

ADAPTIVE REGRESSION BY MIXING

by

Yuhong Yang
Iowa State University

ABSTRACT

Adaptation over different procedures is of practical importance. Different procedures perform well under different conditions. In many practical situations, it is rather hard to assess which conditions are (approximately) satisfied so as to identify the best procedure for the data at hand. Thus automatic adaptation over various scenarios is desirable.

A practically feasible method, named Adaptive Regression by Mixing (ARM) is proposed to convexly combine general candidate regression procedures. Under mild conditions, the resulting estimator is theoretically shown to perform optimally in rates of convergence without knowing which of the original procedures work the best.

Simulations are conducted in several settings, including comparing a parametric model with nonparametric alternatives, comparing a neural network with a projection pursuit in multi-dimensional regression, and combining bandwidths in kernel regression. The results clearly support the theoretical property of ARM.

The ARM algorithm assigns weights on the candidate models/procedures via proper assessment of performance of the estimators. The data is split into two parts, one for estimation and the other for measuring behavior in prediction. While there are many plausible ways to assign the weights, ARM has a connection with information theory, which ensures the desired adaptation capability. Indeed, under mild conditions, we show that the squared L_2 risk of the estimator based on ARM is basically bounded above by the risk of each candidate procedure plus a small penalty term of order $1/n$. Minimizing over the procedures gives the automatically optimal rate of convergence for ARM.

Model selection often induces unnecessarily large variability in estimation. Alternatively, a proper weighting of the candidate models can be more stable, resulting in a smaller risk. Simulations suggest that ARM works better than model selection using AIC or BIC when the error variance is not very small.

Keywords: Adaptive estimation, combining procedures, nonparametric regression.