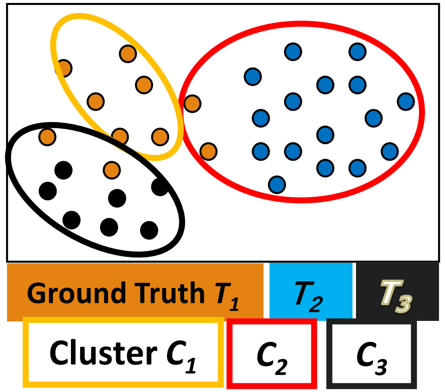
Entropy-Based Measures (I): Conditional Entropy

- **Entropy of clustering *C*:** $H(\mathcal{C}) = -\sum_{i=1}^{r} p_{C_i} \log p_{C_i}$ $\quad p_{C_i} = \frac{n_i}{n}$ (i.e., the probability of cluster C_i)
- **Entropy of partitioning *T*:** $H(\mathcal{T}) = -\sum_{j=1}^{k} p_{T_j} \log p_{T_j}$
- **Entropy of *T* with respect to cluster C_i:** $H(\mathcal{T}|C_i) = -\sum_{j=1}^{k} (\frac{n_{ij}}{n_i}) \log(\frac{n_{ij}}{n_i})$
- **Conditional entropy of *T* with respect to clustering *C*:**

$$H(\mathcal{T}|\mathcal{C}) = -\sum_{i=1}^{r} (\frac{n_i}{n}) H(\mathcal{T}|C_i) = -\sum_{i=1}^{r}\sum_{j=1}^{k} p_{ij} \log(\frac{p_{ij}}{p_{C_i}})$$

  - The more a cluster's members are split into different partitions, the higher the conditional entropy
  - For a perfect clustering, the conditional entropy value is 0, where the worst possible conditional entropy value is *log k*



$$\begin{aligned} H(\mathcal{T}|\mathcal{C}) &= -\sum_{i=1}^{r}\sum_{j=1}^{k} p_{ij}(\log p_{ij} - \log p_{C_i}) = -\sum_{i=1}^{r}\sum_{j=1}^{k} p_{ij} \log p_{ij} + \sum_{i=1}^{r} (\log p_{C_i} \sum_{j=1}^{k} p_{ij}) \\ &= -\sum_{i=1}^{r}\sum_{j=1}^{k} p_{ij} \log p_{ij} + \sum_{i=1}^{r} (p_{C_i} \log p_{C_i}) = H(\mathcal{C},\mathcal{T}) - H(\mathcal{C}) \end{aligned}$$