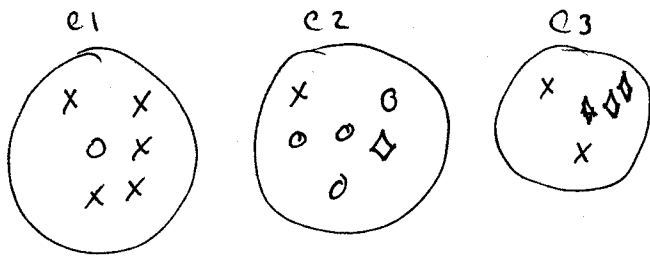


External Criteria Evaluation of Clustering - Rand Index.



$$RI = \frac{TP + TN}{TP + TN + FP + FN}$$

Step 1: get TP + FP. This is the total # of pairs. So $\binom{c_1 N}{2} + \binom{c_2 N}{2} + \dots + \binom{c_k N}{2}$

$$TP + FP = \binom{6}{2} + \binom{6}{2} + \binom{5}{2} \rightarrow \text{Recall: } \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Total # of items in each cluster. (pointing to 6, 6, 5)
pairs (pointing to 2, 2, 2)

AND The pairs that are appropriately grouped together.
TP = $\binom{5}{2} + \binom{4}{2} + \binom{3}{2} + \binom{2}{2} = 20$ \therefore FP = 20
x's o's d's x's

= 40

Step 2 get total # of pairs.

$$= N(N-1)/2 = 17(16)/2 = 136$$

Step 3

Continue filling in the blanks by solving for the 'negatives'

$$FN + TN = \text{Total Pairs} - (TP + FP)$$

$$= 136 - 40$$

$$= 96$$

Step 4

Solve for False Negative: $\sum d_{ca} \cdot d_{cb}$

x

document count by type in each cluster.

$$c_1, c_2 = 5 \times 1 = 5$$

$$c_1, c_3 = 5 \times 2 = 10$$

$$c_2, c_3 = 1 \times 2 = 2$$

$$\left. \begin{array}{l} c_1, c_2 = 5 \\ c_1, c_3 = 10 \\ c_2, c_3 = 2 \end{array} \right\} \Sigma = 24$$

$$\text{So, } FN = 24. \therefore TN = 96 - 24 = 72$$

o

$$c_1, c_2 = 1 \times 4 = 4$$

$$c_2, c_3 = 1 \times 3 = 3$$

$$\text{So, } RI = \frac{20 + 72}{20 + 72 + 20 + 24} = 0.68$$