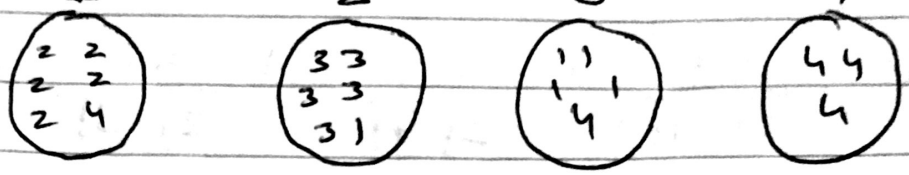
Clusters 1 2 3 4



Problem 1

1) Purity $(C, \Omega) = \frac{1}{N} \sum_k \max_j |C_k \cap \omega_j|$

$= \frac{1}{20} (5+5+4+3)$

$= \frac{17}{20}$

Purity = 0.85

Calculating TP, TN, FP, FN.

Since there are Total 20 data rows and there are 2 possibilities, for example (1,2) (1,3) so on ~~till~~ we get

$TP + FP + FN + TN = {}^{20}C_2 = \binom{20}{2} = \frac{20!}{2!(18!)}$

$\therefore TP + FP + FN + TN = 190.$

Now we calculate TP + FP since we know ~~that~~ the points belonging to same cluster.

$TP + FP = \binom{6}{2} + \binom{6}{2} + \binom{5}{2} + \binom{3}{2}$

$$TP+FP = \frac{6!}{2!(4!)} + \frac{6!}{2!\cdot 4!} + \frac{5!}{2!\cdot 3!} + \frac{3!}{2!}$$

$$= \frac{6\times 5\times 4!}{2\times 4!} + \frac{6\times 5\times 4!}{2\times 4!} + \frac{5\times 4\times 3!}{2\times 3!} + \frac{3\times 2!}{2!}$$

$$= 15 + 15 + 10 + 3$$

$$= 43.$$

$$TP = (5C_2) + (5C_2) + (4C_2) + (3C_2)$$

$$= \frac{5!}{2!\cdot 3!} + \frac{5!}{2!3!} + \frac{4!}{2!\cdot 2!} + \frac{3!}{2!}$$

$$= 29$$

$FP = 43 - TP = 43 - 29 = 14.$

$\therefore FN + TN = 190 - 43 = 147.$

Now we can calculate False negative which is same class, different cluster. For example we see that cluster 1 has a data point with class 4. and cluster 3 has a data point with class 4 so they form 1 point. This way we get 7 possible points for "4" and 4 possible points for class 1.

$FN = 4+4+3 = 11$

$TN = 147 - 11 = 136.$

Precision $P = \frac{TP}{TP+FP} = \frac{29}{43} = 0.674$

Recall $R = \frac{TP}{TP+FN} = \frac{29}{29+11} = 0.725$

F-measure $= \frac{2P*R}{P+R} = \frac{2*(0.674 * 0.725)}{0.674 + 0.725}$

$= 0.699$

NMI

$$H(\Omega) = -\sum_k \frac{|w_k|}{N} \log \frac{|w_k|}{N}$$

$$= -\left(\frac{5}{20}\log\frac{5}{20} + \frac{5}{20}\log\frac{5}{20} + \frac{5}{20}\log\frac{5}{20} + \frac{5}{20}\log\frac{5}{20}\right)$$

$$= -(0.25(-2) + 0.25(-2) + 0.25(-2) + (0.25)(-2))$$
$$= -(-0.5 - 0.5 - 0.5 - 0.5)$$
$$= -(-2)$$
$$= 2.$$

$$H(C) = -\sum_{k} \frac{|C_k|}{N} \log \frac{|C_k|}{N}$$

$$= -\left(\frac{6}{20} \log \frac{6}{20} + \frac{6}{20} \log \frac{6}{20} + \frac{5}{20} \log \frac{5}{20} + \frac{3}{20} \log \frac{3}{20} \right)$$

$$= -(0.3(-1.7369) + 0.3(-1.7369) + 0.25(-2) + 0.15(-2.73$$

$$= 1.953$$

$$I(\Omega, C) = \sum_{k} \sum_{j} \frac{|\omega_k \cap c_j|}{N} \log \frac{N \cdot |\omega_k \cap c_j|}{|\omega_k| \, |c_j|}$$

$$= \frac{1}{20} \log \frac{20 \times 1}{5 \times 6} + \frac{4}{20} \log \frac{20 \times 4}{5 \times 5} + \frac{5}{20} \log \frac{20 \times 5}{5 \times 6} + \frac{5}{20} \log \frac{20 \times 5}{5 \times 6}$$

$$+ \frac{1}{20} \log \frac{20 \times 1}{5 \times 6} + \frac{1}{20} \log \frac{20 \times 1}{5 \times 5} + \frac{3}{20} \log \frac{20 \times 3}{5 \times 3}$$

$$= 0.05 \log(2/3) + 0.2 \log(16/5) + 0.25 \log(10/3) + 0.25 \log(10/3)$$

$$+ 0.05 \log(2/3) + 0.05 \log(4/5) + \frac{3}{20} \log(4)$$

$$= 1.4295$$

$$NMI(\Omega, C) = I(\Omega, C) / \sqrt{H(\Omega) \cdot H(\Omega)}$$

$$= 1.4925 / [\sqrt{1.953 \cdot 2}]$$

$$= 1.4925 / \sqrt{3.906}$$

$$= 0.723$$