The background of the slide is a complex, abstract composition. It features a network of thin, reddish-brown lines forming a web-like structure. Scattered throughout are numerous small, colored dots in shades of green, blue, and orange. On the left side, there is a vertical strip containing a grid of small, light-colored squares. The overall aesthetic is technical and data-driven.

External Measures II: Entropy-Based Measures

Entropy-Based Measures (I): Conditional Entropy

□ Entropy of clustering \mathcal{C} : $H(\mathcal{C}) = - \sum_{i=1}^r p_{C_i} \log p_{C_i}$ $p_{C_i} = \frac{n_i}{n}$ (i.e., the probability of cluster C_i)

□ Entropy of partitioning \mathcal{T} : $H(\mathcal{T}) = - \sum_{j=1}^k p_{T_j} \log p_{T_j}$

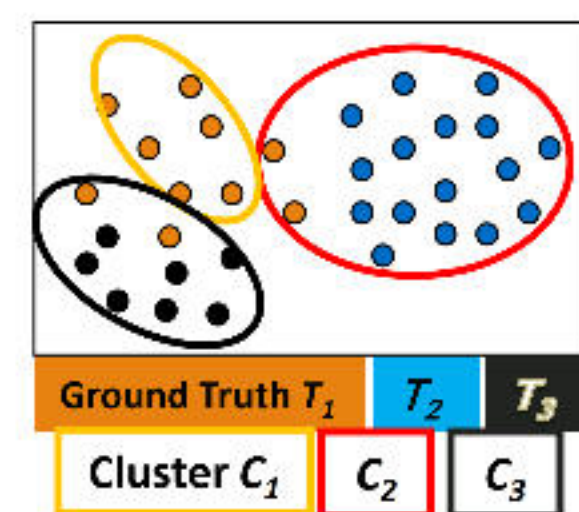
□ Entropy of \mathcal{T} with respect to cluster C_i : $H(\mathcal{T}|C_i) = - \sum_{j=1}^k \left(\frac{n_{ij}}{n_i}\right) \log \left(\frac{n_{ij}}{n_i}\right)$

□ Conditional entropy of \mathcal{T} with respect to clustering \mathcal{C} : $H(\mathcal{T}|\mathcal{C}) = - \sum_{i=1}^r \left(\frac{n_i}{n}\right) H(\mathcal{T}|C_i) = - \sum_{i=1}^r \sum_{j=1}^k p_{ij} \log \left(\frac{p_{ij}}{p_{C_i}}\right)$

□ 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 $\log k$

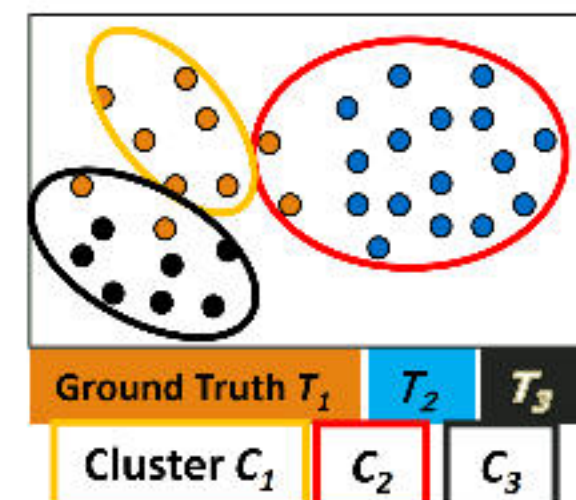
$$\begin{aligned} H(\mathcal{T}|\mathcal{C}) &= - \sum_{i=1}^r \sum_{j=1}^k p_{ij} (\log p_{ij} - \log p_{C_i}) = - \sum_{i=1}^r \sum_{j=1}^k p_{ij} \log p_{ij} + \sum_{i=1}^r (\log p_{C_i} \sum_{j=1}^k p_{ij}) \\ &= - \sum_{i=1}^r \sum_{j=1}^k p_{ij} \log p_{ij} + \sum_{i=1}^r (p_{C_i} \log p_{C_i})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



Entropy-Based Measures (II): Normalized Mutual Information (NMI)

■ Mutual information:

- Quantifies the amount of shared info between the clustering C and partitioning T $I(C, T) = \sum_{i=1}^r \sum_{j=1}^k p_{ij} \log\left(\frac{p_{ij}}{p_{C_i} \cdot p_{T_j}}\right)$
(Handwritten: $I(C, T) = H(C) - H(C|T)$)
- Measures the dependency between the observed joint probability p_{ij} of C and T , and the expected joint probability $p_{C_i} \cdot p_{T_j}$ under the independence assumption
- When C and T are independent, $p_{ij} = p_{C_i} \cdot p_{T_j}$, $I(C, T) = 0$. However, there is no upper bound on the mutual information



■ Normalized mutual information (NMI)

$$NMI(C, T) = \sqrt{\frac{I(C, T)}{H(C)} \cdot \frac{I(C, T)}{H(T)}} = \frac{I(C, T)}{\sqrt{H(C) \cdot H(T)}}$$

- Value range of NMI: $[0, 1]$. Value close to 1 indicates a good clustering