

CDGO 2007 The Second International Conference on Complementarity,  
Duality, and Global Optimization in Science and Engineering  
February 28 - March 2, 2007  
Gainesville, Florida

---

**Bayesian Ying-Yang Learning: A Statistical Theory for Machine  
Learning with Model Selection**

**Lei Xu**

Chinese Univ Hong Kong, IEEE Fellow,  
Fellow of International Association for Pattern Recognition,  
and Member of European Academy of Sciences  
Hong Kong  
lxu@cse.cuhk.edu.hk

Conventionally, the tasks of learning include a parameter learning part for determining unknown parameters and a model selection part for selecting an appropriate scale for a model that accommodates these parameters, implemented in a two-phase procedure with a vast computing cost. First, a number of models of a same architecture but in different scales are enumerated, with the unknown parameters estimated under the maximum likelihood (ML) principle. Second, one of typical learning theories, being different from a ML principle, is applied to select the best model. Alternatively, we introduce a statistical theory that integrates parameter learning, model selection, and regularization from a unified perspective under the name of Bayesian Ying-Yang (BYY) learning. The parameter learning part of BYY learning consists of determining not only unknown parameters but also an inner code of each sample, while the model selection part is searching a BYY system with both a scale for each inner code and a scale for accommodating unknown parameters. Both parameter learning and model selection are made under a same learning principle called best Ying-Yang harmony, implemented by the EM-like adaptive algorithms. The separation of the inner coding scale from the other scales leads to three favorable features. First, we get a criterion that outperforms typical model selection criteria in a two-phase implementation. Second, on several typical learning tasks, model selection can be achieved automatically during parameter learning and save the computing cost significantly. Third, it provides a framework to integrate the roles of regularization and model selection, and mutually one can compensate certain weakness caused by the other. Experiments on several typical learning tasks are given to demonstrate these features.