

# **Bayesian Model Class Selection on Regression Problems**

by  
**Mu He-Qing**



**Master of Science in Civil Engineering**

**July 2010**



**Faculty of Science and Technology  
University of Macau**

University of Macau

Abstract

BAYESIAN MODEL CLASS SELECTION  
ON REGRESSION PROBLEMS

by He-Qing Mu

Thesis Supervisor: Prof. Ka-Veng Yuen

Regression analysis, a technique used for modeling the relationship between input and output data, is widely used for prediction and forecasting. But in most cases, especially the case for constructing the empirical model, it's very hard to select the model for prediction among a set of model class candidate. Even for same set of data, different people may use different assumption and different model class to do the regression analysis. So, based on different selected model, it's not surprising that people may get the prediction result with significant difference.

There are two levels of regression problem to be considered although they are strongly related. The first level is parametric identification with a specified model class. A metric is defined to measure the distances between observed and predicted data, by minimizing this distances, the optimal parameter value can be obtained. The second level is on the selection of model class. Quite a number of techniques have been developed and used to select the model based on the data.

In this thesis, Bayesian model class selection technique is used to identify the optimal parameter value as well as their uncertainty, and to select the most plausible model class based on the data. Two different methods developed from this technique will be adopted to analyze the regression problems. The first one is the discrete Bayesian model selection. The second one uses the machine learning theory, and performs continuous model class selection to the linear regression model. These two methods are used to analyze the strong-motion record of China.

# Table of Contents

<b>List of Tables</b>	<b>iv</b>
<b>List of Figures</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Regression Problems	1
1.1.1 Linear Regression Problems	1
1.1.2 Nonlinear Regression Problems	2
1.2 Model Class Selection	2
1.2.1 Other Methods	3
1.2.2 Bayesian Model Class Selection	3
1.3 Organization of this Thesis	6
<b>2 Bayesian Inference</b>	<b>7</b>
2.1 Bayesian Inference	7
2.1.1 Prior Distribution	9
2.1.2 Bayesian Parametric Identification and Model Identifiability	12
2.1.3 Posterior Robust Predictive PDFs	13
2.1.4 Inference on Additional Data	14
2.2 Regression Problems	15
2.2.1 Linear Regression Problems	15

2.2.2	Nonlinear Regression Problems . . . . .	19
<b>3</b>	<b>Bayesian Model Class Selection</b>	<b>21</b>
3.1	Bayesian Model Class Selection . . . . .	22
3.2	Calculation of the evidence . . . . .	23
3.2.1	Information-theoretic interpretation and comparison with other information criterion . . . . .	24
3.2.2	Bayesian model averaging . . . . .	25
3.3	Model Class Selection for Regression Problems . . . . .	26
3.3.1	Linear Regression Problems . . . . .	26
3.3.2	Nonlinear Regression Problems . . . . .	31
<b>4</b>	<b>Bayesian Inference and Model Selection in Machine Learning</b>	<b>34</b>
4.1	Automatic Relevance Determination Prior . . . . .	34
4.2	Sparse Bayesian Learning for Regression Problem . . . . .	35
<b>5</b>	<b>Peak Ground Acceleration Estimation by Linear and Nonlinear Regression Equations</b>	<b>39</b>
5.1	Introduction . . . . .	39
5.2	Candidates of Predictive Model Classes . . . . .	41
5.3	Selection of the Predictive Model Class . . . . .	43
5.3.1	Selection of the prior PDF . . . . .	44
5.3.2	Computation of the evidence integral . . . . .	45
5.4	Analysis of the Strong Ground Motion Records . . . . .	49
5.4.1	Description of the database . . . . .	49
5.4.2	Tangshan region . . . . .	51
5.4.3	Xinjiang region . . . . .	60
5.4.4	Full set of data . . . . .	66
5.5	Conclusion . . . . .	74

<b>6 Conclusion and Future Work</b>	<b>75</b>
6.1 Conclusion . . . . .	75
6.2 Future Work . . . . .	76
<b>A Posterior PDF and Evidence by using Bayes' Theorem</b>	<b>77</b>
<b>B Hyperparameter <math>\alpha</math> Optimization</b>	<b>80</b>
<b>C Noise Variance <math>\sigma^2</math> Optimization</b>	<b>82</b>
<b>Bibliography</b>	<b>85</b>

