The choice of smoothing parameter in kernel regression smoothing

Orathai Polsen* & Charles C. Taylor

Department of Statistics, University of Leeds

We consider the kernel estimator, called the Nadaraya-Watson (N-W), as an estimate of a regression curve. This estimator was first introduced by Nadaraya (1964) and Watson (1964) and the accuracy of this estimator depends mainly on a smoothing parameter h. Therefore, it is necessary to choose a value for the smoothing parameter. Cross-validation (CV) is a basic technique for choosing the smoothing parameter. In this paper, the two alternative theoretical smoothing choices are proposed in an attempt to estimate regression curve; in one case h is fixed, and in the other depends on x. Specifically, a plug-in rule is obtained when data has a bivariate normal distribution. Some simulations and examples were used in order to demonstrate the comparison between these two theoretical smoothing parameters and the one from Cross-validation. The theoretical smoothing parameters perform better than the one from Cross-validation in most settings.

Keywords: The Nadaraya-Watson estimator; Smoothing parameter; Cross-validation.