Ward's Formula :-

$$\therefore \Delta(A,B) = \left(\frac{n_A \cdot n_B}{n_A + n_B}\right) \|\mu_A - \mu_B\|^2$$

→ When we merge A and B

$$\therefore \mu_{A \cup B} = \frac{n_A \mu_A + n_B \mu_B}{(n_A + n_B)}$$

$$\therefore D(A \cup B, C) = \frac{(n_A + n_B) \cdot n_c}{(n_A + n_B + n_c)} \cdot \|\mu_{A \cup B} - \mu_c\|^2$$

expanding, $\mu_{A \cup B} - \mu_c$,

$$\therefore \mu_{A \cup B} - \mu_c = \frac{n_A (\mu_A - \mu_c) + n_B (\mu_B - \mu_c)}{(n_A + n_B)}$$

$$\therefore \|\mu_{A \cup B} - \mu_c\|^2 = \frac{n_A^2 \|\mu_A - \mu_c\|^2 + n_B^2 \|\mu_B - \mu_c\|^2 + 2 n_A n_B (\mu_A - \mu_c)(\mu_B - \mu_c)}{(n_A + n_B)^2}$$

From Euclidean geometry,

$$\therefore \|\mu_A - \mu_B\|^2 = \|\mu_A - \mu_c\|^2 + \|\mu_B - \mu_c\|^2 - 2(\mu_A - \mu_c)(\mu_B - \mu_c).$$

$$\therefore 2(\mu_A - \mu_c)(\mu_B - \mu_c) = \|\mu_A - \mu_c\|^2 + \|\mu_B - \mu_c\|^2 - \|\mu_A - \mu_B\|^2$$

Then,

$$\|\mu_{A \cup B} - \mu_c\|^2 = \frac{n_A^2 \|\mu_A - \mu_c\|^2 + n_B^2 \|\mu_B - \mu_c\|^2 + n_A n_B \left(\|\mu_A - \mu_c\|^2 + \|\mu_B - \mu_c\|^2 - \|\mu_A - \mu_B\|^2\right)}{(n_A + n_B)^2}$$

$$\therefore \cancel{\|\mu_{A \cup B} - \mu_c\|^2 = n_A^2 \|\mu_A - \mu_c\|^2 + n_B^2 \|\mu_B - \mu_c\|^2}$$

$$\therefore \|\mu_{A \cup B} - \mu_c\|^2 = \frac{n_A (n_A + n_B) \|\mu_A - \mu_c\|^2 + n_B (n_A + n_B) \|\mu_B - \mu_c\|^2 - n_A n_B \|\mu_A - \mu_B\|^2}{(n_A + n_B)^2}$$

$$\therefore \boxed{\|\mu_{A \cup B} - \mu_c\|^2 = \frac{n_A \|\mu_A - \mu_c\|^2}{(n_A + n_B)} + \frac{n_B \|\mu_B - \mu_c\|^2}{(n_A + n_B)} - \frac{n_A n_B \|\mu_A - \mu_B\|^2}{(n_A + n_B)^2}}$$

(i)

We know,

$$D(A\cup B, C) = \frac{(n_A+n_B).n_c}{(n_A+n_B+n_c)} \|\mu_{A\cup B} - \mu_c\|^2$$

Multiplying eq^n (i) by $\frac{(n_A+n_B)n_c}{(n_A+n_B+n_c)}$

$$\therefore D(A\cup B, C) = \frac{(n_A+n_B).n_c}{(n_A+n_B+n_c)}\left[\frac{n_A\|\mu_A-\mu_c\|^2+n_B\|\mu_B-\mu_c\|^2}{(n_A+n_B)} - \frac{n_An_B}{(n_A+n_B)^2}.\|\mu_A-\mu_B\|^2\right]$$

$$= \frac{n_c}{(n_A+n_B+n_c)}\left[n_A\|\mu_A-\mu_c\|^2+n_B\|\mu_B-\mu_c\|^2\right]$$

$$- \frac{\cancel{(n_A+n_B)}\,n_c.n_An_B}{(n_A+n_B+n_c).(n_A+n_B)^{\cancel{2}}}\|\mu_A-\mu_B\|^2.$$

$$\boxed{\therefore D(A\cup B, C) = \frac{n_c}{(n_A+n_B+n_c)}\left[n_A\|\mu_A-\mu_c\|^2+n_B\|\mu_B-\mu_c\|^2\right] - \frac{n_An_Bn_c}{(n_A+n_B)(n_A+n_B+n_c)}\|\mu_A-\mu_B\|^2}$$

Now, we know,

$$\therefore \|\mu_A-\mu_c\|^2 = \left(\frac{n_A+n_c}{n_A.n_c}\right)D(A,C)$$ → from ward's method.

$$\therefore \|\mu_B-\mu_c\|^2 = \left(\frac{n_B+n_c}{n_B.n_c}\right).D(B,C).$$

$$\therefore \|\mu_A-\mu_B\|^2 = \left(\frac{n_A+n_B}{n_A.n_B}\right).D(A,B)$$

$$\therefore D(A,C) = \left(\frac{n_A.n_c}{n_A+n_c}\right)\|\mu_A-\mu_c\|^2$$

$$\Rightarrow \|\mu_A-\mu_c\|^2 = \left(\frac{n_A+n_c}{n_A.n_c}\right).D(A,C)$$

1st term :-

$$\therefore \frac{n_c}{(n_A+n_B+n_c)}\left[n_A\|\mu_A-\mu_c\|^2+n_B\|\mu_B-\mu_c\|^2\right].$$

$$\Rightarrow \frac{\cancel{n_c}}{(n_A+n_B+n_c)}\cdot\left[\cancel{n_A}\cdot\frac{(n_A+n_c)}{\cancel{n_A}\cdot\cancel{n_c}}D(A,C)+\cancel{n_B}\cdot\frac{(n_B+n_c)}{\cancel{n_B}\cdot\cancel{n_c}}D(B,C)\right].$$

$$\Rightarrow \frac{(n_A+n_c)\cdot D(A,C)+(n_B+n_c)D(B,C)}{(n_A+n_B+n_c)}.$$

2nd term :-

$$\therefore \frac{n_A\cdot n_B\cdot n_c}{(n_A+n_B)(n_A+n_B+n_c)}\|\mu_A-\mu_B\|^2=\frac{\cancel{n_A}\cdot\cancel{n_B}\cdot n_c}{\cancel{(n_A+n_B)}\cdot(n_A+n_B+n_c)}\cdot\frac{\cancel{(n_A+n_B)}\cdot D(A,B)}{\cancel{n_A}\cdot\cancel{n_B}}$$

$$=\frac{n_c\cdot D(A,B)}{(n_A+n_B+n_c)}.$$

Put Both terms together :-

$$\therefore D(A\cup B,C)=\frac{(n_A+n_c)\cdot D(A,C)+(n_B+n_c)D(B,C)}{(n_A+n_B+n_c)}-\frac{n_c\,D(A,B)}{(n_A+n_B+n_c)}$$

$$\Rightarrow \boxed{D(A\cup B,C)=\frac{(n_A+n_c)\cdot D(A,C)+(n_B+n_c)D(B,C)-n_c\cdot D(A,B)}{(n_A+n_B+n_c)}}.$$

↳ This is Lance William's Update for Ward's Formula.

- **Strengths** :- Less suspectible to noise and outliers.
- **Limitations** :- Biased towards globular clusters.

Ward's Method :-

- Similarity of two clusters is based on the increase in squared error (SSE) when two clusters are merged.

$$\therefore \boxed{\text{Inter-cluster Similarity} = \left(\frac{n_A \cdot n_B}{|n_A| + |n_B|}\right) \times \|\bar{X}_A - \bar{X}_B\|^2}$$

where, $\bar{X}_A \rightarrow$ mean of cluster-A
$\bar{X}_B \rightarrow$ mean of cluster-B
$n_A \rightarrow$ # of pts in cluster A.
$n_B \rightarrow$ # of pts in cluster-B.

ex:-

	1	2	3	4	5	6
1.	0	0.24	0.22	0.37	0.34	0.23
2.	0.24	0	0.15	0.20	0.14	0.25
3	0.22	0.15	0	0.15	0.28	0.11
4	0.37	0.20	0.15	0	0.29	0.22
5	0.34	0.14	0.28	0.29	0	0.39
6	0.23	0.25	0.11	0.22	0.39	0

→ First {3, 6} forms a cluster.

$\therefore \cancel{d((3,6),1) = (|3|+|1|)\,d(3,1) + (|6|+|1|)\,d(6,1) - |1|\,d(3,6)}$

	1	2	{3,6}	4	5
1	0.	0.24	0.26	0.37	0.34
2	0.24	0	0.23	0.20	0.14
{3,6}	0.26	0.23	0	0.21	0.41
4	0.37	0.20	0.21	0	0.29
5	0.34	0.14	0.41	0.29	0

$\therefore D((3,6),1) = \frac{(1+1)D(3,1) + (1+1)D(6,1) - 1.D(3,6)}{1+1+1}$

$\Rightarrow D((3,6),1) = \frac{2\times0.22 + 2\times0.23 - 1\times0.11}{3} = 0.26.$

$\therefore D((3,6),4) = \frac{(1+1)D(3,4) + (1+1).D(6,4) - 1.D(3,6)}{(1+1+1)}$

$= \frac{2\times0.15 + 2\times0.22 - 0.11}{3} = 0.21.$

$\therefore D((3,6),2) = \frac{(1+1)D(3,2) + (1+1)D(6,2) - 1.D(3,6)}{(1+1+1)}$

$= \frac{2\times0.15 + 2\times0.25 - 0.11}{3} = 0.23.$

$\therefore D((3,6),5) = \frac{(1+1)D(3,5) + (1+1)D(6,5) - 1.D(3,6)}{1+1+1}$

$= \frac{2\times0.28 + 2\times0.39 - 0.11}{3} = 0.41.$

Next, {2,5} forms a cluster.

	1	{2,5}	{3,6}	4
1	0	0.34	0.26	0.37
{2,5}	0.34	0	0.41	0.28
{3,6}	0.26	0.41	0	0.21
4	0.37	0.28	0.21	0

$\therefore D((2,5),1) = \frac{(1+1)D(2,1) + (1+1).D(5,1) - 1.D(2,5)}{(1+1+1)}$

$= \frac{2\times0.24 + 2\times0.34 - 0.14}{3} = 0.34.$

$\therefore D((2,5),4) = \frac{(1+1).D(2,4) + (1+1).D(5,4) - 1.D(2,5)}{(1+1+1)}$

$= \frac{2\times0.20 + 2\times0.29 - 1\times0.14}{3} = 0.28$

$$d((2,5),(3,6)) = \frac{(1+2).D(2,(3,6)) + (1+2).D(5,(3,6)) - 2.D(2,5)}{(2+2)}$$

$$= \frac{3(0.23) + 3(0.41) - 2(0.14)}{4} = \frac{1.64}{4} = \underline{0.41}$$

Next, {3,6} &(4) forms a cluster.

	1	{2,5}	{3,4,6}
1	0	0.34	0.3275
{2,5}	0.34	0	0.41
{3,4,6}	0.3275	0.41.	0

$$\therefore d((4,(3,6)),1) = \frac{(1+1).D(4,1) + (2+1).D((3,6),1) - 1.D(4,(3,6))}{(3+1)}$$

$$= \frac{2\times(0.37) + 3(0.26) - 1.\times 0.21}{4} = 0.3275.$$

$$\therefore d((4,(3,6)),(2,5)) = \frac{(1+2).D(4,(2,5)) + (2+2).D((3,6),(2,5)) - 2.D(4,(3,6))}{(3+2)}$$

$$= \frac{3\times 0.28 + 4\times 0.41 - 2\times 0.21}{5.} = 0.41$$

Next, {3,4,6} & 1 forms a cluster.
And lastly {1,3,4,6} & {2,5} forms a clusters.