Structure Search in Gaussian Process Models

Anonymous Author 1
Unknown Institution 1

Anonymous Author 2
Unknown Institution 2

Anonymous Author 3
Unknown Institution 3

Abstract

Gaussian process (GP) models are used widely and successfully. However, their effectiveness depends critically on choosing an appropriate family of kernels. This aspect of GP modeling has been sorely underdeveloped. In this paper, we introduce a procedure for automatically and efficiently searching through a large space of GP models.

1 INTRODUCTION

Similar searches over large model classes have been succesfully used in machine vision [cite Cox + Pinto]. In general, learning the model class from data seems superior proposing the model beforehand. In high dimensional problems, it is also hard for a practitioner to propose an appropriate model even after examining a dataset closely. Choosing a kernel family is also a stumbling block for non-experts who wish to use Gaussian Process models.

2 GP STRUCTURE

[Standard GP intro Rasmussen & Williams (2006)]

3 MODEL-BASED SEARCH OVER MODELS

Bayesian optimization for hyper-parameter search: Snoek et al. (2012)

3.1 Bayesian Optimization

3.2 A Kernel between kernels

Hyperkernels Ong et al. (2002)

Preliminary work. Under review by AISTATS 2013. Do not distribute.

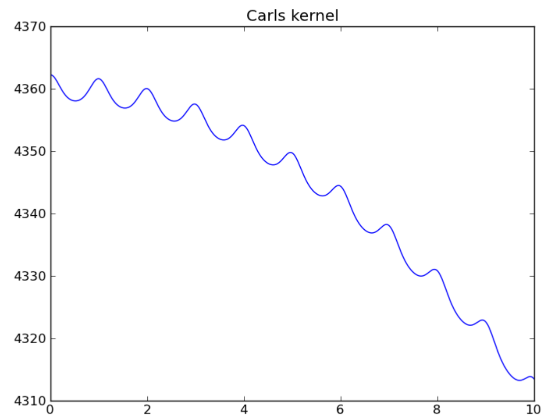

Figure 1: Carl's kernel function on the Mauna dataset.

4 EXPERIMENTS

4.1 Maunal Loa Atmoshperic Carbon Dioxide

As an example of a GP modeling problem where choosing an appropriate structure is critical, we revisit a dataset explored in Rasmussen & Williams (2006), pages 120-126, where a kernel was hand-tailored to fit a GP model to the dataset.

4.2 Load forecasting

4.3 Synthetic experiments

4.4 Real datasets

5 RELATED WORK

Compositional Model search for unsupervsied learning: Grosse et al. (2012)

Additive Gaussian Processes Duvenaud et al. (2011)

Hyperkernels Ong et al. (2002)

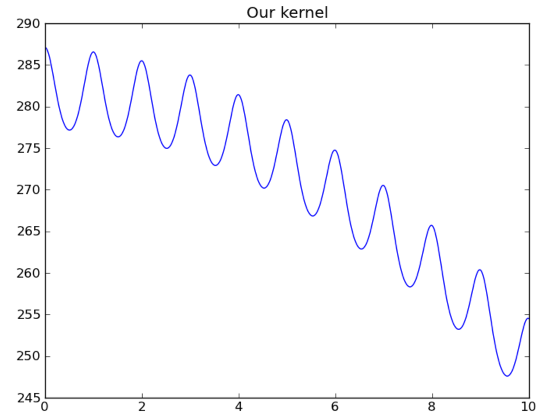

Figure 2: Our best kernel function on the Mauna dataset.

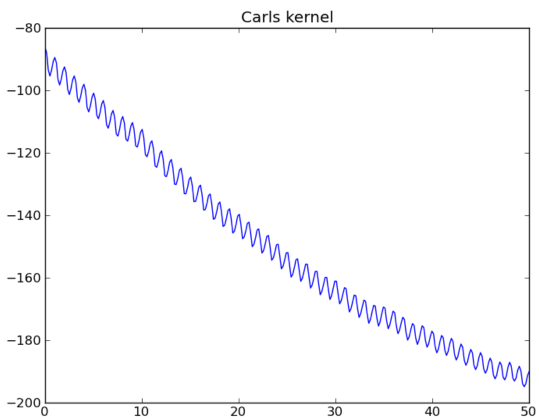

Figure 3: A draw from Carl's kernel.

5.1 Genetic Searches

Evolving kernel functions for SVMs by genetic programming: Diosan et al. (2007)

A Genetic Programming based kernel construction and optimization method for Relevance Vector Machines: Bing et al. (2010)

5.2 Equation Learning

Equation discovery with ecological applications Dzeroski et al. (1999)

6 DISCUSSION

Machine learning can be more data-driven, analogous to the high-thoughput approaches being used in biology.

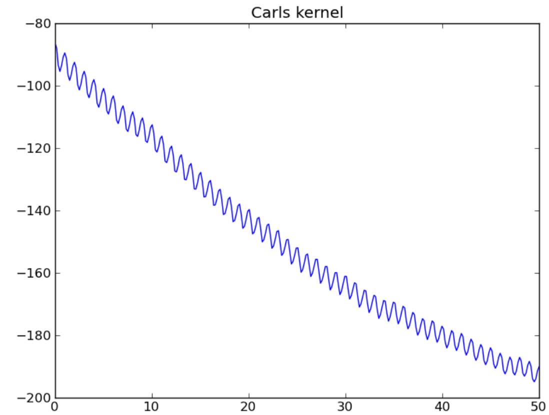

Figure 4: Another draw from Carl's kernel.

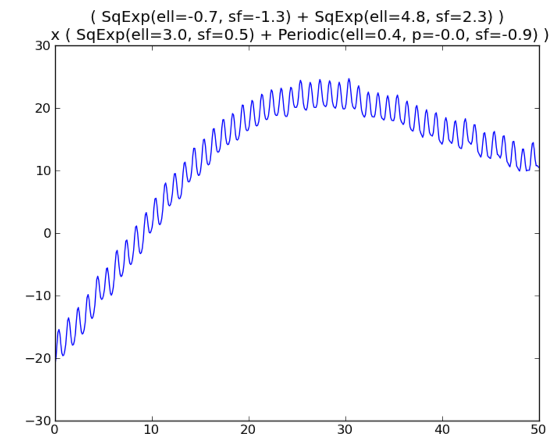

Figure 5: Another draw from our kernel.

7 CONCLUSION

Acknowledgements

References

Bing, W., Wen-qiong, Z., Ling, C., and Jia-hong, L. A gp-based kernel construction and optimization method for rvm. In *Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on*, volume 4, pp. 419–423. IEEE, 2010.

Diosan, L., Rogozan, A., and Pecuchet, J.P. Evolving kernel functions for svms by genetic programming. In *Machine Learning and Applications, 2007. ICMLA 2007. Sixth International Conference on*, pp. 19–24. IEEE, 2007.

Duvenaud, David, Nickisch, Hannes, and Rasmussen, Carl Edward. Additive Gaussian processes. In *Advances in Neural Information Processing Systems 25*, pp. 1–8, Granada, Spain, December 2011.

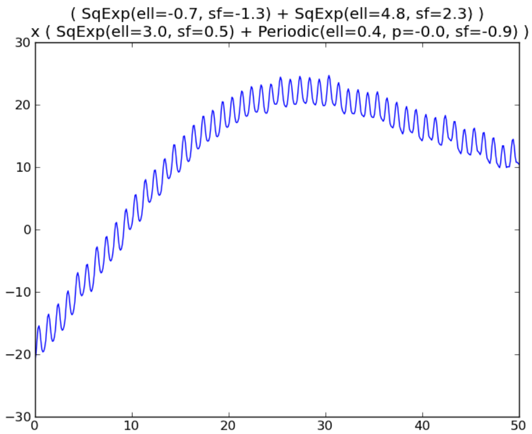

Figure 6: Another draw from our kernel.

Dzeroski, S., Todorovski, L., Bratko, I., Kompare, B., and Krizman, V. Equation discovery with ecological applications. *Machine learning methods for ecological applications*, pp. 185–207, 1999.

Grosse, R.B., Salakhutdinov, R., and Tenenbaum, J.B. Exploiting compositionality to explore a large space of model structures. In *Uncertainty in Artificial Intelligence*, 2012.

Ong, C.S., Smola, A.J., and Williamson, R.C. Hyperkernels. *Advances in neural information processing systems*, 15:478–485, 2002.

Rasmussen, C.E. and Williams, CKI. Gaussian Processes for Machine Learning. *The MIT Press, Cambridge, MA, USA*, 2006.

Snoek, J., Larochelle, H., and Adams, R.P. Practical bayesian optimization of machine learning algorithms. *arXiv preprint arXiv:1206.2944*, 2012.