Matching-Based Measures (II): F-Measure

- **Precision**: The fraction of points in C_i from the majority partition T_{j_i} (i.e., the same as purity), where j_i is the partition that contains the maximum # of points from C_i

  - Ex. For the green table

    $$prec_i = \frac{1}{n_i} \max_{j=1}^{k} \{n_{ij}\} = \frac{n_{ij_i}}{n_i}$$

    - $prec_1 = 30/50$; $prec_2 = 20/25$; $prec_3 = 25/25$

- **Recall**: The fraction of point in partition T_{j_i} shared in common with cluster C_i, where $m_{j_i} = |T_{j_i}|$
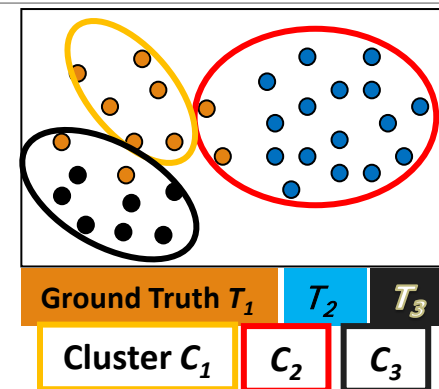
  - Ex. For the green table

    $$recall_i = \frac{n_{ij_i}}{|T_{j_i}|} = \frac{n_{ij_i}}{m_{j_i}}$$

    - $recall_1 = 30/35$; $recall_2 = 20/40$; $recall_3 = 25/25$

- **F-measure** for C_i: The harmonic means of $prec_i$ and $recall_i$: $F_i = \frac{2n_{ij_i}}{n_i + m_{j_i}}$

- F-measure for clustering C: average of all clusters: $F = \frac{1}{r}\sum_{i=1}^{r} F_i$

  - Ex. For the green table

    - $F_1 = 60/85$; $F_2 = 40/65$; $F_3 = 1$; $F = 0.774$

Ground Truth T_1	T_2	T_3
Cluster C_1	C_2	C_3

$C \backslash T$	T_1	T_2	T_3	Sum
C_1	0	20	30	50
C_2	0	20	5	25
C_3	25	0	0	25
m_j	25	40	35	100

3