Robust Space Transformations for Distance-based Outliers

Raymond Ng[1]

[1] Department of Computer Science, University of British Columbia, Canada

Abstract

In the first part of this talk, we will present the notion of distance-based outliers. This is a non-parametric approach, and is particularly suitable for high dimensional data. We will show a case study based on video trajectory surveillance.

For distance-based outlier detection, there is an underlying multi-dimensional data space in which each tuple/object is represented as a point in the space. We observe that in the presence of variability, correlation, outliers and/or differing scales, we may get unintuitive results if an inappropriate space is used. The fundamental question addressed in the second half of this talk is: "What then is an appropriate space?". We propose using a robust space transformation called the Donoho-Stahel estimator. We will focus on the computation of the transformation.