بسم الله الرحمن الرحيم

For my parents.

Abstract

Regression analysis is undoubtedly an important tool to understand the relationship between one or more explanatory and independent variables of interest. In this thesis, we explore a novel methodology for fitting a wide-range of parametric and non-parametric regression models, called the I-prior methodology (Bergsma, 2018).

We assume that the regression function belongs to a reproducing kernel Hilbert or Kreĭn space of functions, and by doing so, it allows us to utilise the convenient topologies of these vector spaces. This is important for the derivation of the Fisher information of the regression function, which might be infinite-dimensional. Based on the principle of maximum entropy, an I-prior is an objective Gaussian process prior for the regression function with covariance function proportional to its Fisher information.

Our work focusses on the statistical methodology and computational aspects of fitting I-priors models. We examine a likelihood-based approach (direct optimisation and EM algorithm) for fitting I-prior models with normally distributed errors. The culmination of this work is the R package **iprior** (Jamil, 2017) which has been made publicly available on CRAN. The normal I-prior methodology is subsequently extended to fit categorical response models, achieved by "squashing" the regression functions through a probit sigmoid function. Estimation of I-probit models, as we call it, proves challenging due to the intractable integral involved in computing the likelihood. We overcome this difficulty by way of variational approximations. Finally, we turn to a fully Bayesian approach of variable selection using I-priors for linear models to tackle multicollinearity.

We illustrate the use of I-priors in various simulated and real-data examples. Our study advocates the I-prior methodology as being a simple, intuitive and comparable alternative to similar leading state-of-the-art models.

Keywords: regression, model, reproducing kernel, Hilbert space, Kreĭn space, Fréchet derivative, Gâteaux derivative, Fisher information, binary, multinomial, Gaussian process, variational inference, empirical Bayes, EM algorithm, MCMC, truncated normal

http://phd.haziqj.ml • http://myphdcode.haziqj.ml

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 65,899 words.

I confirm that Chapters 2 and 3 were jointly co-authored with Dr. Wicher Bergsma, and I contributed 60% of these works. Chapter 4 was also jointly co-authored with Dr. Wicher Bergsma, in which I contributed 70% of this work, and we plan to submit this chapter for publication.

Acknowledgements

Wicher Bergsma, guidance, brilliance, hope to emulate, can always count on, friendship

Irini Moustaki second supervision, comforted, smile, guidance an honour and a pleasure social statistics group, i can always rely on administrators (Ian Marshall and Penny Montague)

second floor friends (Ragvir Sabharwal, Andy T. Y. Ho , Phoenix H. Feng, Tayfun Terzi, Alice Pignatelli-di-Cerchiara, Yan Qu, Xioalin Zhu), miss our time together, although i'm not always in the office!

I thank the Government of Brunei Darussalam and The London School of Economics and Political Science for funding and supporting this project.

To my parents whom this thesis is dedicated, and my family—your simultaneous concern and confidence in my ability to persevere is, by some measure a paradox, but more so a testament to your unconditional love and faith you put in me. For all you have given me, and continue to give, I would never be able to repay. To my beloved wife Ireve Ameera and darling daughter Naqiyyah Afrah who was born halfway through my PhD—you both make me whole.