Evolutionary Unsupervised Kernel Regression

Oliver Kramer oliver.kramer@tu-dortmund.de
Department of Computer Science, Technische Universität Dortmund, Germany

Dimension reduction and manifold learning play an important role in robotics, multimedia processing and data mining. For these tasks strong methods like Unsupervised Kernel Regression [4, 7] or Gaussian Process Latent Variable Models [5, 6] have been proposed in the last years. But many methods suffer from numerous local optima and crucial parameter dependencies. We use advanced methods from stochastic search to solve the embedded optimization problems of Unsupervised Kernel Regression. Furthermore, we apply a technique from Design of Experiments, i.e. Sequential Parameter Optimization, to tune the parameters and improve the algorithm's performance.

1 Stochastic Search in Unsupervised Kernel Regression

Machine learning comprises methods for classification, clustering, regression and dimension reduction, and has shown outstanding success in the last years. Nevertheless, many optimization problems in machine learning suffer from numerous local optima. First results have shown that machine learning optimization problems can successfully been solved by means of stochastic search [2, 8, 9, 11].

Unsupervised Kernel Regression is an approach for the learning of principal manifolds and has been introduced as an unsupervised counterpart of the Nadaraya-Watson kernel regression estimator [7, 4]. Various optimization problems are necessary for learning the Unsupervised Kernel Regression model:

- Search for optimal scale factors,
- candidate selection,
- the homotopy problem,
- CV-error minimization, and
- computation of the density threshold.

Because of the limited space for this abstract we refer to [7] for an introduction of the optimization problems and the kernel regression on latent variables. We make use of Covariance Matrix Adaptation techniques [10, 3] to efficiently solve these optimization tasks. The idea of Covariance Matrix Adaptation is to adapt the distribution of the mutation operator such that the probability to reproduce steps that led to the actual population increases. This idea is similar to estimation of distribution approaches. Covariance Matrix Adaptation algorithms approximate the inverse Hessian matrix and are the state-of-the-art optimizers in stochastic optimization. The advantage of stochastic optimization is robustness to local optima. Furthermore, evolutionary algorithms are embarrassingly parallelizable and thus fairly efficient search methodologies in distributed computing scenarios. Experimental analyses and statistical tests reveal that the evolutionary algorithm improves the results of the kernel regression method significantly.

2 Parameter Tuning of Unsupervised Kernel Regression with Sequential Parameter Optimization

Statistical tools like design of experiments support the parameter tuning process. An experimental design is the layout of a detailed experimental plan in advance of doing the experiment. Design of Experiments starts with the determination of the objectives of an experiment and the selection of the parameters (factors) for the study. The quality of the experiment (response) guides the search to find appropriate settings. In an experiment, we deliberately change one or more factors in order to observe the effect the changes have on one or more *response* variables. The response can be defined as the quality of the results, e.g. average fitness values at a given generation or convergence ratios. Bartz-Beielstein *et al.* [1] developed a parameter tuning method for stochastically disturbed algorithm output, the Sequential Parameter Optimization. It combines classical regression methods and statistical approaches for stochastic algorithms. We tune the parameters of Unsupervised Kernel Regression and the embedded stochastic optimization algorithms using Sequential Parameter Optimization. Our experimental analysis is focused on a set of known test problems. In our experiments we observe a significant win in performance and accuracy after the tuning process. The significance of the experimental results is evaluated using the non-parametric Wilcoxon test.

References

[1] T. Bartz-Beielstein, C. Lasarczyk, and M. Preuß. Sequential parameter optimization. In B. McKay et al., editors, *Proceedings of the IEEE Congress on Evolutionary Computation - CEC 2005*, volume 1, pages 773–780. IEEE Press, 2005.

[2] F. Gieseke and O. Kramer. Fast evolutionary maximum margin clutering. In *Submitted to International Conference on Machine Learning - ICML 2009*, 2009.

[3] N. Hansen. The CMA evolution strategy: A tutorial. Technical report, TU Berlin, ETH Zürich, 2005.

[4] S. Klanke and H. Ritter. Variants of unsupervised kernel regression: General cost functions. *Neurocomputing*, 70(7-9):1289–1303, 2007.

[5] N. Lawrence and A. Hyvärinen. Probabilistic non-linear principal component analysis with gaussian process latent variable models. *Journal of Machine Learning Research*, 6:1783–1816, 2005.

[6] N. D. Lawrence. Gaussian process latent variable models for visualisation of high dimensional data. In *Proceesdings of NIPS*, 2004.

[7] P. Meinicke, S. Klanke, R. Memisevic, and H. Ritter. Principal surfaces from unsupervised kernel regression. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(9):1379–1391, 2005.

[8] I. Mierswa. Evolutionary learning with kernels: A generic solution for large margin problems, 2006.

[9] O. Morell, T. Bernholt, R. Fried, J. Kunert, and R. Nunkesser. An evolutionary algorithm for lts-regression: A comparative study. 2008.

[10] A. Ostermeier, A. Gawelczyk, and N. Hansen. A derandomized approach to self adaptation of evolution strategies. *Evolutionary Computation*, 2(4):369–380, 1994.

[11] R. Stoean, D. Dumitrescu, M. Preuss, and C. Stoean. Evolutionary support vector regression machines. In *SYNASC*, pages 330–335, 2006.