Statistics > Methodology

Efficient Gaussian Process Regression for Large Data Sets

Anjishnu Banerjee, David Dunson, Surya Tokdar

(Submitted on 28 Jun 2011)

Gaussian processes (GPs) are widely used in nonparametric regression, classification and spatio-temporal modeling, motivated in part by a rich literature on theoretical properties. However, a well known drawback of GPs that limits their use is the expensive computation, typically O(n^3) in performing the necessary matrix inversions with n denoting the number of data points. In large data sets, data storage and processing also lead to computational bottlenecks and numerical stability of the estimates and predicted values degrades with n. To address these problems, a rich variety of methods have been proposed, with recent options including predictive processes in spatial data analysis and subset of regressors in machine learning. The underlying idea in these approaches is to use a subset of the data, leading to questions of sensitivity to the subset and limitations in estimating fine scale structure in regions that are not well covered by the subset. Motivated by the literature on compressive sensing, we propose an alternative random projection of all the data points onto a lower-dimensional subspace. We demonstrate the superiority of this approach from a theoretical perspective and through the use of simulated and real data examples. Some Keywords: Bayesian; Compressive Sensing; Dimension Reduction; Gaussian Processes; Random Projections; Subset Selection

Comments: 23 pages, 4 tables
Subjects: **Methodology (stat.ME)**
Cite as: **arXiv:1106.5779v1 [stat.ME]**

Submission history

From: Anjishnu Banerjee [view email]
[v1] Tue, 28 Jun 2011 19:55:22 GMT (24kb)

Which authors of this paper are endorsers?