

Biweight Correlation as a Measure of Distance between Genes on a Microarray

Abstract: The underlying goal of microarray experiments is to identify genetic patterns across different experimental conditions. Genes contained in a particular pathway or that respond similarly to experimental conditions should be coregulated and show similar patterns of expression on a microarray. Using any of a variety of clustering methods or gene network analyses, we can partition genes of interest into groups, clusters, or modules based on measures of similarity. Typically, Pearson correlation is used to measure distance (or similarity) before implementing a clustering algorithm. Pearson correlation is susceptible to outliers, however, an unfortunate characteristic when dealing with microarray data (well-known to be typically noisy.) We propose a robust similarity metric based on Tukey's biweight estimate of multivariate scale and location.