index.KL(clusterSim)

**Krzanowski and Lai index**

(for metric data only)

$$KL(u) = \left| \frac{DIFF_u}{DIFF_{u+1}} \right|,$$

$$DIFF_u = (u-1)^{2/m} tr\mathbf{W}_{u-1} - u^{2/m} tr\mathbf{W}_u,$$

where: $\mathbf{W}$ – within-group dispersion matrix,
$u$ – number of clusters ($u = 2,\ldots,n-2$),
$n$ – number of objects,
$m$ – number of variables.

The value of $u$, which maximizes $KL(u)$, is regarded as specifying the number of clusters.

**References**

Krzanowski , W.J., Lai, Y.T. (1988), *A criterion for determining the number of groups in a data set using sum of squares clustering*, "Biometrics", 44, no. 1, 23-34.

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.