

Clustering Evaluation

Mahdi Roozbahani
Georgia Tech



Clustering Evaluation

- Clustering evaluation aims at quantifying the goodness or quality of the clustering.
- Two main categories of measures:
 - **External measures**: employ external ground-truth
 - **Internal measures**: derive goodness from the data itself

Outline

- External measures for clustering evaluation 
 - Matching-based measures
 - Entropy-based measures
 - Pairwise measures
- Internal measures for clustering evaluation
 - Graph-based measures
 - Davies-Bouldin Index
 - Silhouette Coefficient

External Measures

External measures assume that the correct or ground-truth clustering is known *a priori*, which is used to evaluate a given clustering.

Let $\mathbf{D} = \{\mathbf{x}_i\}_{i=1}^n$ be a dataset consisting of n points in a d -dimensional space, partitioned into k clusters. Let $y_i \in \{1, 2, \dots, k\}$ denote the ground-truth cluster membership or label information for each point.

The ground-truth clustering is given as $\mathcal{T} = \{T_1, T_2, \dots, T_k\}$, where the cluster T_j consists of all the points with label j , i.e., $T_j = \{\mathbf{x}_i \in \mathbf{D} | y_i = j\}$. We refer to $\mathcal{T}$ as the ground-truth *partitioning*, and to each T_j as a *partition*.

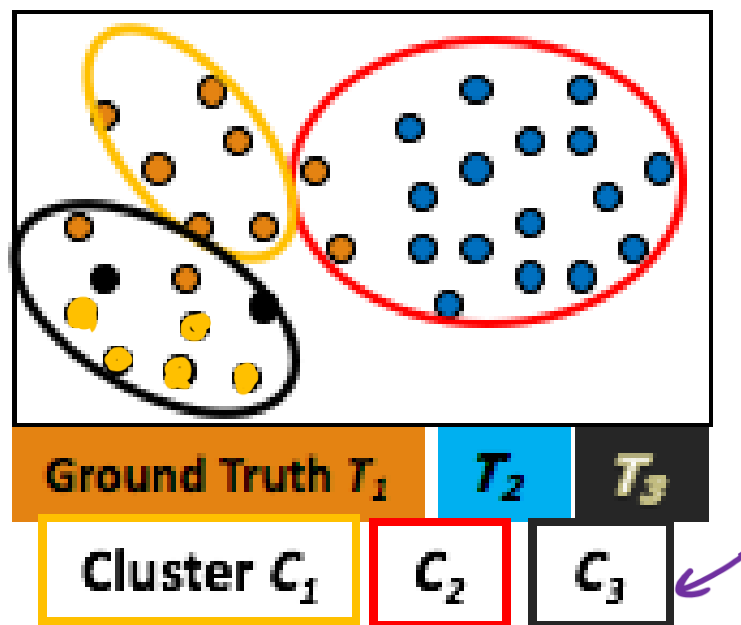
Let $\mathcal{C} = \{C_1, \dots, C_r\}$ denote a clustering of the same dataset into r clusters, obtained via some clustering algorithm, and let $\hat{y}_i \in \{1, 2, \dots, r\}$ denote the cluster label for $\mathbf{x}_i$.

So **k** is the number of ground truth partitions (**T**) and **r** is the number of clusters (**C**) obtained by algorithm

n_{ij} = Number of data points in cluster **i** which are also in ground truth partition **j**

Matching-Based Measures (I): Purity

- **Purity**: Quantifies the extent that cluster C_i contains points only from one (ground truth) partition:



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

Partition
cluster

$$purity_3 = \frac{1}{n_3} \max(n_{31}, n_{32}, n_{33}) = \frac{7}{9}$$

$$= \frac{1}{9} \max(2, 0, 7) = \frac{7}{9}$$

The Total purity of clustering C is the weighted sum of the cluster-wise purity:

$$purity = \sum_{i=1}^r \frac{n_i}{n} purity_i = \frac{1}{n} \sum_{i=1}^r \max_{j=1}^k \{n_{ij}\}$$

What is purity value for a perfect clustering?

Purity = 1

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

$$purity = \sum_{i=1}^r \frac{n_i}{n} purity_i = \frac{1}{n} \sum_{i=1}^r \max_{j=1}^k \{n_{ij}\}$$

n_{ij} Table

Partians



Example:

$$purity_1 = 30/50;$$

$$purity_2 = 20/25;$$

$$purity_3 = 25/25;$$

$$purity = (30 + 20 + 25)/100 = 0.75$$

clusters



C\T	T ₁	T ₂	T ₃	Sum
C ₁	0	20	30	⇒ 50
C ₂	0	20	(5) n_{23}	⇒ 25
C ₃	25	0	0	⇒ 25
m_j	↓ 25	↓ 40	↓ 35	100

Two clusters may be matched to the same partition

C1 is more paired with T3
C2 is more paired with T2

$C \backslash T$	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

$$\text{purity} = (30 + 20 + 25)/100 = 0.75$$

C1 is more paired with T2
C2 is more paired with T2

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

$$\text{purity} = (30 + 20 + 25)/100 = 0.75$$

Maximum weight matching: Only one cluster can match one partition

Ex. If C1 is more paired with T2 **THEN** C2 and C3 cannot paired with T2

<i>C\T</i>	<i>T</i> ₁	<i>T</i> ₂	<i>T</i> ₃	Sum
<i>C</i> ₁	0	30	20	50
<i>C</i> ₂	0	20	5	25
<i>C</i> ₃	25	0	0	25
<i>m</i> _{<i>j</i>}	25	50	25	100

C1 is more paired with T2 = $\frac{30+5+25}{100} = 0.6$

C1 is more paired with T3 = $\frac{20+20+25}{100} = 0.65$

MAX

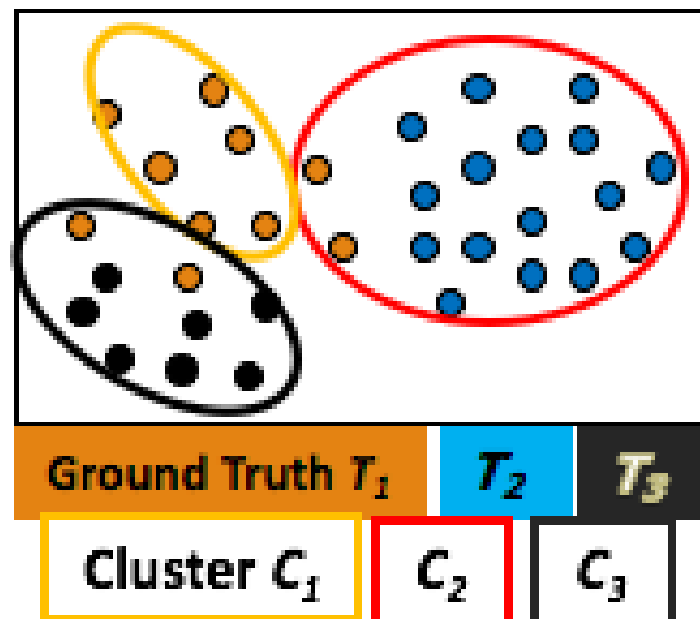
Purity = 0.65

Matching-Based Measures (II): Maximum Matching

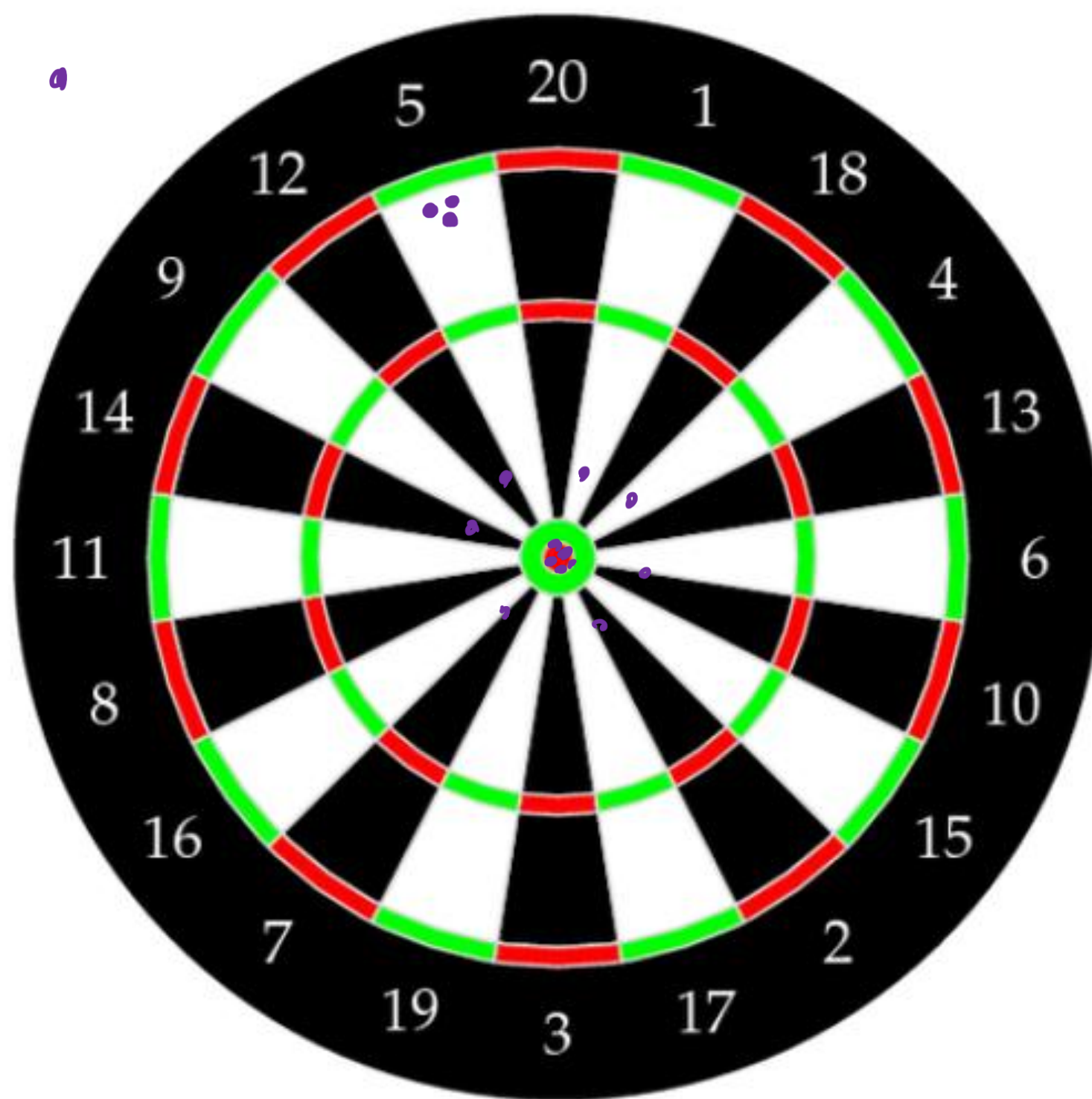
- **Drawback of purity**: two clusters may be matched to the same partition.
- **Maximum matching**: the maximum purity under the one-to-one matching constraint.
 - Examine all possible pairwise matching between C and T and choose the best (the maximum)

Example:

Maximum matching = $0.65 > 0.6$



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





In a general context: **Precision, Recall and Accuracy**

Correct
prediction

Wrong
prediction

Positive → cat
Negative → dog

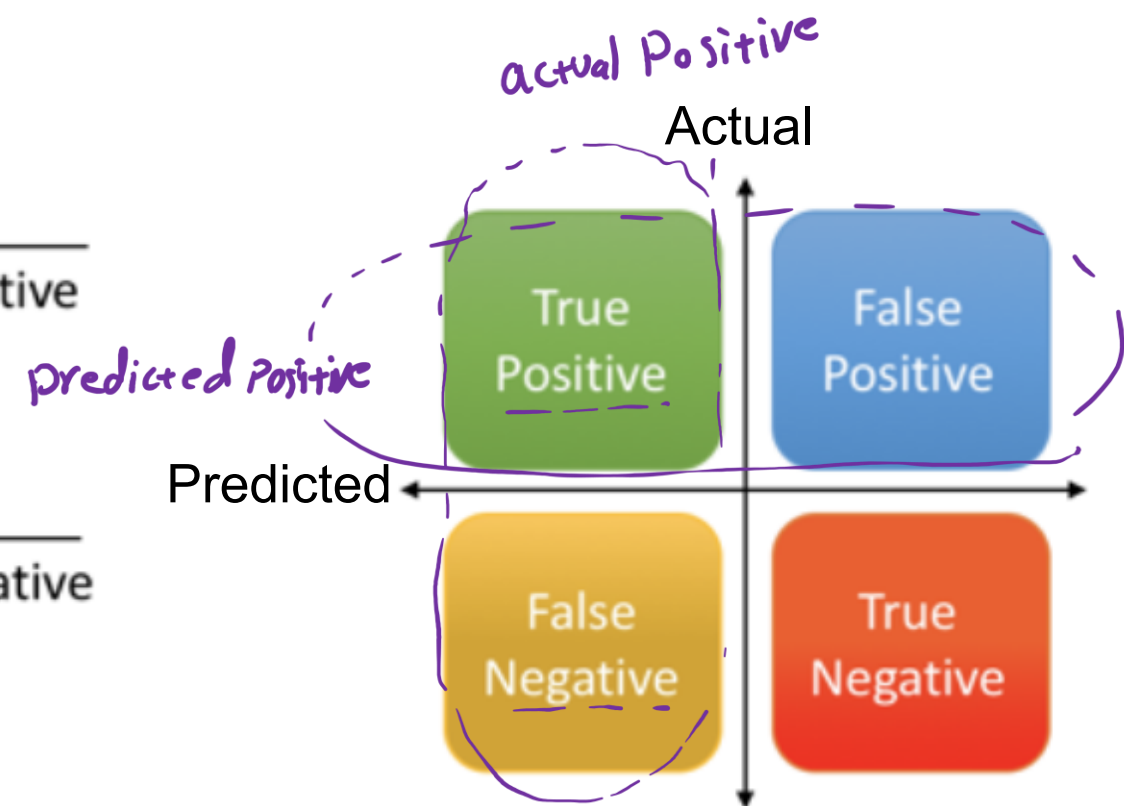
Number of predicted “positive” labeled data = True Positive + False Positive

Number of predicted “negative” labeled data = True Negative + False Negative

$$\text{Precision} = \frac{\text{True Positive}}{\text{Predicted Results}} \text{ or } \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{Actual Results}} \text{ or } \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total}}$$



False positive is also called false alarm

Matching-Based Measures (II): F-Measure

- **Precision**: which measure *quality*, is the same as purity:
 - How precisely does each cluster represent the ground truth?

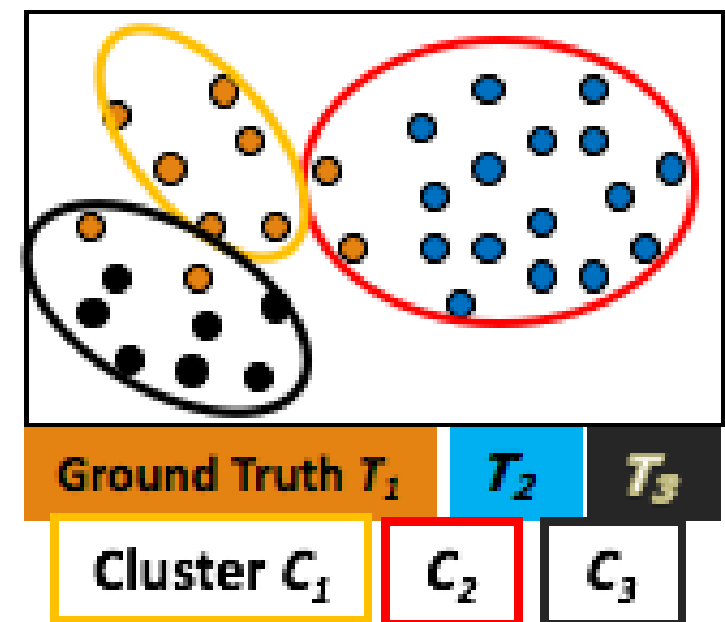
$$prec_i = \frac{1}{n_i} \max_{j=1}^k \{n_{ij}\} = \frac{n_{ij_i}}{n_i}$$

- **Recall**: measures completeness $recall_i = \frac{n_{ij_i}}{|T_{j_i}|} = \frac{n_{ij_i}}{m_{j_i}}$
 - How completely does each cluster recover the ground truth?

The Fraction of point in partition T_j shared in common with cluster C_i

$$Prec_1 = \frac{6}{6}$$

$$Recall_1 = \frac{6}{10}$$



Precision and Recall

(Precision here is same as the purity)

Precision:

$$\text{prec}_1 = 30/50;$$

$$\text{prec}_2 = 20/25;$$

$$\text{prec}_3 = 25/25$$

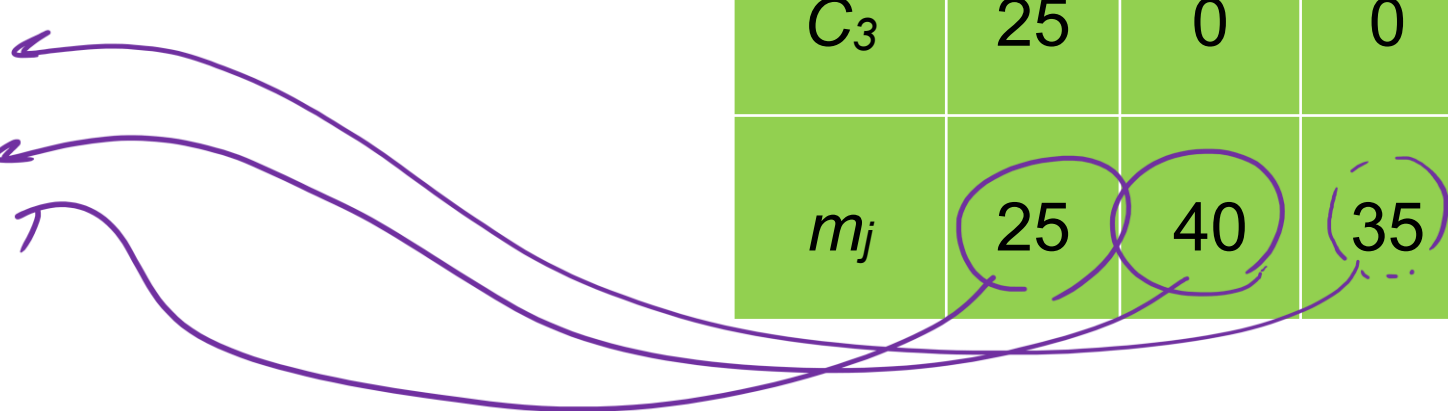
Recall:

$$\text{recall}_1 = 30/35;$$

$$\text{recall}_2 = 20/40;$$

$$\text{recall}_3 = 25/25$$

$C \backslash T$	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



Matching-Based Measures (II): F-Measure

- **F-Measure**: the harmonic mean of precision and recall
 - Take into account both *precision* and *completeness*

$$F_i = \frac{2}{\frac{1}{\text{prec}_i} + \frac{1}{\text{recall}_i}} = \frac{2 \cdot \text{prec}_i \cdot \text{recall}_i}{\text{prec}_i + \text{recall}_i} = \frac{2 n_{ij_i}}{n_i + m_{j_i}}$$

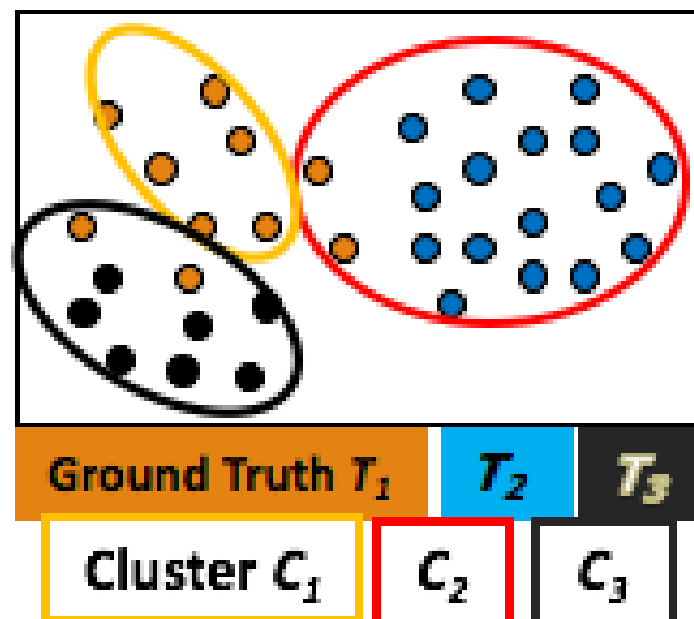
The F-measure for the clustering $\mathcal{C}$ is the mean of clusterwise F-measure values:

$$P(C_1) = \frac{50}{100} \quad P(T_1) = \frac{25}{100}$$

$$P(n_{23}^{C,T}) = \frac{5}{100} = \frac{n_{ij}}{n}$$

$$\begin{aligned} F_1 &= 60/85; \\ F_2 &= 40/65; \\ F_3 &= 1; \\ F &= 0.774 \end{aligned}$$

$$F = \frac{1}{r} \sum_{i=1}^r F_i$$



C \ 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

↓
m_j m₁ = 25

$$I(x) = \log_2 \frac{1}{p(x)} \quad E[I(x)] = H(x) = \sum p(x) I(x) = \sum p(x) \log_2 \frac{1}{p(x)} = - \sum p(x) \log_2 p(x)$$

Entropy-Based Measures (I): Conditional Entropy

Amount of information orderliness in different partitions

- The entropy for clustering C and partition T is

$$H(C) = - \sum_{i=1}^r p_{C_i} \log p_{C_i} \quad H(T) = - \sum_{j=1}^k p_{T_j} \log p_{T_j}$$

$$\text{where } p_{C_i} = \frac{n_i}{n} \text{ and } p_{T_j} = \frac{m_j}{n}$$

i.e., The probability of cluster C_i
 $n_i = n_{i1} + n_{i2} + \dots + n_{ik}$

i.e., The probability of ground truth T_j

$$p(T|C_i) = \frac{p(T, C_i)}{p(C_i)} = \frac{\frac{n_{ij}}{n}}{\frac{n_i}{n}} = \frac{n_{ij}}{n_i}$$

- Conditional Entropy:** The cluster-specific entropy, namely the conditional entropy of T with respect to cluster C_i :

$$H(T|C) = \begin{cases} H(T|C_1) \\ H(T|C_2) \\ H(T|C_3) \end{cases} \quad H(T|C_i) = - \sum_{j=1}^k \left(\frac{n_{ij}}{n_i} \right) \log \left(\frac{n_{ij}}{n_i} \right)$$

How ground truth is distributed within each cluster

n_{ij} ← Ground truth (T)
 Cluster (C)

Entropy-Based Measures (I): Conditional Entropy

- The conditional entropy of $\mathcal{T}$ given clustering $\mathcal{C}$ is defined as the weighted sum:

$$\begin{aligned} H(\mathcal{T}|\mathcal{C}) &= \sum_{i=1}^r \frac{n_i}{n} H(\mathcal{T}|\mathcal{C}_i) = - \sum_{i=1}^r \sum_{j=1}^k p_{ij} \log \left(\frac{p_{ij}}{p_{\mathcal{C}_i}} \right) \\ &= H(\mathcal{C}, \mathcal{T}) - H(\mathcal{C}) \end{aligned}$$

$\frac{p_{ij}}{p_{\mathcal{C}_i}} \rightarrow \frac{n_{ij}}{n}$
 $\frac{p_{ij}}{p_{\mathcal{C}_i}} \rightarrow \frac{n_i}{n}$

The more clusters members are split into different partitions, the higher the conditional entropy
(not a desirable condition and the max value is $\log k$)

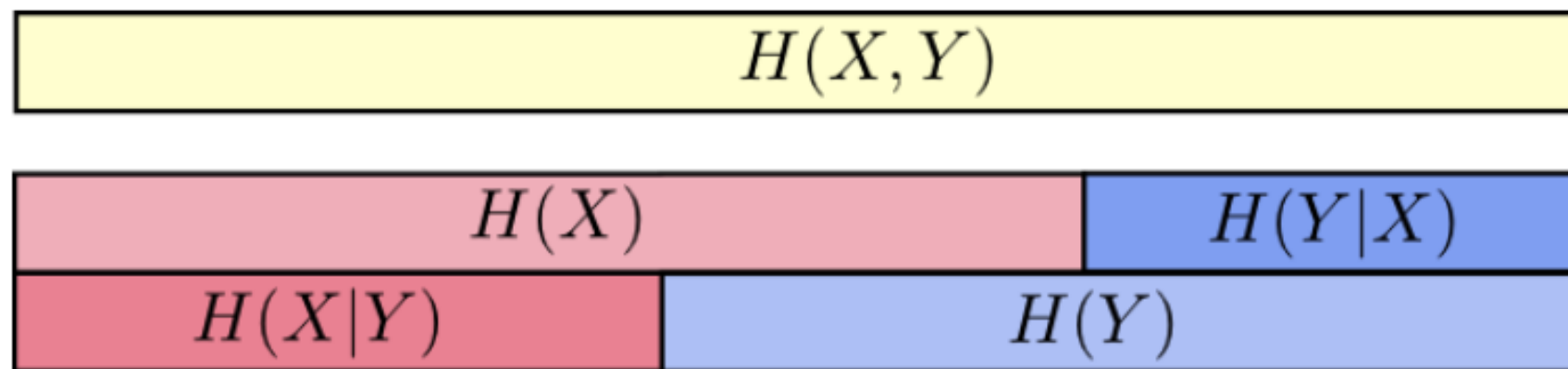
$H(\mathcal{T}|\mathcal{C}) = 0$ if and only if $\mathcal{T}$ is completely determined by $\mathcal{C}$, corresponding to the ideal clustering. If $\mathcal{C}$ and $\mathcal{T}$ are independent of each other, then $H(\mathcal{T}|\mathcal{C}) = H(\mathcal{T})$.

$$H(Y|X) = \sum_{x \in X} p(x) H(Y|X = x)$$

Fresh your memory:

$$H(Y|X) = H(X, Y) - H(X)$$

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



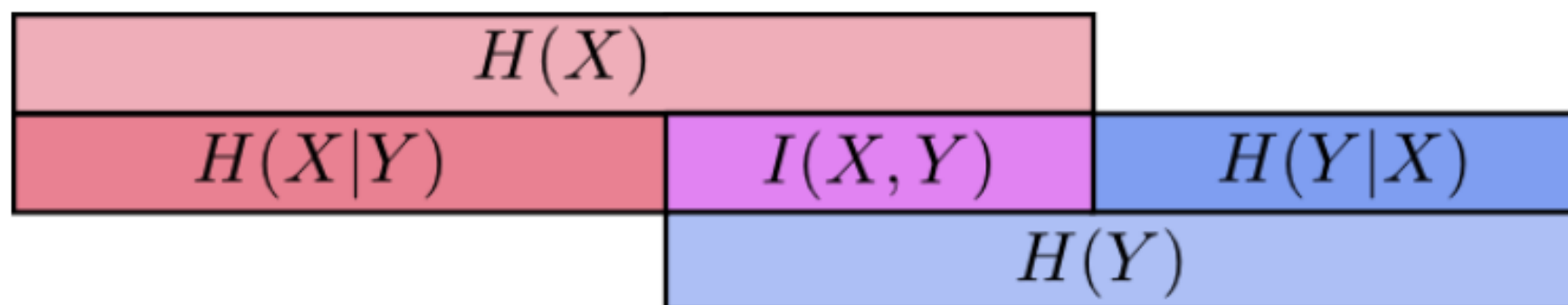
Entropy-Based Measures (I): Mutual Information

The *mutual information* tries to quantify the amount of shared information between the clustering $\mathcal{C}$ and partitioning $\mathcal{T}$, and it is defined as

$$H(X) \geq 0 \quad I(\mathcal{C}, \mathcal{T}) = \sum_{i=1}^r \sum_{j=1}^k p_{ij} \log \left(\frac{p_{ij}}{p_{C_i} \cdot p_{T_j}} \right) = H(\mathcal{T}) - H(\mathcal{T}|\mathcal{C})$$

When $\mathcal{C}$ and $\mathcal{T}$ are independent then $p_{ij} = p_{C_i} \cdot p_{T_j}$, and thus $I(\mathcal{C}, \mathcal{T}) = 0$. However, there is no upper bound on the mutual information.

We should do something about this



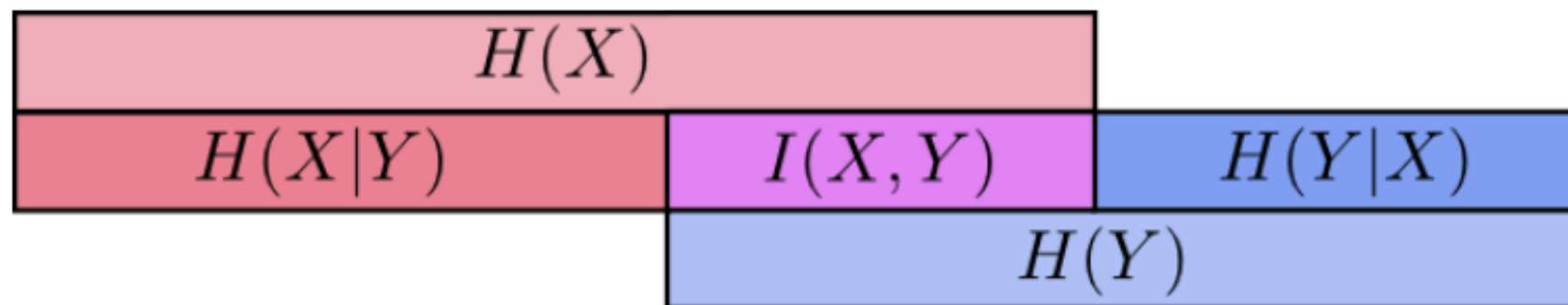
We measure the dependency between the observed joint probability p_{ij} of $\mathcal{C}$ and $\mathcal{T}$, and the expected joint probability $p_{ci} \cdot p_{Tj}$ under the independence assumption

Entropy-Based Measures (I): Mutual Information

The *normalized mutual information* (NMI) is defined as the geometric mean:

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

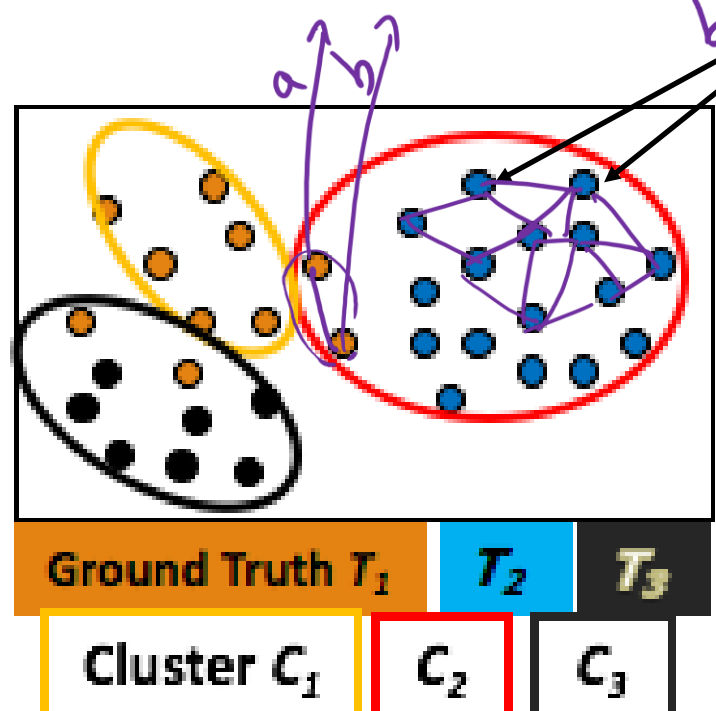
The NMI value lies in the range $[0, 1]$. Values close to 1 indicate a good clustering.



$\binom{n}{2} = \frac{n(n-1)}{2}$ Pairwise Measures

Given clustering $\mathcal{C}$ and ground-truth partitioning $\mathcal{T}$, let $\mathbf{x}_i, \mathbf{x}_j \in \mathbf{D}$ be any two points, with $i \neq j$. Let y_i denote the true partition label and let $\hat{y}_i$ denote the cluster label for point $\mathbf{x}_i$.

True Positives: $\mathbf{x}_i$ and $\mathbf{x}_j$ belong to the same partition in $\mathcal{T}$, and they are also in the same cluster in $\mathcal{C}$. The number of true positive pairs is given as



$$TP = |\{(\mathbf{x}_i, \mathbf{x}_j) : y_i = y_j \text{ and } \hat{y}_i = \hat{y}_j\}|$$

$n_{23} = 5$

Same partition Same cluster

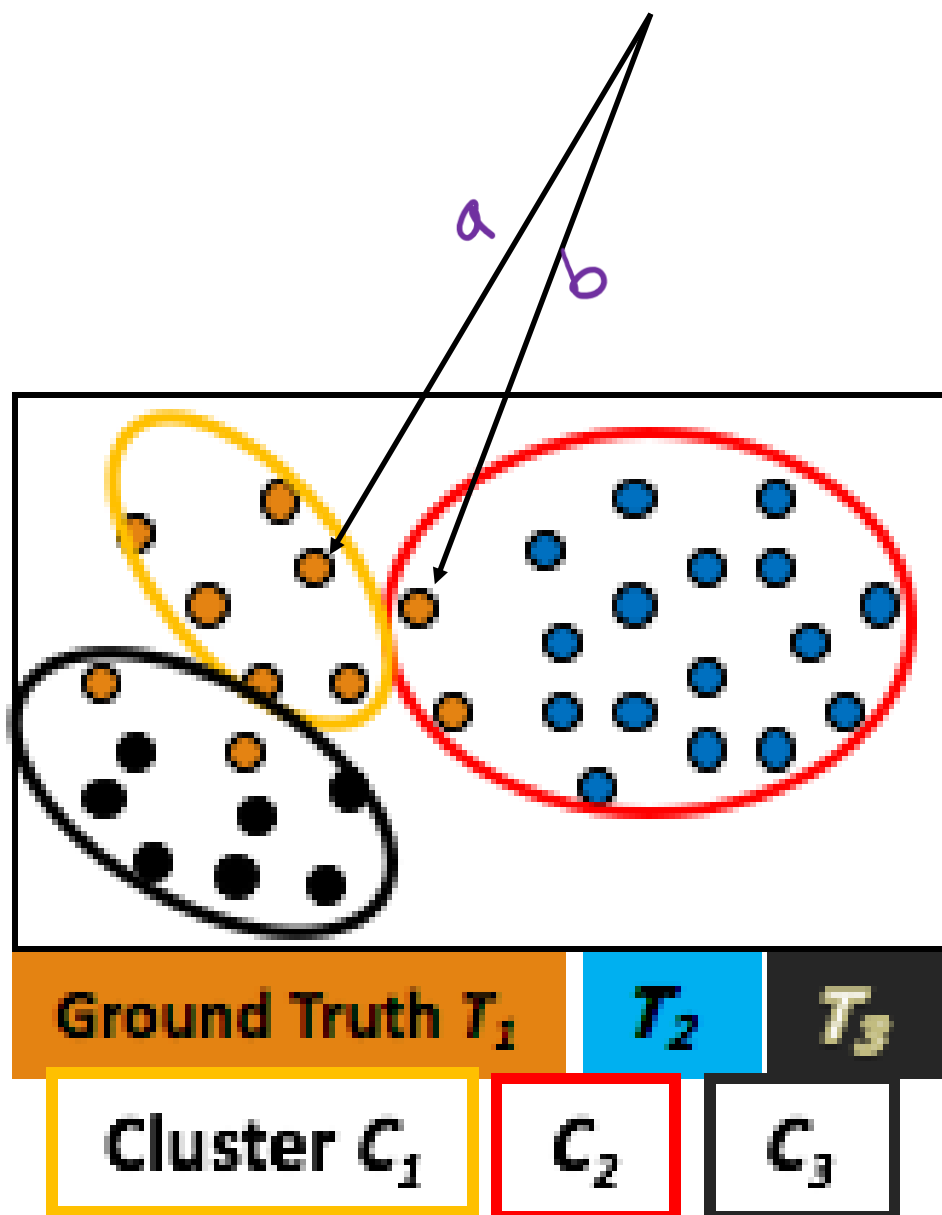
$\binom{0}{2} + \binom{20}{2} + \binom{30}{2} + \binom{0}{2} + \binom{20}{2}$
 $+ \binom{5}{2} + \binom{25}{2} + \binom{0}{2} + \binom{0}{2}$

$\mathcal{C} \backslash \mathcal{T}$	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

False Negatives: $\mathbf{x}_i$ and $\mathbf{x}_j$ belong to the same partition in $\mathcal{T}$, but they do not belong to the same cluster in $\mathcal{C}$. The number of all false negative pairs is given as

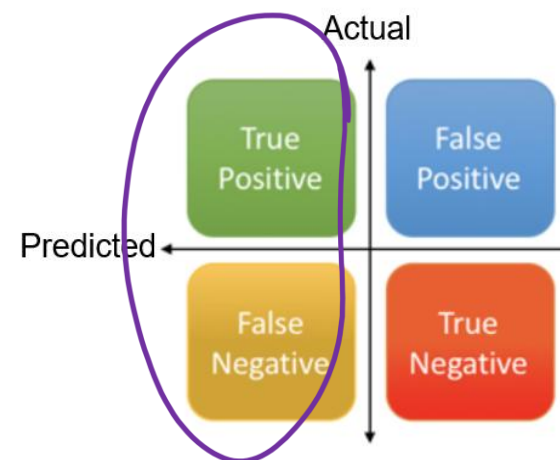
$$FN = |\{(\mathbf{x}_i, \mathbf{x}_j) : y_i = y_j \text{ and } \hat{y}_i \neq \hat{y}_j\}|$$

Same partition Different cluster



$(TP) + (FN) = \text{Actual Positive Results}$

$$\binom{25}{2} + \binom{40}{2} + \binom{35}{2}$$

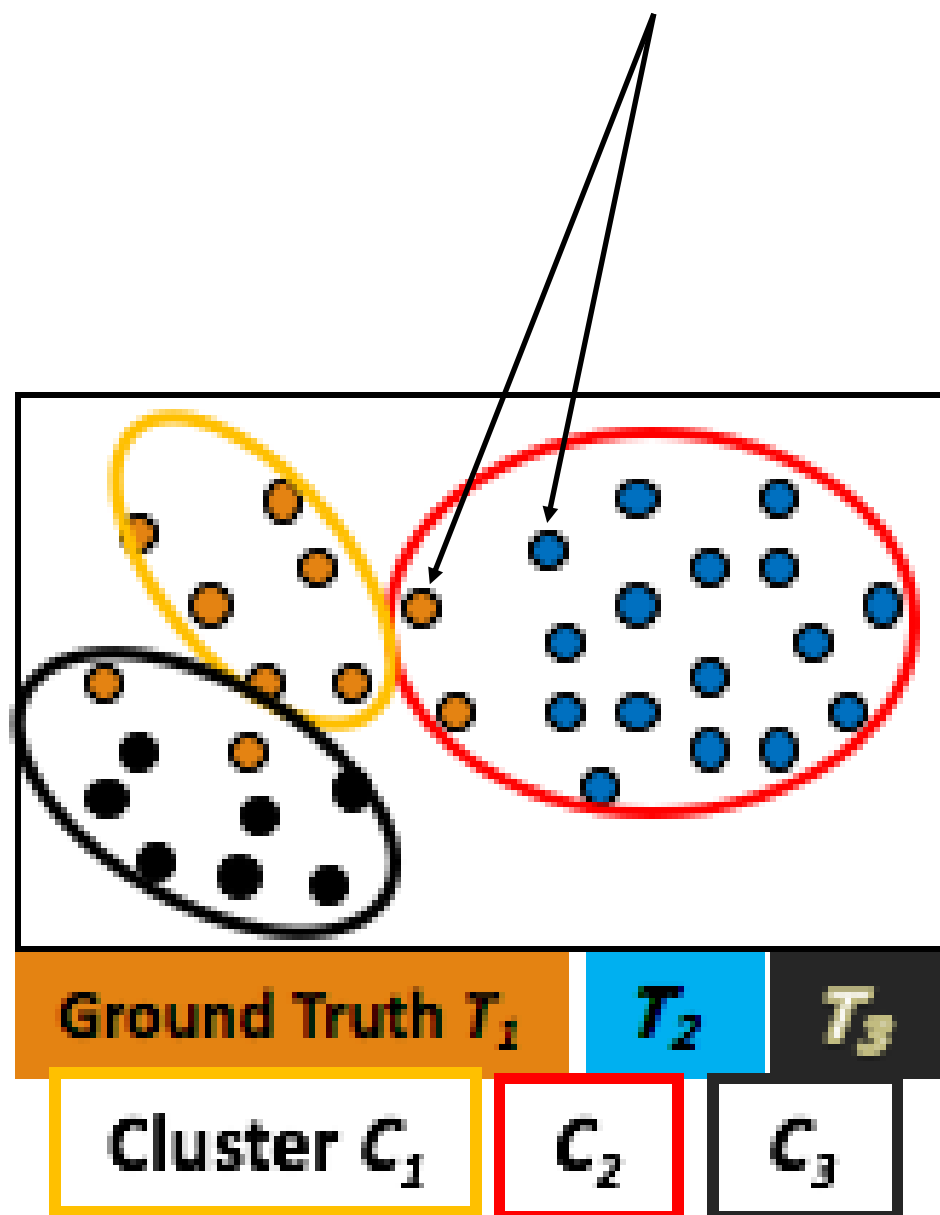


C \ T	T ₁	T ₂	T ₃	Sum
C ₁	0	20	30	50
C ₂	0	20	5	25
C ₃	25	0	0	25
m _j	25	40	35	100

False Positives: $\mathbf{x}_i$ and $\mathbf{x}_j$ do not belong to the same partition in $\mathcal{T}$, but they do belong to the same cluster in $\mathcal{C}$. The number of false positive pairs is given as

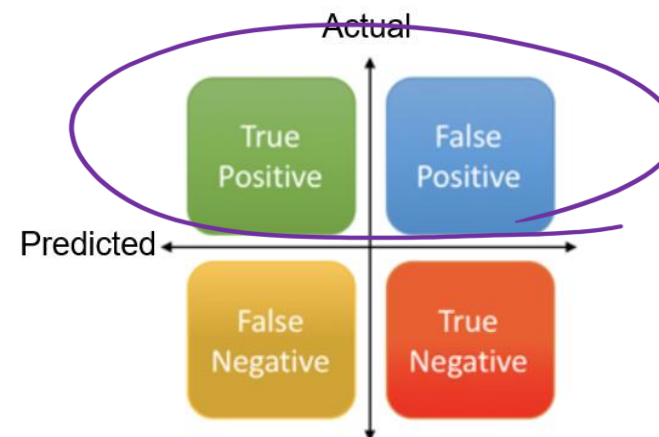
$$FP = |\{(\mathbf{x}_i, \mathbf{x}_j) : y_i \neq y_j \text{ and } \hat{y}_i = \hat{y}_j\}|$$

Different partition Same cluster



$(TP) + FP = \text{Predicted Positive results}$

$\checkmark + FP = \binom{50}{2} + \binom{25}{2} + \binom{25}{2}$



$C \setminus T$	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

True Negatives: $\mathbf{x}_i$ and $\mathbf{x}_j$ neither belong to the same partition in $\mathcal{T}$, nor do they belong to the same cluster in $\mathcal{C}$. The number of such true negative pairs is given as

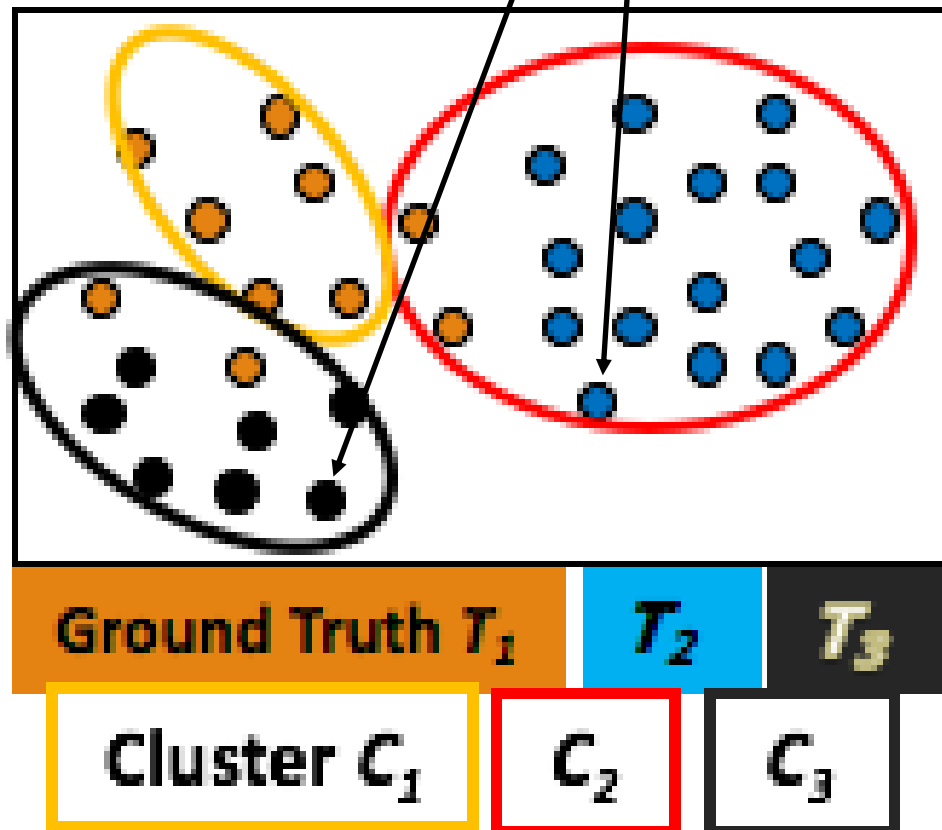
$$N_{\text{edges}} = \binom{N}{2} = \binom{100}{2}$$

$$TN = |\{(\mathbf{x}_i, \mathbf{x}_j) : y_i \neq y_j \text{ and } \hat{y}_i \neq \hat{y}_j\}|$$

Different partition

Different cluster

$$N_{\text{edges}} = TP + FP + FN + TN$$



Pairwise Measures

Because there are $N = \binom{n}{2} = \frac{n(n-1)}{2}$ pairs of points, we have the following identity:

$$N = TP + FN + FP + TN$$

$$TP = \sum_{i=1}^r \sum_{j=1}^k \binom{n_{ij}}{2} = \frac{1}{2} \left(\sum_{i=1}^r \sum_{j=1}^k (n_{ij}^2 - n_{ij}) \right) = \frac{1}{2} \left(\left(\sum_{i=1}^r \sum_{j=1}^k n_{ij}^2 \right) - n \right)$$

$$FN = \sum_{j=1}^k \binom{m_j}{2} - TP$$

$$FP = \sum_{i=1}^r \binom{n_i}{2} - TP$$

$$TN = N - (TP + FN + FP)$$

$C \backslash T$	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

$$FP = \sum_{i=1}^r \binom{n_i}{2} - TP$$

$n_{12} = 20$ Points which have same Cluster one and same Partition two

Pairwise Measures

Jaccard Coefficient: measures the fraction of true positive point pairs, but after ignoring the true negative:

$$Jaccard = \frac{TP}{TP + FN + FP}$$

Perfect clustering = 1

Rand Statistic: measures the fraction of true positives and true negatives over all point pairs:

$$Rand = \frac{TP + TN}{N}$$

Perfect clustering = 1 (like accuracy)

Fowlkes-Mallows Measure: Define the overall *pairwise precision* and *pairwise recall* values for a clustering $\mathcal{C}$, as follows:

$$prec = TP / (TP + FP)$$

$$recall = TP / (TP + FN)$$

The Fowlkes–Mallows (FM) measure is defined as the geometric mean of the pairwise precision and recall

$$FM = \sqrt{prec \cdot recall} = \frac{TP}{\sqrt{(TP + FN)(TP + FP)}}$$

Higher value means a better clustering

Outline

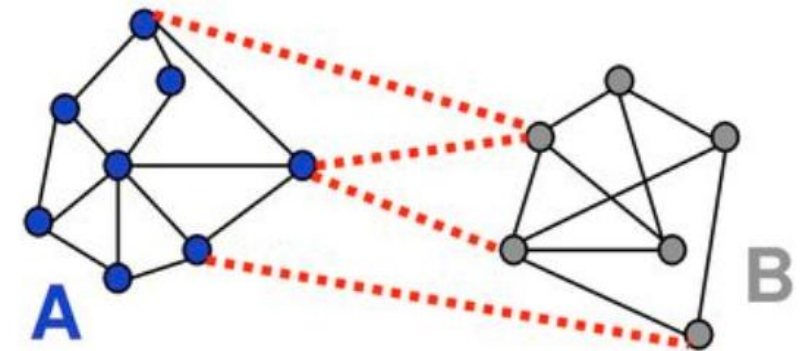
- External measures for clustering evaluation
 - Matching-based measures
 - Entropy-based measures
 - Pairwise measures
- Internal measures for clustering evaluation 
 - Graph-based measures
 - Davies-Bouldin Index
 - Silhouette Coefficient

We want intra-cluster datapoints to be as close as possible to each other and inter-clusters to be as far as possible from each other

The Beta-CV Measure

- Let W be the pair-wise distance matrix for all the given points. For any two point sets S and R , we define:

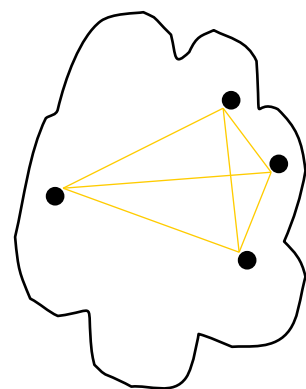
$$W(S, R) = \sum_{\mathbf{x}_i \in S} \sum_{\mathbf{x}_j \in R} w_{ij}$$



The sum of all the intracluster and intercluster weights are given as

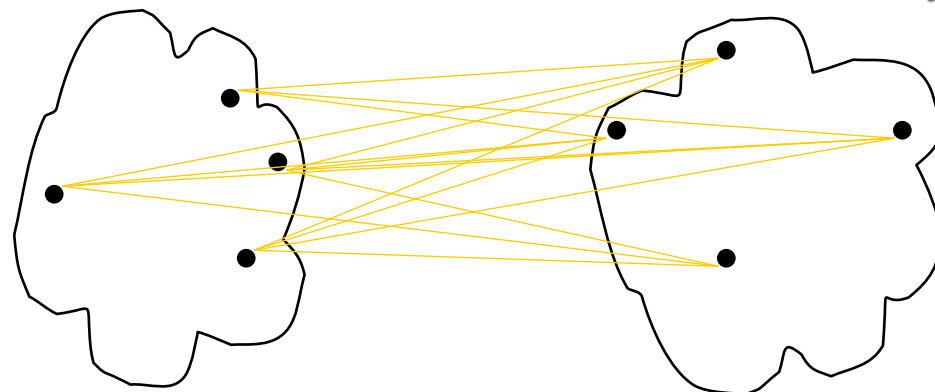
$$W_{in} = \frac{1}{2} \sum_{i=1}^k W(C_i, C_i)$$

The distance of each point is measured two times



cohesion

$$W_{out} = \frac{1}{2} \sum_{i=1}^k W(C_i, \bar{C}_i) = \sum_{i=1}^{k-1} \sum_{j>i} W(C_i, C_j)$$



separation

The Beta-CV Measure

The number of distinct intracluster and intercluster edges is given as

$$N_{in} = \sum_{i=1}^k \binom{n_i}{2} \qquad N_{out} = \sum_{i=1}^{k-1} \sum_{j=i+1}^k n_i \cdot n_j$$

BetaCV Measure: The BetaCV measure is the ratio of the mean intracluster distance to the mean intercluster distance:

$$BetaCV = \frac{W_{in}/N_{in}}{W_{out}/N_{out}} = \frac{N_{out}}{N_{in}} \cdot \frac{W_{in}}{W_{out}} = \frac{N_{out}}{N_{in}} \frac{\sum_{i=1}^k W(C_i, C_i)}{\sum_{i=1}^k W(C_i, \overline{C_i})}$$

The smaller the BetaCV ratio, the better the clustering.

Normalized Cut

Normalized cut:
$$NC = \sum_{i=1}^k \frac{W(C_i, \bar{C}_i)}{vol(C_i)} = \sum_{i=1}^k \frac{W(C_i, \bar{C}_i)}{W(C_i, V)} = \sum_{i=1}^k \frac{W(C_i, \bar{C}_i)}{W(C_i, C_i) + W(C_i, \bar{C}_i)} = \sum_{i=1}^k \frac{1}{\frac{W(C_i, C_i)}{W(C_i, \bar{C}_i)} + 1}$$

where $vol(C_i) = W(C_i, V)$ is the volume of cluster C_i

The higher normalized cut value, the better the clustering

$W(C_i, C_i)$



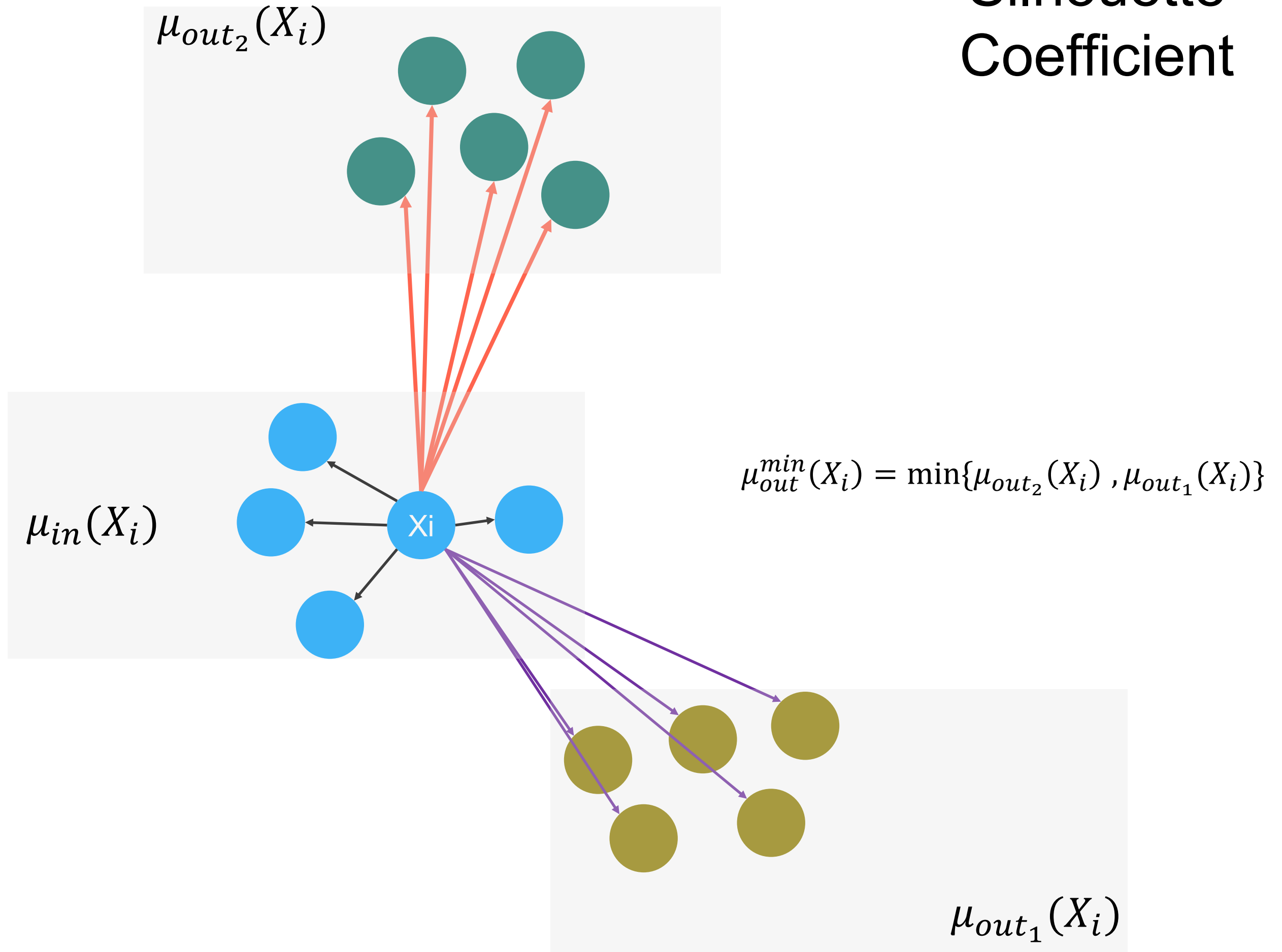
Intra-cluster distance

$W(C_i, \bar{C}_i)$



Inter-cluster distance

Silhouette Coefficient



Silhouette Coefficient

Define the silhouette coefficient of a point $\mathbf{x}_i$ as

$$s_i = \frac{\mu_{out}^{\min}(\mathbf{x}_i) - \mu_{in}(\mathbf{x}_i)}{\max\{\mu_{out}^{\min}(\mathbf{x}_i), \mu_{in}(\mathbf{x}_i)\}}$$

where $\mu_{in}(\mathbf{x}_i)$ is the mean distance from $\mathbf{x}_i$ to points in its own cluster $\hat{y}_i$:

$$\mu_{in}(\mathbf{x}_i) = \frac{\sum_{\mathbf{x}_j \in C_{\hat{y}_i}, j \neq i} \delta(\mathbf{x}_i, \mathbf{x}_j)}{n_{\hat{y}_i} - 1}$$

and $\mu_{out}^{\min}(\mathbf{x}_i)$ is the mean of the distances from $\mathbf{x}_i$ to points in the closest cluster:

$$\mu_{out}^{\min}(\mathbf{x}_i) = \min_{j \neq \hat{y}_i} \left\{ \frac{\sum_{\mathbf{y} \in C_j} \delta(\mathbf{x}_i, \mathbf{y})}{n_j} \right\}$$

The Silhouette Coefficient for clustering C: $SC = \frac{1}{n} \sum_{i=1}^n s_i$.

SC close to 1 implies a good clustering (Points are close to their own clusters but far from other clusters)

The Davies-Bouldin Index

Let μ_i denote the cluster mean

$$\mu_i = \frac{1}{n_i} \sum_{\mathbf{x}_j \in C_i} \mathbf{x}_j$$

Let σ_{μ_i} denote the dispersion or spread of the points around the cluster mean

$$\sigma_{\mu_i} = \sqrt{\frac{\sum_{\mathbf{x}_j \in C_i} \delta(\mathbf{x}_j, \mu_i)^2}{n_i}} = \sqrt{\text{var}(C_i)}$$

The Davies–Bouldin measure for a pair of clusters C_i and C_j is defined as the ratio

Calculate the DB of i cluster from other clusters

$$DB_{ij} = \frac{\sigma_{\mu_i} + \sigma_{\mu_j}}{\delta(\mu_i, \mu_j)} \quad D_i = \max_{i \neq j} DB_{ij}$$

DB_{ij} measures how compact the clusters are compared to the distance between the cluster means. The Davies–Bouldin index is then defined as

$$DB = \frac{1}{k} \sum_{i=1}^k D_i$$

a lower value means that the clustering is better

Summary

- External measures for clustering evaluation
 - Matching-based measures
 - Entropy-based measures
 - Pairwise measures
- Internal measures for clustering evaluation
 - Graph-based measures
 - Davies-Bouldin Index
 - Silhouette Coefficient