Penalized Logistic Regression Machines: New methods for statistical prediction

Kunio Tanabe

The Institute of Statistical Mathematics
4–6–7 Minamiazabu, Minatoku, Tokyo, 106–0047, Japan.

Abstract

Support Vector Machines(Vapnik,1979,'95,'98) have been recognized as powerful method for learning certain structures from data for prediction. Their success is due to intrinsic combination of the quadratic programming models with *Kernel Method*. The machine, however, does not seem to accomodate very well the cases where the mechanism of generating data is largely of stochastic nature. By employing the penalized logistic regression model, we make a statistical attempt to construct multiclass discrimination machines which can handle much noisier stochastic data to be competetive with SVM in such an environment. It is shown that *by penalizing the likelihood in a specific way, we can intrinsically combine the logistic regression model with the kernel methods.* In particular, *a new class of penalty functions and associated normalized projective kernels are introduced to gain a versatile induction power of our learning machines.* The closed formulas are given for the first and second derivetives of the log penalized logistic regression likelihood, whose Hessian matrix is shown to be positive definite and uniformly bounded. *Dual classes of globally convergent* learning machines(algorithms) are given for obtaining the optimal parameters for both probabilistic and deterministic prediction. Analysis of the rate of convergence is given for each class of machines. The type-II(or marginal) likelihood and Generalized Information Criteria are also given in closed form for determining the optimal value of hyperparameters in the model so that the machines have a *due induction capacity to the size and the quality of an available training data set.*

Key words: Prediction, Multiclass Discrimination, Penalized Logistic Regression, Neural Network, Kernel Method, Normalized Projective Kernel, Dual Learning Machines, Induction Capacity, Type-II Likelihood, Marginal Likelihood, Generalized Information Criterion