To classify 5 students into clusters according to only two variables: x_1 expenditures on food for the student, and x_2 expenditures on communications by the student, according to the data shown in the following table:

Code	x_1	x_2
1 = *a*	2	4
2 = b	8	2
3 = c	9	3
4 = d	1	5
5 = e	8.5	1

Let us first assume that each of the five students alone forms a special cluster, then we calculate the Euclidean distances between each pair according to the relationship:

$$d_{jk} = \sqrt{(x_{1j} - x_{1k})^2 + (x_{2j} - x_{2k})^2}$$

We find the distance between students 1 and 2

$$d_{12} = \sqrt{(2-8)^2 + (4-2)^2} = 6.325$$

We find the distance between students 1 and 3

$$d_{13} = \sqrt{(2-9)^2 + (4-3)^2} = 7.071$$

Continuing to calculate the rest of the Euclidean distances, it is possible to formulate the distance matrix as follows:

$$D = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 0 & 6,325 & 7.071 & 1.414 & 7.159 \\ & 0 & 1.414 & 7.616 & 1.116 \\ & & 0 & 8.246 & 2.062 \\ & & & 0 & 8.500 \\ & & & & 0 \end{bmatrix} \end{array}$$

We note that the smallest element of this matrix is 1.116, which corresponds to students 2 and 5, because they are the most similar in expenses, so we can create the first cluster from them.

Therefore, the previous matrix will be updated by merging the components of the first cluster together into one column and one row, and the values included in the new column will be updated according to the following equation:

$$d_{(u,v)_j} = \frac{1}{2}\left[d_{uj} + d_{vj}\right]$$

or

$$D(P_k, (P_i, P_j)) = \frac{1}{2}\left((P_k, P_i) + (P_k, P_j)\right)$$

Accordingly, the following points will be updated:

$$D(P_1,(P_2,P_5)), D((P_2,P_5),P_3), D((P_2,P_5),P_4).$$

The new matrix will be as follows:

$$
D = \begin{array}{c} \\ 1 \\ 2,5 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2,5 & 3 & 4 \end{array} \\ \begin{bmatrix} 0 & ? & 7.071 & 1.414 \\ & 0 & ? & ? \\ & & 0 & 8.246 \\ & & & 0 \end{bmatrix} \end{array}.
$$

That is, the values bearing the question mark in the above matrix, which are located in the same row and same column for the last cluster, will be counted. As for the rest of the values, they are the same and do not change if both of the above matrices are compared.
It will be calculated as follows:

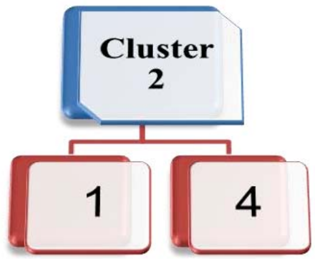
$$
\begin{aligned}
D(P_1,(P_2,P_5)) &= \frac{1}{2}((P_1,P_2)+(P_1,P_5)) \\
&= \frac{1}{2}(6.325+7.159) \\
&= \frac{1}{2}(13.484) = 6.742
\end{aligned}
$$

$$D((P_2,P_5),P_3) = \frac{1}{2}((P_2,P_3)+(P_3,P_5))$$
$$= \frac{1}{2}(1.414+2.062)$$
$$= \frac{1}{2}(3.476) = 1.738$$
$$D((P_2,P_5),P_4) = \frac{1}{2}((P_2,P_4)+(P_4,P_5))$$
$$= \frac{1}{2}(7.616+8.500)$$
$$= \frac{1}{2}(16.116) = 8.058$$

So the new matrix will be as follows:

$$\begin{array}{c} \\ 1 \\ D = 2,5 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2,5 & 3 & 4 \end{array} \\ \begin{bmatrix} 0 & 6.742 & 7.071 & 1.414 \\ & 0 & 1.738 & 8.058 \\ & & 0 & 8.246 \\ & & & 0 \end{bmatrix} \end{array}.$$

We note that the smallest element of this matrix is 1.414, which corresponds to students 1 and 4, because they are the most similar in expenditures after students 2 and 5 were clustered, so we can create the second cluster from them.



Therefore, the previous matrix will be updated by merging the components of the second cluster together into one column and one row, and the values included in the new column will be updated according to the following matrix:

$$D = \begin{array}{c} \\ 1,4 \\ 2,5 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1,4 & 2,5 & 3 \end{array} \\ \begin{bmatrix} 0 & ? & ? \\ & 0 & 1.738 \\ & & 0 \end{bmatrix} \end{array}.$$

That is, the values bearing the question mark in the above matrix, which are located in the same row and same column for the last cluster, will be counted. As for the rest of the values, they are the same and do not change if both of the previous matrices are compared. Accordingly, the following points will be updated:

$$D((P_1, P_4),(P_2, P_5)), D((P_1, P_4), P_3).$$

It will be calculated as follows:

$$D((P_1,P_4),(P_2,P_5)) = \frac{1}{2}((P_1,(P_2,P_5)) + ((P_2,P_5),P_4))$$
$$= \frac{1}{2}(6.742 + 8.058)$$
$$= \frac{1}{2}(14.7978) = 7.3989$$

$$D((P_1,P_4),P_3) = \frac{1}{2}((P_1,P_3) + (P_3,P_4))$$
$$= \frac{1}{2}(7.071 + 8.246)$$
$$= \frac{1}{2}(15.317) = 7.6585$$

So the new matrix will be as follows:

$$\begin{array}{cc} & \begin{array}{ccc} 1,4 & 2,5 & 3 \end{array} \\ D = \begin{array}{c} 1,4 \\ 2,5 \\ 3 \end{array} & \begin{bmatrix} 0 & 7.399 & 7.659 \\ & 0 & 1.738 \\ & & 0 \end{bmatrix}. \end{array}$$

So the matrix will be. We note that the smallest element of this matrix is 1.738, which corresponds to the cluster of students (2,5) and student 3, because they are the most similar in expenditures after students 2 and 5 were clustered in one cluster and students 1 and 4 were clustered in one cluster, so we can create from them The new third cluster is as follows:

Therefore, the previous matrix will be updated by merging the components of the third cluster together into one column and one row, and the values included in the new column will be updated according to the following matrix:

$$D = \begin{array}{c} \\ 1,4 \\ (2,5),3 \end{array} \begin{array}{c} \begin{array}{cc} 1,4 & (2,5),3 \end{array} \\ \begin{bmatrix} 0 & ? \\ & 0 \end{bmatrix} \end{array}.$$

That is, the value bearing the question mark in the above matrix, which is located in the same row and column of the last cluster, will be calculated. Accordingly, the following point will be updated:

$$D((P_1, P_4), [(P_2, P_5), P_3]).$$

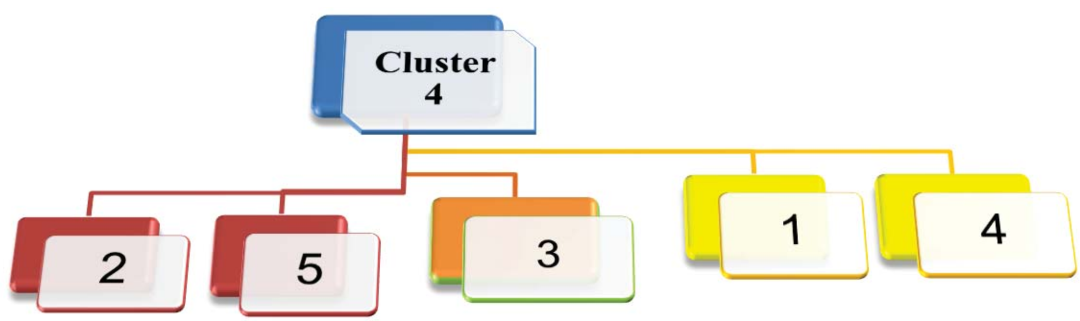
It will be calculated as follows:

$$D((P_1,P_4),[(P_2,P_5),P_3]) = \frac{1}{2}(((P_1,P_4),(P_2,P_5)) + ((P_1,P_4),P_3))$$
$$= \frac{1}{2}(7.399 + 7.659)$$
$$= \frac{1}{2}(15.058) = 7.529$$

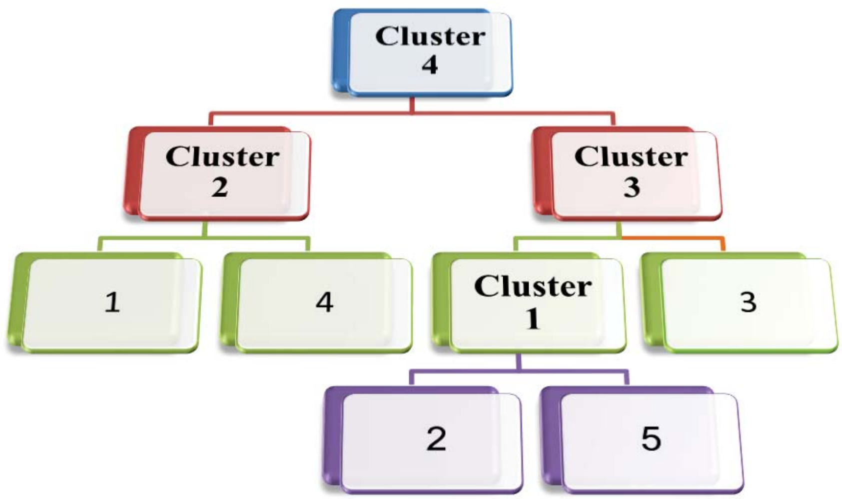
So the new matrix will be as follows:

$$\begin{array}{ccc} & 1,4 & (2,5),3 \\ D = \quad 1,4 & \left[\begin{array}{l} 0 \\ \\ \end{array}\right. & \left.\begin{array}{c} 7.529 \\ 0 \end{array}\right]. \\ (2,5),3 & & \end{array}$$

Therefore, we can create the fourth cluster from them.



Therefore, at the end of the problem, the hierarchical cluster tree will be as follows:


Cluster 4
Cluster 2
Cluster 3
1
4
Cluster 1
3
2
5