

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

□ Mutual information:

- Quantifies the amount of shared info between the clustering C and partitioning T
$$I(C, T) = \sum_{i=1}^r \sum_{j=1}^k p_{ij} \log\left(\frac{p_{ij}}{p_{C_i} \cdot p_{T_j}}\right)$$
- 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

