

Ward's method

Contributors to Wikimedia projects : 5-6 minutes : 3/10/2011

DOI: [10.1007/s00357-015-9167-1](https://doi.org/10.1007/s00357-015-9167-1), [Show Details](#)

From Wikipedia, the free encyclopedia

In [statistics](#), **Ward's method** is a criterion applied in [hierarchical cluster analysis](#). **Ward's minimum variance method** is a special case of the [objective function](#) approach originally presented by Joe H. Ward, Jr.^[1] Ward suggested a general [agglomerative hierarchical clustering](#) procedure, where the criterion for choosing the pair of clusters to merge at each step is based on the optimal value of an objective function. This objective function could be "any function that reflects the investigator's purpose." Many of the standard clustering procedures are contained in this very general class. To illustrate the procedure, Ward used the example where the objective function is the [error sum of squares](#), and this example is known as *Ward's method* or more precisely *Ward's minimum variance method*.

The [nearest-neighbor chain algorithm](#) can be used to find the same clustering defined by Ward's method, in time proportional to the size of the input [distance matrix](#) and space linear in the number of points being clustered.

The minimum variance criterion[[edit](#)]

Ward's minimum variance criterion minimizes the total within-cluster variance. To implement this method, at each step find the pair of clusters that leads to minimum increase in total within-cluster variance after merging. This increase is a weighted squared distance between cluster centers. At the initial step, all clusters are singletons (clusters containing a single point). To apply a [recursive algorithm](#) under this [objective function](#), the initial distance between individual objects must be (proportional to) squared [Euclidean distance](#).

The initial cluster distances in Ward's minimum variance method are therefore defined to be the squared Euclidean distance between points:

$$d_{ij} = d(\{X_i\}, \{X_j\}) = \|X_i - X_j\|^2.$$

Note: In software that implements Ward's method, it is important to check whether the function arguments should specify Euclidean distances or squared Euclidean distances.

Lance–Williams algorithms[\[edit\]](#)

Ward's minimum variance method can be defined and implemented recursively by a Lance–Williams algorithm. The Lance–Williams algorithms are an infinite family of agglomerative hierarchical clustering algorithms which are represented by a recursive formula for updating cluster distances at each step (each time a pair of clusters is merged). At each step, it is necessary to optimize the objective function (find the optimal pair of clusters to merge). The recursive formula simplifies finding the optimal pair.

Suppose that clusters C_i and C_j were next to be merged. At this point all of the current pairwise cluster distances are known. The recursive formula gives the updated cluster distances following the pending merge of clusters C_i and C_j . Let

An algorithm belongs to the Lance-Williams family if the updated cluster distance $d_{(ij)k}$ can be computed recursively by

$$d_{(ij)k} = \alpha_i d_{ik} + \alpha_j d_{jk} + \beta d_{ij} + \gamma |d_{ik} - d_{jk}|,$$

where $\alpha_i, \alpha_j, \beta$, and γ are parameters, which may depend on cluster sizes, that together with the cluster distance function d_{ij} determine the clustering algorithm. Several standard clustering algorithms such as [single linkage](#), [complete linkage](#), and group average method have a recursive formula of the above type. A table of parameters for standard methods is given by several authors. [\[2\]\[3\]\[4\]](#)

Ward's minimum variance method can be implemented by the Lance-Williams formula. For disjoint clusters C_i, C_j , and C_k with sizes n_i, n_j , and n_k respectively:

$$d(C_i \cup C_j, C_k) = \frac{n_i + n_k}{n_i + n_j + n_k} d(C_i, C_k) + \frac{n_j + n_k}{n_i + n_j + n_k} d(C_j, C_k) - \frac{n_k}{n_i + n_j + n_k} d(C_i, C_j).$$

Hence Ward's method can be implemented as a Lance-Williams algorithm with

$$\alpha_i = \frac{n_i + n_k}{n_i + n_j + n_k}, \quad \alpha_j = \frac{n_j + n_k}{n_i + n_j + n_k}, \quad \beta = \frac{-n_k}{n_i + n_j + n_k}, \quad \gamma = 0.$$

Variations[\[edit\]](#)

The popularity of the Ward's method has led to variations of it. For instance, Ward_p introduces the use of cluster specific feature weights, following the intuitive idea that features could have different degrees of relevance at different clusters. [\[5\]](#)

References[\[edit\]](#)

- [^] Ward, J. H., Jr. (1963), "Hierarchical Grouping to Optimize an Objective Function", *Journal of the American Statistical Association*, 58, 236–244.
- [^] Cormack, R. M. (1971), "A Review of Classification", *Journal of the Royal Statistical Society*, Series A, 134(3), 321-367.
- [^] Gordon, A. D. (1999), *Classification, 2nd Edition*, Chapman and Hall, Boca Raton.
- [^] Milligan, G. W. (1979), "Ultrametric Hierarchical Clustering Algorithms", *Psychometrika*, 44(3), 343–346.
- [^] R.C. de Amorim (2015). "[Feature Relevance in Ward's Hierarchical Clustering Using the Lp Norm](#)" (PDF). *Journal of Classification*. **32** (1): 46–62. [doi:10.1007/s00357-015-9167-1](#). [S2CID 18099326](#).

Further reading[\[edit\]](#)

- Everitt, B. S., Landau, S. and Leese, M. (2001), *Cluster Analysis, 4th Edition*, Oxford University Press, Inc., New York; Arnold, London. [ISBN 0340761199](#)
- Hartigan, J. A. (1975), *Clustering Algorithms*, New York: Wiley.
- [Jain, A. K.](#) and Dubes, R. C. (1988), *Algorithms for Clustering Data*, New Jersey: Prentice–Hall.

- Kaufman, L. and Rousseeuw, P. J. (1990), *Finding Groups in Data: An Introduction to Cluster Analysis*, New York: Wiley.