Gaussian Process Regression

Introduction, Comparison and Analysis

Tzu-Heng Lin, 2014011054, W42
Department of Electronic Engineering, Tsinghua University, Beijing, China
lzhbrian@gmail.com

ABSTRACT

[1] Gaussian Process Regression is a powerful, non-parametric tool developed based on the Bayesian theory and the Statistical learning theory. Choosing the right *Mean Functions*, *Kernel Functions* as well as the *Likelihood Functions* and the *Inference Methods* have been critical to the performance of the model. However, these works are often hard and time-consuming.

In this paper, we first give an introduction on the overall process of the Gaussian Process Regression. We also summarize some of the recent works which emphasize on the automatic construction of the *Kernel Function*. In addition, we implement sufficient number of experiments to systematically analyze the performance of different *Mean Functions*, *Likelihood Functions* and the *Inference Methods*. Our experiments are conducted on two interesting datasets. We seek to provide an comprehensive practical overview on the field of Gaussian Process Regression.

1. INTRODUCTION

Machine learning has been a heated research topic these days. With this amazing tool, we are now capable of predicting the price of the stock price based on history, doing the classifying by just inputing the pixels of images.

Supervised learning is one of the most important sections for machine learning. And Regression is possibly the core of supervised learning.

In Gaussian Process Regression, we take advantage of the flexibility and simplicity of Gaussian Process and implement it into a regression problem.

The structure of this pape is as follows: In section 3 we provide an overview of Gaussian Process Regression. Section 4 describes some of the recent works on the methods for auto-construction of the kernels. In section 5 we conduct two experiments. Conclusions are drawn in section 8.

2. RELATED WORK

3. GAUSSIAN PROCESS REGRESSION

Regression Regression is probably one of the most fundamental problems in a wide range of fields including *Statistics*, *Signal Processing* and *Machine Learning*, etc. A regression problem is usually formulated as follows: Given a training set $D = \{(\mathbf{x}_i, y_i) | i = 1, 2, ..., n\}$, we assume that $\mathbf{x}_i, y_i$ have the following relationship:

$$y_i = f(\mathbf{x}_i) + e \tag{1}$$

where e is the error noise. By finding such $f(\cdot)$, we can predict what a corresponding y^* is in some test case $\mathbf{x}^*$. Note that $\mathbf{x}$ can either be a vector or a scalar.

Generalized Linear Model A widely used regression model is called Generalized Linear Model(GLM)[5], in which a regression function can be expressed as a linear combination:

$$f(\mathbf{x}) = \Sigma_{i=1}^{M} w_i \phi_i(\mathbf{x}) \tag{2}$$

where $\phi_i(x)$ is called the basis function. In a regular GLM analysis, we have to firstly determine what our basis functions we are going to use, and subsequently can we use the training dataset to derive the parameters in the basis functions and the coefficients in the regression function.

Mean Square Error We use a measurement called the Mean Square Error(MSE) to evaluate the performance of the regression function. It is defined as follows:

$$MSE = \frac{1}{m} \Sigma_{i=1}^{m} (f(\mathbf{x}^*) - y_i^*)^2 \tag{3}$$

where $f(x)$ represents the regression function. A smaller MSE represents a better regression function on a test set.

Gaussian Process

Gaussian Process Regression Gauss Process Regression(GPR)[9] is a popular regression method these years, the key of this method is to model the regression function $\{f(\mathbf{x}) | \mathbf{x} \in S\}$ as a Gauss Process $N\{m(\mathbf{x}), K(\mathbf{x}, \mathbf{x}^{'})\}$, where $m(\mathbf{x})$ is the *Mean function* and $K(\mathbf{x}, \mathbf{x}^{'})$ is the *Kernel Function*. In a GPR, we don't have to derive the exact form of the regression function, we just need to determine the form of the above two functions. By calculating the posterior probability of the desired $f(\mathbf{x}^*)$, *i.e.* $p(f(\mathbf{x}^*) | \mathbf{x}^*, D)$, We can derive the mean value of the estimation, along with the standard deviation of this estimation. In fact, the *Kernel Function* $K(\mathbf{x}, \mathbf{x}^{'})$ represents the correlation between $f(\mathbf{x})$ and $f(\mathbf{x}^{'})$.

On the other hand, calculating the posterior probability is a intractable work when we have a high-dimensional dataset. It is a need that we introduce several inference method to estimate this work. Some popular works include MCMC[3], Expectation Propagation[6], Variational Bayes[8, 7] and Laplace Approximation[10], etc.

[1] Tzu-Heng Lin is currently an undergraduate student in the Department of Electronic Engineering, Tsinghua University. His research interests include Big Data Mining, Machine Learning, etc. For more information about me, please see my personal website. Please feel free to contact me at any time via email

4. AUTOMATIC CONSTRUCTION OF KERNEL FUNCTIONS

Choosing an appropriate *Kernel Function* in a regression problem has always been a

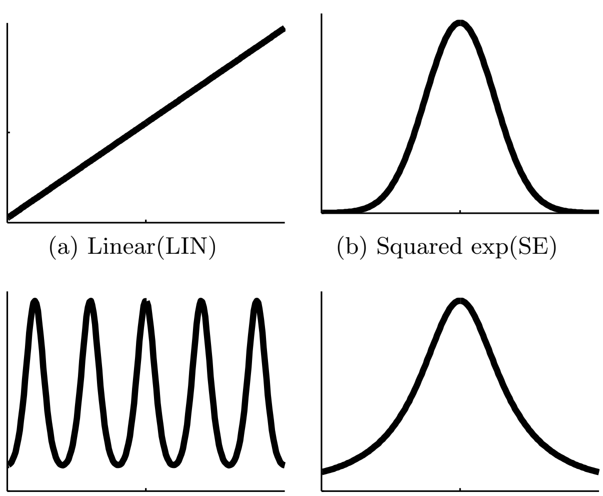
This section is based on the work of Duvenaud and Lloyd et. al [2, 4, 1]

Base Kernels As we all know, different kernels can represent different kinds of data. In Figure 1, we show 4 different kinds of base kernels, each represents a specific kind of data characteristics (Table 1). Each function of the kernel is shown in Equation 4

$$\begin{cases} k_{LIN}(x, x^{'}) = \sigma_b^2 + \sigma_v^2(x-l)(x^{'}-l) \\ k_{SE}(x, x^{'}) = \sigma^2 exp(-\dfrac{(x-x^{'})^2}{2l^2}) \\ k_{PER}(x, x^{'}) = \sigma^2 exp(-\dfrac{2sin^2(\pi(x-x^{'})/p)}{l^2}) \\ k_{RQ}(x, x^{'}) = \sigma^2(1+\dfrac{(x-x^{'})^2}{2\alpha l^2})^{-\alpha} \end{cases} \quad (4)$$

Kernel	**Data Characteristic**
Linear(LIN)	linear functions
Squared exponential(SE)	local variation
Periodic(PER)	repeating structure
Rational quadratic(RQ)	multi-scale variation

Table 1: Different kinds of kernels and its represented data characteristics.



(a) Linear(LIN) (b) Squared exp(SE)

(c) Periodic(PER) (d) Rational quadratic(RQ)

Figure 1: Base Kernels

Compositional Kernels

Automatic Construction

5. EXPERIMENTS

Our experiments are conducted on two interesting datasets,

6.

7.

8. CONCLUSION

In this paper, we discuss about the Gaussian Process Regression.

We systematically compare three methods of partition function estimation which are crucial works in training a Restricted Bolztmann Machine or a Deep Belief Network.

As future work, we would like to join more methods to the comparison and if could, propose some improvement to the algorithms available.

9. ACKNOWLEDGEMENT

I would like to thank Yuanxin Zhang, XueChao Wang, for the discussion with me on the algorithms. Without them, I wouldn't have the possibility to accomplish this work in such a short time. This paper is a project of Stochastic Process Course in Tsinghua University, taught by Prof. Zhijian Ou.

Source Code and Dataset Source Code to perform all experiments, along with the dataset in this paper can be found in my github repository[2].

[2]Available at http://github.com/lzhbrian/gpr

10. REFERENCES

[1] D. Duvenaud. *Automatic model construction with Gaussian processes*. PhD thesis, University of Cambridge, 2014.

[2] D. K. Duvenaud, J. R. Lloyd, R. B. Grosse, J. B. Tenenbaum, and Z. Ghahramani. Structure discovery in nonparametric regression through compositional kernel search. In *ICML (3)*, pages 1166–1174, 2013.

[3] D. Gamerman. Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing*, 7(1):57–68, 1997.

[4] J. R. Lloyd, D. Duvenaud, R. Grosse, J. B. Tenenbaum, and Z. Ghahramani. Automatic construction and natural-language description of nonparametric regression models. *arXiv preprint arXiv:1402.4304*, 2014.

[5] P. McCullagh. Generalized linear models. *European Journal of Operational Research*, 16(3):285–292, 1984.

[6] T. P. Minka. Expectation propagation for approximate bayesian inference. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pages 362–369. Morgan Kaufmann Publishers Inc., 2001.

[7] H. Nickisch and M. W. Seeger. Convex variational bayesian inference for large scale generalized linear models. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 761–768. ACM, 2009.

[8] J. Palmer, D. Wipf, K. Kreutz-Delgado, and B. Rao. Variational em algorithms for non-gaussian latent variable models. *Advances in neural information processing systems*, 18:1059, 2006.

[9] C. E. Rasmussen. Gaussian processes for machine learning. 2006.

[10] L. Tierney and J. B. Kadane. Accurate approximations for posterior moments and marginal densities. *Journal of the american statistical association*, 81(393):82–86, 1986.