

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
UNIVERSITY OF SOUTHERN DENMARK, ODENSE

COMPUTER SCIENCE COLLOQUIUM

Local Intrinsic Dimensionality II: Multivariate Analysis and Distributional Support

Michael E. Houle

National Institute of Informatics, Japan

Tuesday, 16 July, 2019 at 14:15

IMADA's Seminar Room

Abstract:

Researchers have long considered the analysis of similarity applications in terms of the intrinsic dimensionality (ID) of the data. This presentation is concerned with a generalization of a discrete measure of ID, the expansion dimension, to the case of smooth functions in general, and distance distributions in particular. A local model of the ID of smooth functions is first proposed and then explained within the well-established statistical framework of extreme value theory (EVT). Moreover, it is shown that under appropriate smoothness conditions, the cumulative distribution function of a distance distribution can be completely characterized by an equivalent notion of data discriminability. As the local ID model makes no assumptions on the nature of the function (or distribution) other than continuous differentiability, its generality makes it ideally suited for the learning tasks that often arise in data mining, machine learning, and other AI applications that depend on the interplay of similarity measures and feature representations. An extension of the local ID model to a multivariate form will also be presented, that can account for the contributions of different distributional components towards the intrinsic dimensionality of the entire feature set, or equivalently towards the discriminability of distance measures defined in terms of these feature combinations. The talk will conclude with a discussion of recent applications of local ID to deep learning.

Host: Arthur Zimek