On the selection of the order of a polynomial model

CHRIS S. WALLACE †  Monash University, AU

A recent report by Cherkassky, Mulier and Vapnik has compared the performance of several methods for selecting the order of a polynomial approximation to a function \( t(x) \) given only the values of \( t(x) \) for some set of \( x \)-values, where the given values are corrupted by i.i.d. Gaussian noise of unknown variance. They compare various "classical" methods with a new method based on the concept of Vapnik-Chervonenkis (VC) dimension (Vapnik, 1995, personal communication) and conclude that the latter gives the most reliable prediction of the value of \( t(x) \) at unseen values of \( x \). The current work replicates their investigation and extends it by including a method based on minimum message length (MML) estimation (Wallace and Freeman, 1987, Journal of the Royal Statistical Society Series B 49, 3, 223-265; Wallace and Boulton, 1968, Computation Journal 11, 184-94). The results largely confirm the previous results, but show that MML is generally superior to the VC method in terms of average squared prediction error.

† POSTAL: School of Computer Science and Software Engineering, Monash University, Clayton VIC 3168, AUSTRALIA.
FAX: —
EMAIL: csw@cs.monash.edu.au
WWW: ---