Deep Multi-Fidelity Gaussian Process for 2-Dimensional Cases

Zilin Chen & Liu Yang

November 20, 2017

Abstract

We develop a 2-dimensional multi-fidelity structure framework, based on the work Deep Multi-Fidelity Gaussian Process. We managed to handle 2-dimensional discontinuities with information from different fidelities. The neural network captures the discontinuities and affine transformations in the data, while the gaussian process presents the global regression. To show the power of our technique, several examples with different kind of discontinuities are presented.

1 Motivation

Multi-fidelity gaussian process has an advantage in merging informatin from different sources with a few samples, which is promising to be used in imaging. One limitation of gaussian process is the disability of handling discontinuities. Neural network makes up this weakpoint and greatly enlarge the flexibility of the whole model. A special layer in neural network is also introduced to deal with the affine difference between data from different sources. Combing these, our model succeeds in dealing with some simple discontinuities and shows a forseeable usage in meany areas.

2 Introduction

Gaussian process is a useful supervised learning method assuming the problems obey gaussian distribution. It is powerful when we are approximating smooth functions using few samples. To combine it with neural network, we mainly focus on Manifold Gaussian Process which increases the flexibility of elementary model. The main benefit of Manifold Gaussian Process is that the data to be learned will first be transformed into a feature space which in our framework is learned supervisedly and then a normal gaussian process is utilized. This helps our framework find expressive representation of data. The feature space we applied is a deep neural network which can explain various relationships. Since multi-fidelity information are always available, multi-fidelity gaussian process is essential for merging such information. A special affine layer should be added on different fidelity data in order to make the framework robust to accept information with a minor transformation.

2.1 Manifold Gaussian Process

Normal Gaussian Process is a distribution over functions

$$f \sim \mathcal{GP}(0, k) \tag{1}$$

determined by its covariance function k.
We can get predictive distribution of an input x_* with available information $\mathbb{D}$

$$p(f(x_*)|\mathbb{D}, x_*) = \mathcal{N}(\mathbf{k}_*^T(\mathbf{K} + \sigma_w^2\mathbf{I})^{-1}\mathbf{Y}, k_{**} - \mathbf{k}_*^T(\mathbf{K} + \sigma_w^2\mathbf{I})^{-1}\mathbf{k}_*) \tag{2}$$

with training data $\mathbb{D} = (\mathbf{X}, \mathbf{Y})$, kernel matrix $\mathbf{K}$ and measurement noise variance σ_w^2.The kernel matrix $\mathbf{K}$ is defined by covariance function k, where $\mathbf{K}_{ij} = k(x_i, x_j)$. The most common choice of covariance function is squared exponential covariance function

$$k_{SE}(x_p, x_q) = \sigma_f^2 exp(-\frac{1}{2}(x_p - x_q)^T\Lambda^{-1}(x_p - x_q)) \tag{3}$$

with $\Lambda = diag(l_1^2, l_2^2, \dots, l_n^2)$. σ_f,σ_w and $l_1, l_2, \dots, l_n$ are all the hyperparameters to be trained. We denote these hypermeters as θ.
To train θ, we minimize the Negative Log Marginal Likelihood(NLML)

$$\text{NLML} = \frac{1}{2}f^T\mathbf{K}^{-1}f + \frac{1}{2}log|\mathbf{K}| + \frac{n}{2}log2\pi \tag{4}$$

Above is the whole step of an elementary Gaussian Process. What Manifold Gaussian Process do is first transform the data into a feature space.

$$x \to h(x)$$

And we do the Gaussian Process with $h(\mathbf{X})$ and $\mathbf{Y}$. This opeartion greatly inceases the flexibility of Gaussian Process.

2.2 Multi-Fidelity Gaussian Process

Without loss of gemerality, we talk about two levels of fidelity. The merit of Multi-Fidelity Gaussian Process is that it can merge information from different fidelity. Namely, we can use less expensive information and more cheap information to get good performance compared to elementary Gaussian Process. The two fidelities work as follow

$$f_2(x) = \rho f_1(x) + \delta_2(x) \tag{5}$$

with two independent Gaussian process $f_1(x)$ and $\sigma_2(x)$

$$f_1(x) \sim \mathcal{GP}(0, k_1), \delta_2(x) \sim \mathcal{GP}(0, k_2) \tag{6}$$

After computation, we know

$$f_2(x) \sim \mathcal{GP}(0, \rho^2 k_1 + k_2) \tag{7}$$

and the joint distribution function is

$$\begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} = \mathcal{GP}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix} \right) \tag{8}$$

The training is the same as the elementary Gaussian Process.

2.3 Deep Multi-Fidelity with Affine Layer

In our framework, we use deep neural network as the transformation from data to feature space. The deep neural network we use is a simple feedforward neural network with four layers and six units in each layer. To deal with some affine transformation in low fidelity data, we add an affine layer to capture these transformation in our model. We denote h as the feedforward network and a as the affine layer. Then the relationship of different fidelity data is

$$f_2(h(x) = \rho f_1(h(a(x))) + \delta_2(h(x)) \tag{9}$$

We use stochastic optimization based on Keras.

3 Experiments

We demonstrate several artificial examples to show the power of our method. We first show that the affine layer managed to capture the affine transformation as we have expected. Namely, the trained affine layer's parameter is very similar to what we set by hands. Then we show the power of method in two dimensional cases. We test both linear discontinuity and circular discontinuity in two example without affine transformation. Finally, we test a circular discontinuity function with affine transformation in low fidelity data.

3.1 1-Dimensional Mulitifidelity with Affine Transformation

To be finished.

3.2 2-Dimensional Step Function

The high fidelity data is

$$f_h(x_1, x_2) = \chi_{>0.5}(x_1) + \chi_{>0.5}(x_2) \tag{10}$$

And the low fidelity data is

$$f_l(x_1, x_2) = 0.8 \cdot f_h(x_1, x_2) - 4 \tag{11}$$

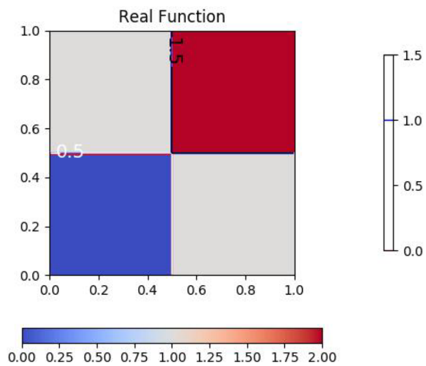

Figure 1: Real Function of 2-Dimensional Step Function

The plot of high fidelity function is Figure 1. The domain is $[0, 1] \times [0, 1]$, and to generate the training data, we choose points evenly from the data. We demonstrate the result with 4^2 high fidelity sample and 10^2 low fidelity sample. The red points are high fidelity samlples and the yellow points are low fidelity samples.

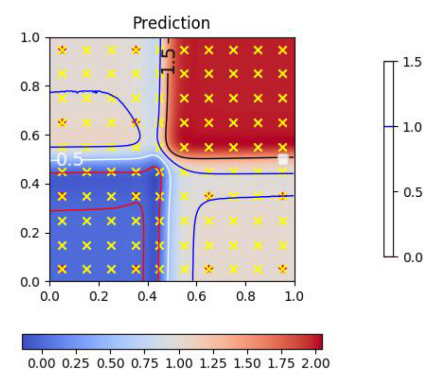

Figure 2: Prediction of 2-Dimensional Step Function

The squared error is Figure 3. The most error happens around the discontinuity ,and the other places look good.

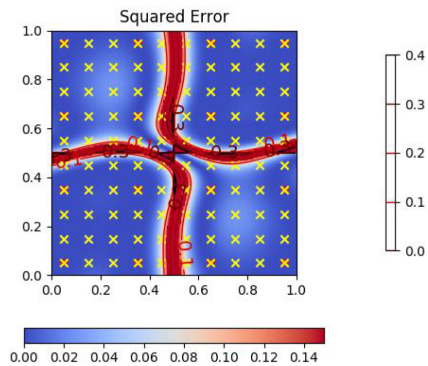

Figure 3: Squared Error of 2-Dimensional Step Function

3.3 2-dimensional Function with Circular Discontinuity

The low fidelity data is

$$f_l(x_1, x_2) = \begin{cases} x_1 \cdot x_2 + cos(2\pi \cdot x_1^2) - sin(2\pi \cdot x_1^2) + 10 \cdot x_1 - 5 & r > 0.5 \\ x_1 \cdot x_2 + cos(2\pi \cdot x_1^2) - sin(2\pi \cdot x_1^2) & r \leq 0.5 \end{cases} \tag{12}$$

And the high fidelity data is

$$f_h(x_1, x_2) = 0.2x_1 \cdot x_2 + 0.7 \cdot f_l(x_1, x_2) \tag{13}$$

All is added a noise $\sigma \sim \mathcal{N}(0, 0.5)$. The plot of high fidelity function is Figure 4. The domain is $[-1, 1] \times [-1, 1]$, and to generate the training data, we choose

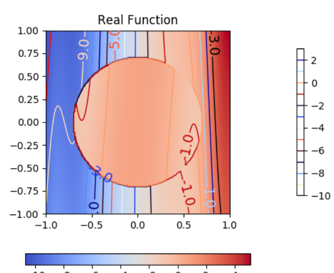

Figure 4: Real Function of Circular Discontinuity Function

points evenly from the data. We demonstrate the result with 12^2 high fidelity sample and 14^2 low fidelity sample. The red points are high fidelity samlples and the yellow points are low fidelity samples. The squared error is Figure 6.

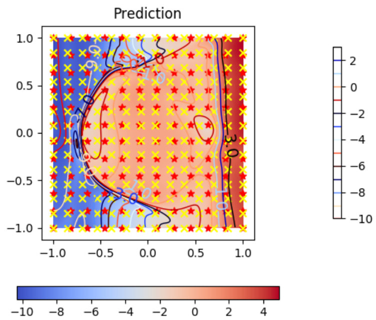

Figure 5: Prediction of Circular Discontinuity Function

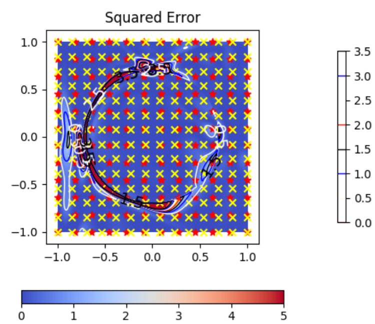

Figure 6: Squared Error of 2-Dimensional Step Function

3.4 2-dimensional Function with Circular discontinuity with Affine Transformation

To be finished.

References

[1] Maziar Raissi, George Karniadakis, Deep Multi-Fidelity Gaussian Process, arXiv preprint:arXiv:1604.07484v1, 2016

[2] C. E. Rasmussen, C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006