

index.H(clusterSim)

### **Hartigan index**

(for metric data only)

$$H(u) = \left( \frac{\text{tr}\mathbf{W}_u}{\text{tr}\mathbf{W}_{u+1}} - 1 \right) (n - u - 1),$$

where:  $\mathbf{W}$  – within-group dispersion matrix,

$u$  – number of clusters ( $u = 1, \dots, n - 2$ ),

$n$  – number of objects.

The estimated number of clusters is the smallest  $u \geq 1$  such that  $H(u) \leq 10$ .

### **References**

Hartigan, J. (1975), *Clustering algorithms*, Wiley, New York.

Milligan, G.W., Cooper, M.C. (1985), *An examination of procedures of determining the number of cluster in a data set*, "Psychometrika", vol. 50, no. 2, 159-179.

Tibshirani R., Walther G., Hastie T. (2001), *Estimating the number of clusters in a data set via the gap statistic*, „Journal of the Royal Statistical Society”, ser. B, vol. 63, part 2, 411-423.