2nd International Symposium on
Information Geometry and its Applications
December 12-16, 2005, Tokyo

INTEGRABILITY OF WEAK LEARNER ON BOOSTING

TAKAFUMI KANAMORI

Boosting algorithms for multi-class classification problems are studied. Mainly we are concerned with the relation between statistical models and weak learners. Boosting is a common learning method that can find a highly accurate classifier by using weak learner iteratively, where weak learner denotes a simple learning algorithm such as decision-stumps.

From the viewpoint of information geometry, estimation by boosting is interpreted as moment matching estimation in positive measure space, and weak learner corresponds to a vector field on statistical models. When the vector fields defined by weak learner is integrable, boosting algorithm is identical to the coordinate descent method for some loss function, and then, convergence property of boosting is assured. In general, the vector field defined by weak learner is not necessarily integrable on statistical models. In this talk, we study statistical models on which the vector field is integrable. In binary classification problems, integrability is satisfied on wide range of statistical models. This is a reason why there is a high degree of freedom for the design of loss functions for binary classification problems. On the other hand, when the number of labels is more than two, statistical models with integrable vector field are quite restricted. If weak learner has some symmetry property, these statistical models have one-dimensional parameter which is useful to deal with noisy data.

A typical way to derive boosting algorithms is to apply the coordinate descent methods for some loss function. However, underlying statistical models are not necessarily known explicitly. On the other hand, integrability condition provides explicit representation of statistical models on which moment matching estimation is consistent. Thus, our approach gives a simple way to estimate conditional probability of labels. This is a practical advantage of our methods over the other learning methods for multi-class problems.