

Generalizations of Ward's method, or, norms and netball - what's the connection?

Alan Lee and Bobby Willcox

University of Auckland, email: lee@stat.auckland.ac.nz

Keywords: distance matrix, Minkowski distance, Ward's method.

In this talk, we consider several generalizations of the popular Ward's method for agglomerative hierarchical clustering. Our work was motivated by clustering software, such as the R function `hclust`, which accepts a distance matrix as input and applies Ward's definition of inter-cluster distance to produce a clustering. The standard version of Ward's method uses squared Euclidean distance to form the distance matrix. We explore the properties and effect on the clustering of using other definitions of distance, such as the Minkowski distance and powers of the Minkowski distance. We explore the effect of these on several examples and find that using powers of the Manhattan metric is particularly effective.