

From Kernels to Mixtures to Clusters

William F. Szewczyk (National Security Agency)

`wfszewc@afterlife.ncsc.mil`

David W. Scott

Abstract

Most clustering algorithms are based on the heuristic: lots of points close together are more interesting than points far apart. Different algorithms will define “close together” differently. One reasonable way to find points close together is to search for regions of high density and define those as clusters. In this talk we will show how starting with a non-parametric density estimate one is able to identify clusters by a simple recursive simplification of the estimate.