Distance Methods

Bret Larget

Departments of Botany and of Statistics
University of Wisconsin—Madison

September 22, 2009

UPGMA

- UPGMA is an acronym for Unweighted Pair-Group Method with Arithmetic Mean.
- UPGMA produces an ultrametric tree from a symmetric distance matrix.
- The depth of each node is the average of all of the pairwise distances between joined subtrees from the original distance matrix.
- The algorithm joins the pair with the smallest distance and then recomputes the distance from the new group to others.
- Continue until there is only one group.

UPGMA Algorithm

1. Find the i and j with the smallest distance D_{ij}.
2. Create a new group (ij) which has $n_{(ij)} = n_i + n_j$ members.
3. Connect i and j on the tree to a new node (ij). Give the edges connecting i to (ij) and j to (ij) each length so that the depth of group (ij) is $D_{ij}/2$.
4. Compute the distance between the new group and all other groups except i and j by using

$$D_{(ij),k} = \left(\frac{n_i}{n_i + n_j}\right) D_{ik} + \left(\frac{n_j}{n_i + n_j}\right) D_{jk}$$

5. Delete columns and rows corresponding to i and j and add one for (ij). If there are two or more groups left, go back to the first step.

Example

	Dog	Bear	Raccoon	Weasel
Dog	0	32	48	52
Bear	32	0	26	34
Raccoon	48	26	0	42
Weasel	52	34	42	0

Join Bear and Raccoon, depth is $26/2 = 13$.

Example

	Dog	B/R	Weasel
Dog	0	40	52
B/R	40	0	38
Weasel	52	38	0

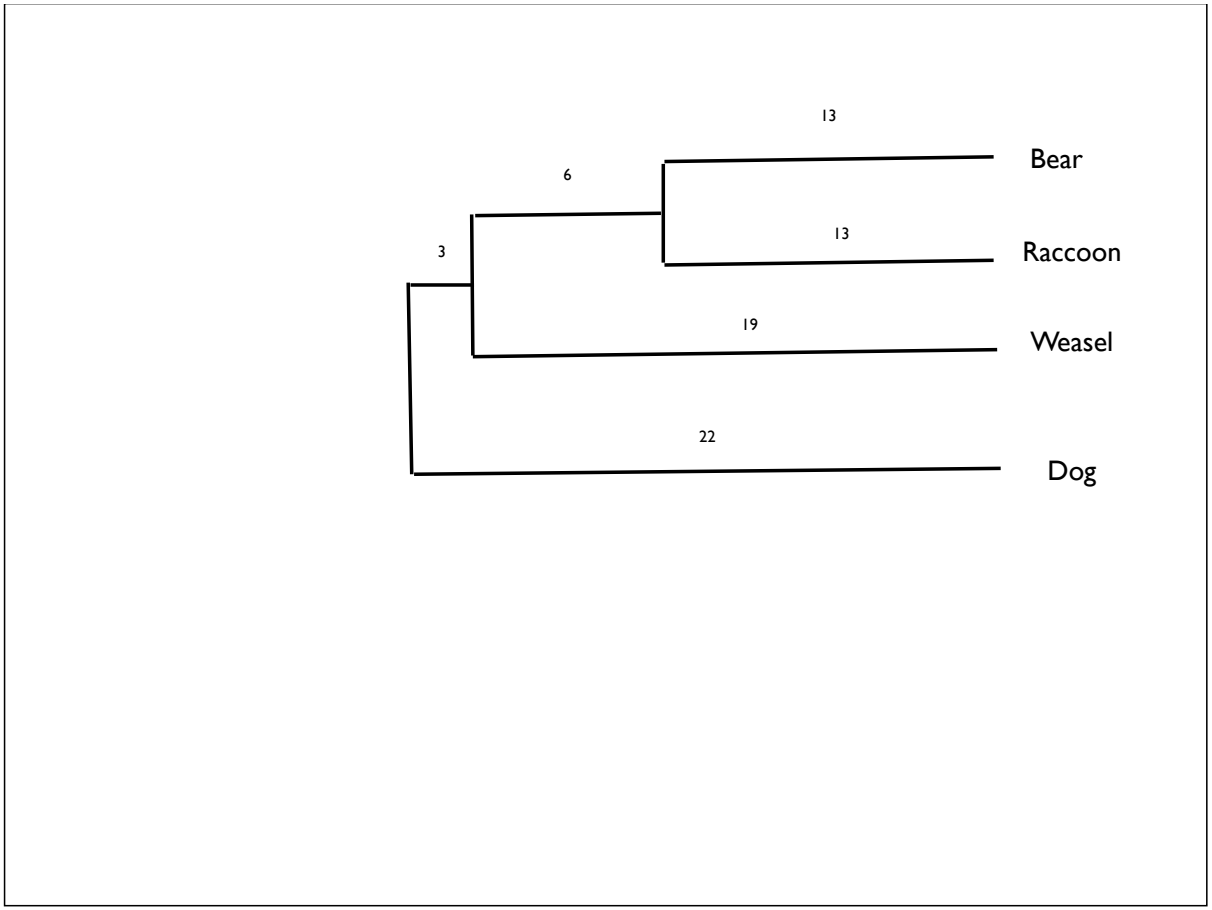
Join B/R and Weasel, depth is $38/2 = 19$.

Example

	Dog	B/R/W
Dog	0	44
B/R/W	44	0

Join Dog and B/R/W, depth is $44/2 = 22$.

UPGMA Tree



Neighbor-joining

- Neighbor-Joining creates an unrooted tree which will be exact if original distance matrix matches an additive tree.
- Choice of the selected pair to join depends on adjusting distances to account for possible unequal rates.
- The adjustments result in negative "distances", and the smallest of these is selected.
- Once there are three remaining groups, the same tree results regardless which pair is selected to join next.
- UPGMA and Neighbor-joining can lead to different tree topologies.

Neighbor-joining Algorithm

1. For each leaf, compute $u_i = \sum_{j \neq i} D_{ij}/(n-2)$.
2. Choose the i and j for which $D_{ij} - u_i - u_j$ is smallest.
3. Join i and j to a new node with lengths $(D_{ij} + u_i - u_j)/2$ to node i and $(D_{ij} + u_j - u_i)/2$ to node j.
4. Compute the distance to the new node (ij) and the other groups as

$$D_{(ij),k} = \frac{D_{ik} + D_{jk} - D_{ij}}{2}$$

5. Delete columns and rows corresponding to i and j and add one for (ij). If there are three or more groups left, go back to the first step. Otherwise, connect the two remaining nodes with their distance.

Example

	D	B	R	W	u_i
Dog	0	32	48	52	66
Bear	32	0	26	34	46
Raccoon	48	26	0	42	58
Weasel	52	34	42	0	64
u_j	66	46	58	64	

	D	B	R	W
Dog	—	−80	−76	−78
Bear	−80	—	−78	−76
Raccoon	−76	−78	—	−80
Weasel	−78	−76	−80	—

- Can choose to join either D/B or R/W because of tie.
- New edge to dog has length $(32 + 66 - 46)/2 = 26$.
- New edge to bear has length $(32 + 46 - 66)/2 = 6$.
- Note these edges sum to 32, but are not equal.

Example

	D/B	R	W	u_i
D/B	0	21	27	48
Raccoon	21	0	42	63
Weasel	27	42	0	69
u_j	48	63	69	

	D/B	R	W
D/B	—	−90	−90
Raccoon	−90	—	−90
Weasel	−90	−90	—

- For the last three, you can always join any pair.
- Simply use the equation from step 4 of the algorithm for the distances.
- Note that in computing u_i, we now use $n = 3$ as there are n groups now.

Neighbor-Joining Tree