

CCSU  
DEPARTMENT OF MATHEMATICAL SCIENCES  
DATA MINING THESIS PRESENTATION

Friday, March 6  
3:00 – 4:00 pm in MS 101

**Extending the Data Mining  
Software Packages  
SAS Enterprise Miner and  
SPSS Clementine to  
Handle Fuzzy Cluster Membership:  
Implementation with Examples**

Don Wedding

**SAS Institute**

(MS candidate in Data Mining)

**Abstract**

*Hard clustering* and *fuzzy clustering* are both commonly used by data miners for predictive modeling and explaining data and results. In hard clustering, membership is absolute. A record is completely in a cluster or it is completely out of a cluster so that membership is mutually exclusive. In fuzzy clustering, a record is permitted to have partial membership. For example, a record might be 60% in cluster 1 and 40% in cluster 2. Although both types of clustering are used in research, only hard clustering is provided by the two most popular commercially available data mining development tools: SAS Enterprise Miner and SPSS Clementine.

This research presents a method of approximating fuzzy clusters in both SAS Enterprise Miner and SPSS Clementine by extending their hard clustering tools to incorporate fuzzy membership values. Although these fuzzy memberships of hard clusters are only approximations of fuzzy clusters, it is shown that they can improve accuracy over hard clustering in some situations.

***For further information:***

[gotchevi@ccsu.edu](mailto:gotchevi@ccsu.edu) (860) 832-2839