Bayesian Nonparametric Regression with Local Models

Jo-Anne Ting[1], Stefan Schaal[1,2]

[1]Computer Science, University of Southern California, Los Angeles, CA 90089
[2]ATR Computational Neuroscience Laboratories, Kyoto 619-0288, Japan

We propose a Bayesian nonparametric regression algorithm with locally linear models for high-dimensional, data-rich scenarios where real-time, incremental learning is necessary. Nonlinear function approximation with high-dimensional input data is a nontrivial problem. For example, real-time learning of internal models for compliant control may be needed in a high-dimensional movement system like a humanoid robot. Fortunately, many real-world data sets tend to have locally low dimensional distributions, despite having high dimensional embedding [1, 2]. A successful algorithm, thus, must avoid numerical problems arising potentially from redundancy in the input data, eliminate irrelevant input dimensions, and be computationally efficient to allow for incremental, online learning.

Several methods have been proposed for nonlinear function approximation, such as Gaussian process regression [3], support vector regression [4] and variational Bayesian mixture models [5]. However, these global methods tend to be unsuitable for fast, incremental function approximation. Atkeson et al. [6] have shown that in such scenarios, learning with spatially localized models is more appropriate, particularly in the framework of locally weighted learning. In recent years, Vijayakumar et al. [7] have introduced a learning algorithm designed to fulfill the fast, incremental requirements of locally weighted learning, targeting high-dimensional input domains through the use of local projections. This algorithm, called Locally Weighted Projection Regression (LWPR), performs competitively in its generalization performance with state-of-the-art batch regression methods and has been applied successfully to sensorimotor learning on a humanoid robot.

The major issue with LWPR is that it requires gradient descent (with leave-one-out cross-validation) to optimize the local distance metrics in each local regression model. Since gradient descent search is sensitive to the initial values, we propose a novel Bayesian treatment of locally weighted regression with locally linear models [8] that eliminates the need for any manual tuning of meta parameters, cross-validation approaches or sampling. Combined with variational approximation methods to allow for fast, tractable inference, our algorithm learns the optimal distance metric value for each local regression model. It is able to automatically determine the size of the neighborhood data (i.e., the "bandwidth") that should contribute to each local model. A Bayesian approach offers error bounds on the distance metrics and incorporates this uncertainty in the predictive distributions. By being able to automatically detect relevant input dimensions, our algorithm is able to handle high-dimensional data sets with a large number of redundant and/or irrelevant input dimensions and a large number of data samples. We demonstrate competitive performance of our Bayesian locally weighted regression algorithm with Gaussian Process regression and LWPR on standard benchmark sets.

References

[1] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000.

[2] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, 2000.

[3] Christopher K. I. Williams and Carl Edward Rasmussen. Gaussian processes for regression. In David S. Touretzky, Michael C. Mozer, and Michael E. Hasselmo, editors, *In Advances in Neural Information Processing Systems 8*, volume 8. MIT Press, 1995.

[4] A. Smola and B. Schölkopf. From regularization operators to support vector kernels. In M. I. Jordan, M. J. Kearns, and S. A. Solla, editors, *Advances in Neural Information Processing Systems 10*, pages 343–349, Cambridge, MA, 1998. MIT Press.

[5] Z. Ghahramani and M.J. Beal. Graphical models and variational methods. In D. Saad and M. Opper, editors, *Advanced Mean Field Methods - Theory and Practice.* MIT Press, 2000.

[6] C. Atkeson, A. Moore, and S. Schaal. Locally weighted learning. *AI Review*, 11:11–73, April 1997.

[7] S. Vijayakumar, A. D'Souza, and S. Schaal. Incremental online learning in high dimensions. *Neural Computation*, pages 1–336, 2005.

[8] J. Ting, A. D'Souza, S. Vijayakumar, and S. Schaal. A Bayesian approach to empirical local linearization for robotics. Submitted for publication, 2007.