

***k*-Nearest Neighbor and Kernel-Based Credal Classifiers: Theory and Development**

Mark Ducey (Department of Natural Resources, University of New Hampshire)
mjducey@cisunix.unh.edu

Abstract

Credal classifiers extend probabilistic classification schemes by employing *credal sets*, or sets of probability measures, to describe imprecision in classification. Credal classifiers have previously been developed as classification trees using discrete or discretized attributes. Here, I extend this family of techniques to continuous classification variables using *k*-nearest neighbor, kernel-based, and weighted *k*-nearest neighbor methods. These methods can tolerate imprecision in both training and validation data, and return classification results having logically defensible precision. A test example, employing hemispherical photography for the description of forest architecture, shows results competitive with a traditional voting *k*-nearest neighbor kernel classifier when the training and validation data are precise, and returns useful results even when the classes in the training set are imprecise. Parameter selection rules based on the Total Uncertainty Measure (TUM) suggest the need for post-classification analysis may be reduced with credal classifiers, but in our examples the TUM indicates parameter values that are too conservative. However, further research is needed on accuracy measures, representation invariance properties, and parameter selection for the IDM itself. Additional pragmatic testing is also required to establish whether, and under what circumstances, credal classifiers offer improvements over more traditional methods.