Sparse Regression via the Winner-Take-All Networks

Nan Zhang, Shuqing Zeng and Juyang Weng

Abstract

Sparse representation is a desirable property for machine learning architectures, because it improves the generalization capability of the learning system. In supervised learning, the goal is to derive a mapping based on the training samples. The generalization capability is accomplished by reducing the complexity of the model, which is characterized by the number of non-zero parameters. This problem is formalized as l-0 norm minimization, which is called sparseness. It is known that l-0 norm minimization is NP-hard. Thus, it is usually approximated by assuming a super-Gaussian priori and applying a MAP (maximum a posteriori) procedure. If the Laplace priori is used, this problem is boiled down to LASSO regression, a minimization of residual error with an l-1 norm regularization term.

In this paper, we propose a new perspective to achieve sparseness via the winner-take-all principle for the linear kernel regression and classification task. We form the duality of the LASSO criteria, and transfer an l-1 norm minimization to an l-infinity norm maximization problem. Two solutions are proposed: it can be solved by standard quadratic programming with linear inequality constraints. We show that the number of parameters to be estimated in the l-infinity normed space is half of the parameters in the solution suggested by David Gay for l-1 normed space. Second, we introduce a novel winner-take-all neural network solution derived from gradient descending, which links the sparse representation and the competitive learning scheme. This scheme is a form of unsupervised learning in which each input pattern comes through learning, to be associated with the activity of one or at most a few neurons. However, the lateral interaction between neurons in the same layer is strictly preemptive in this model. This framework is applicable to a variety of problems, such as Independent Component Analysis (ICA), feature selection, and data clustering.