to measure the difference between the predicted and actual values of the model, especially suitable for situations with large errors. By calculating RMSE, we can obtain the average error between model predictions and actual values, and compare multiple models to select the best performing model.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2} \tag{5}$$

3.3. Application

The dataset used is the California Housing dataset, which includes housing prices from different regions of California. The dataset contains geographic, population, and income features, with the median house price as the target variable.

By using Gaussian process regression models with three different combinations of kernel functions for prediction, and observing the impact of different combinations of kernel functions on the prediction results.

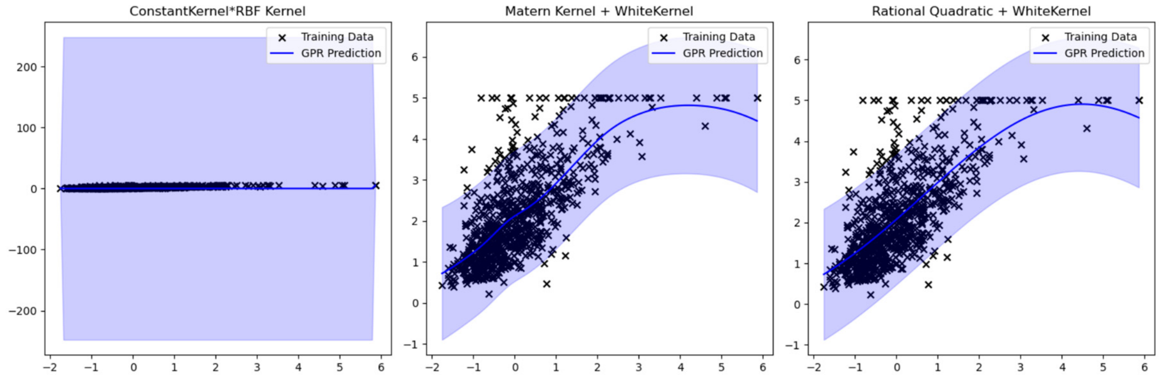

Fig 1. House price prediction performance under different kernel functions

Figure 1 shows the housing price prediction performance under different kernel functions. When the kernel function is ConstantKernel * RBF, the predicted results (blue line) are almost horizontal, indicating that the model did not capture the trend of the data well. This may be because the RBF kernel is too smooth on this dataset, which makes it difficult for the model to fit the complexity of the data well, and the confidence interval (blue shaded area) is very wide, indicating that the model has a high degree of uncertainty in the predicted results. The prediction results of Matern Kernel+ White Kernel capture the trend of the data well, showing a clear nonlinear relationship, and the confidence interval is relatively narrow, indicating that the model has less uncertainty in the prediction results and the prediction is more reliable. The prediction results of Rational Quadratic+ WhiteKernel also capture the trend of the data well and show a non-linear relationship. However, compared with the previous combination kernel function, the fitting effect is slightly different, and the confidence interval is relatively narrow, indicating that the model has less uncertainty in the prediction results and the prediction is more reliable.

Table 1. Forecasting error analysis.

	ConstantKernel * RBF	Matern Kernel + WhiteKernel	Rational Quadratic + WhiteKernel
RMSE	2.2745	0.8225	0.8231

Table 1 presents the RMSE of different models. When using the ConstantKernel * RBF kernel function, the model has a higher prediction error on the test set. In this case, the model's performance may be affected by data noise or feature selection, and it may not be able to capture the patterns of the data well. Both Matern Kernel+WhiteKernel and Rational Quadratic+WhiteKernel perform well on this dataset, with RMSE fluctuating around 0.82. Among them, Matern Kernel+WhiteKernel has the smallest RMSE, indicating that this method is more accurate. In summary, the predicted values of both kernel functions can serve as references.

4. Summary

Single-scale kernels like RBF have limitations in capturing multi-scale variations. Matern and Rational Quadratic kernels provide better flexibility in balancing local details and global trends, while handling noise more effectively. Choosing appropriate kernel combinations significantly impacts the model’s ability to generalize.

This study reveals that different combinations of kernel functions have a significant impact on the fitting performance of models when data exhibits complex nonlinear characteristics. Therefore, choosing the appropriate kernel function is crucial for improving the generalization ability and stability of the model. In addition, Gaussian Process Regression (GPR), as a probabilistic modeling technique, can not only predict but also estimate variance, which is an important reference for market analysis and risk management. Compared with traditional regression methods, the advantage of Gaussian process regression is that it can flexibly adapt to the characteristics of data by selecting different kernel functions, making the prediction results more accurate and easy to interpret.

References

[1] MATTHIAS, SEEGER. Gaussian processes for machine learning [J]. International Journal of Neural Systems, 2004, 14(2): 69-106.

[2] Lyu C,Liu X, Mihaylova L .Review of Recent Advances in Gaussian Process Regression Methods[C].UK Workshop on Computational Intelligence. Springer, Cham, 2024.

[3] Zhao Mengen. Research on Improved Gaussian Process Regression Algorithm and Its Application [D]. Zhejiang Sci-Tech University, 2025.

[4] Campbell C, Ying Y. Learning with Kernels[J]. 2011.

[5] Wilson A G, Adams R P. Gaussian Process Kernels for Pattern Discovery and Extrapolation[J]. Microtome Publishing, 2013, 13 (4): 129-135.

[6] Duvenaud D. Automatic model construction with Gaussian processes [J]. 2014.

[7] Zhou Huiling. Research on Stock Index Volatility Prediction Based on Gaussian Process Regression of Combined Kernel Function [D]. Hunan University, 2021.