In [1]:
```
## settings ↔
```

\usepackage*amssymb*

4.3 Hierarchic Clustering

Hierarchic Clustering leads to a treelike division of clusters. There are two opposite procedures:

- divisive clustering: start with a single cluster that contains all data, stepwise division into smaller clusters (top-down)
- agglomerative clustering: initialize all data points as 1-element clusters, stepwise merging of clusters (bottom-up)

4.3.1 Agglomeratives Clustering

1. Initialization: Regard data points as 1-element cluster $C_1, \ldots, C_n$.
2. Find pair of clusters C_i, C_j with
$$(i^\star, j^\star) = \arg\min_{i<j} d(C_i, C_j)$$
Set $\tilde{d} = d(C_{i^\star}, C_{j^\star})$
3. Replace $C_{i^\star} \leftarrow C_{i^\star} \cup C_{j^\star}$
4. If $j^\star < n$: replace $C_j^\star \leftarrow C_n$
5. Set $n \leftarrow n - 1$
6. If termination condition A($C_1, \ldots, C_n$) =false: ~~goto 2

termination conditions:

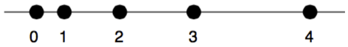
- number of clusters: $n \leq n_{\text{goal}}$
- error: $\sum_{i=1}^{n} \text{Var}(C_i) > E_{goal}$
- distance: & $\tilde{d} > d_{max}$

Different clustering variants result according to the choice of the distance function in step 2

1. **Single Linkage Clustering (SLC)**
$$d(C_i, C_j) = d_1(C_i, C_j)$$
   - SLC is inclined to form 'strings of clusters'
2. **Complete Linkage Clustering (CLC)**
$$d(C_i, C_j) = d_2(C_i, C_j)$$
   - CLC has the tendency to result in compact sphere-shaped clusters

Example

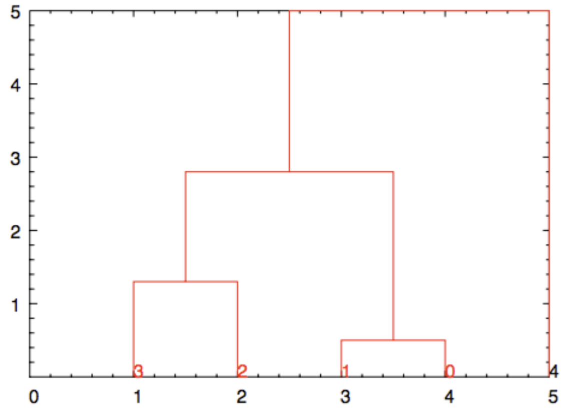
Consider the data set: $D = \{0, 0.5, 1.5, 2.8, 5\}$



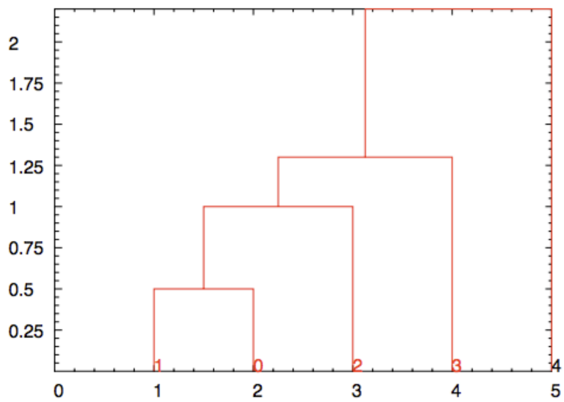
Clustering yields the following dendrograms:

- Note that the x-coordinates of the plots do not match to the data point indices. sorting is always done so in dendrograms that neighbored items can be merged.

CLC:



SLC:



3. Average linkage clustering

$$d(C_i, C_j) = d_3(C_i, C_j)$$

- well balanced, between CLC and SLC

4. Centroid Linkage Clustering

$$d(C_i, C_j) = d_4(C_i, C_j)$$

- Each cluster is represented by the mean of its center vectors
- The computation of means requires real-valued variables
- Attention: When merging two clusters, the center-of-mass of the resulting cluster is dominated by the bigger cluster

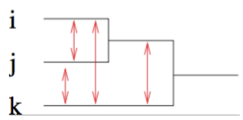
5. Ward's Linkage Clustering

Here, an optimality criterion is used:

- with each step that pair of clusters C_i, C_j is merged that increases the mean standard deviation
$$E = \frac{1}{N} \sum_i \sum_{\vec{x} \in C_i} (\vec{x} - \hat{\mu}_i)^2$$
of data around the cluster centers $\hat{\mu}_i = \langle \vec{x} \rangle_{C_i}$ the least.
- this favors the formation of spherical clusters.

4.3.2. Recursive Distance Measures

All methods described before can be derived from a recursively defined distance measure:
$$d_{k,(i,j)} = \alpha_i d_{ki} + \alpha_j d_{kj} + \beta d_{ij} + \gamma |d_{ki} - d_{kj}|$$



- Note that this Figure is a 90 degree rotated dendrogram, but the cluster distances $d(j,k) = d_{jk}$ and $d(i,k) = d_{ik}$ are not depicted. The red arrows highlight what distances are meant.

According to the choice of parameters $\alpha_i, \alpha_j, \beta$ and γ we receive the above methods 1.-5., and further methods.

α_i	α_j	β	γ	Method
1/2	1/2	0	$-1/2$	Single Linkage
1/2	1/2	0	1/2	Complete Linkage
$\frac{n_i}{n_i+n_j}$	$\frac{n_j}{n_i+n_j}$	0	0	Centroid linkage Clustering
$\frac{n_i}{n_i+n_j}$	$\frac{n_j}{n_i+n_j}$	$-\alpha_i\alpha_j$	0	Average linkage Clustering
$\frac{n_i+n_j}{n_i+n_j+n_k}$	$\frac{n_j+n_k}{n_i+n_j+n_k}$	$-\frac{n_k}{n_i+n_j+n_k}$	0	Ward's linkage Clustering