Kernel bandwidth estimation for non-parametric density estimation: a comparative study

Christiaan M. van der Walt
Modelling and Digital Science, CSIR, Pretoria 0001
Multilingual Speech Technologies Group,
North-West University,
Vanderbijlpark 1900, South Africa
cvdwalt@csir.co.za

Etienne Barnard
Multilingual Speech Technologies Group
North-West University
Vanderbijlpark 1900, South Africa
etienne.barnard@nwu.ac.za

Abstract

We investigate the performance of conventional bandwidth estimators for non-parametric kernel density estimation on a number of representative pattern-recognition tasks, to gain a better understanding of the behaviour of these estimators in high-dimensional spaces. We show that there are several regularities in the relative performance of conventional kernel bandwidth estimators across different tasks and dimensionalities. In particular, we find that the Silverman rule-of-thumb and maximal-smoothing principle estimators consistently perform competitively on most tasks and dimensions for the datasets considered.