CS724: Topics in Algorithms
Evaluation of Clustering Quality

Prof. Dan A. Simovici

The quality of clusterings can be evaluated using two types of criteria:

- criteria unrelated to the data set that is subjected to clustering (external criteria);
- criteria that are derived from the data set (internal criteria)

The Ground Truth

External validation criteria are useful when a "ground truth" is known (as it is typically the case for classification problems) and we seek to evaluate the appropriateness of a clustering algorithm for separating objects into clusters that conform more or less to the existing classification.
The ground truth is captured by a *reference partition* $\theta = \{T_1, \ldots, T_r\}$ of data set D (also known as the *ground-truth* partition). We discuss modalities of comparing a clustering $\kappa = (C_1, \ldots, C_m)$ with the ground truth partition.

Definition

The *contingency matrix* of θ and κ is the matrix $G(\theta, \kappa) \in \mathbb{R}^{r \times m}$, where $g_{ij} = |T_i \cap C_j|$ for $1 \leqslant i \leqslant r$ and $1 \leqslant j \leqslant m$.

UMASS BOSTON

Suppose that the classes of objects of a data set D relative to the partitions θ and κ are described respectively by the **R** -vectors `t` and `k` whose length is $n = |D|$. Then, the contingency matrix $G(\theta, \kappa)$ of partitions θ and κ can be obtained by using `table(t,k)`.

Example

Let D be a data set with $|D| = 12$ and let θ and κ be two partitions of D:

$$\begin{aligned} \theta &= \{\{d_1, d_6, d_{10}\}, \{d_3, d_4, d_7, d_8, d_{12}\}, \{d_2, d_5, d_9, d_{11}\}\}, \\ \kappa &= \{\{d_4, d_6, d_{10}, d_{12}\}, \{d_1, d_3, d_8\}, \{d_2, d_5, d_7, d_9, d_{11}\}\}. \end{aligned}$$

The **R** -vectors that describe these partitions are:

```
t <- c(1,3,2,2,3,1,2,2,3,1,3,2)
k <- c(2,3,2,1,3,1,3,2,3,1,3,1).
```

A call to the function `table` returns the contingency table of partitions θ and κ:

```
> table(t,k)
   k
t   1 2 3
  1 2 1 0
  2 2 2 1
  3 0 0 4
```

Example

The row sums of the matrix equal the sizes of blocks of θ, while the column sums equal the sizes of blocks of κ.
For a contingency matrix $G = (g_{ij}) \in \mathbb{R}^{r\times m}$ for the reference partition $\theta = \{T_1, \ldots, T_r\}$ and clustering $\kappa = \{C_1, \ldots, C_m\}$ we introduce the notations:

$$
\begin{aligned}
n_{i\cdot} &= \sum_{j=1}^{m} g_{ij} = |T_i|, \\
n_{\cdot j} &= \sum_{i=1}^{r} g_{ij} = |C_j|, \\
n_{\cdot\cdot} &= \sum_{i=1}^{r} \sum_{j=1}^{m} g_{ij} = |D|.
\end{aligned}
$$

UMASS BOSTON

A cluster C_j is *θ-pure* if it is included in a block T_i of the reference partition θ.

We denote by T_{i_j} the largest block of the reference partition θ that has the largest intersection with the cluster C_j.

Definition

The *precision* of a cluster C_j is defined as

$$\text{precision}_\theta(C_j) = \frac{1}{|C_j|} \cdot \max\{|C_j \cap T_i| \mid 1 \leqslant i \leqslant r\},$$

and it measures the largest fraction of the cluster in a block of the reference partition.

The *precision* of the clustering κ is the average precision of the clusters $C_1, \ldots, C_m$, that is,

$$\begin{aligned} \text{precision}_\theta(\kappa) &= \sum_{j=1}^{m} \frac{|C_j|}{|D|} \text{precision}_\theta(C_j) \\ &= \frac{1}{|D|} \sum_{j=1}^{m} \max\{|T_i \cap C_j| \mid 1 \leqslant i \leqslant r\}. \end{aligned}$$

If all clusters of κ are pure, then $\text{precision}_\theta(\kappa) = 1$.

Definition

The *recall* of the cluster C_j is defined as

$$\text{recall}_\theta(C_j) = \frac{1}{|T_{i_j}|}|C_j \cap T_{i_j}|$$

and measures the fraction of the largest reference block that has the largest intersection with C_j which is shared with C_j.

The *F-measure* of the cluster C_j is the harmonic average of its precision and recall:

$$F(C_j) = \frac{2}{\frac{1}{\text{precision}_\theta(C_j)} + \frac{1}{\text{recall}_\theta(C_j)}} = 2\frac{n_{i_j\,j}}{n_j + |T_{i_j}|}$$

The F-measure $F(\kappa)$ for the clustering κ is the mean of the F-measures for the clusters:

$$F(\kappa) = \frac{1}{m} \sum_{j=1}^{m} F(C_j).$$

Higher values for the F-measure indicate a better fit between the reference partition θ and the clustering κ.

Example

Let θ and κ be the partitions introduced in above, where $\theta = \{T_1, T_2, T_3\}$, $\kappa = \{C_1, C_2, C_3\}$ and

$$\begin{array}{ll} T_1 = \{d_1, d_6, d_{10}\} & C_1 = \{d_4, d_6, d_{10}, d_{12}\}, \\ T_2 = \{d_3, d_4, d_7, d_8, d_{12}\} & C_2 = \{d_1, d_3, d_8\}, \\ T_3 = \{d_2, d_5, d_9, d_{11}\} & C_3 = \{d_2, d_5, d_7, d_9, d_{11}\}. \end{array}$$

Note that contingency matrix $G(\theta, \sigma)$ can be written as

$$G = \begin{pmatrix} |T_1 \cap C_1| & |T_1 \cap C_2| & |T_1 \cap C_3| \\ |T_2 \cap C_1| & |T_2 \cap C_2| & |T_2 \cap C_3| \\ |T_3 \cap C_1| & |T_3 \cap C_2| & |T_3 \cap C_3| \end{pmatrix} = \begin{pmatrix} 2 & 1 & 0 \\ 2 & 2 & 1 \\ 0 & 0 & 4 \end{pmatrix}$$

Thus, the blocks of the reference partitions that have the largest intersection with the clusters C_1, C_2 and C_3 are T_2, again T_2 and T_3, respectively.

Example

The precision of the clusters of κ relative to θ are

$$\text{precision}_\theta(C_1) = \frac{2}{4}, \text{precision}_\theta(C_2) = \frac{2}{3}, \text{precision}_\theta(C_3) = \frac{4}{5},$$

so C_3 has the largest precision.
The precision of κ is

$$\begin{aligned}\text{precision}_\theta(\kappa) &= \sum_{j=1}^{m} \frac{|C_j|}{|D|}\text{precision}_\theta(C_j) \\ &= \frac{4}{12}\frac{2}{4} + \frac{3}{12}\frac{2}{3} + \frac{5}{12}\frac{4}{5} = \frac{2}{3}.\end{aligned}$$

The recalls of the clusters are

$$\text{recall}_\theta(C_1) = \frac{2}{5}, \text{recall}_\theta(C_2) = \frac{2}{5}, \text{recall}_\theta(C_3) = \frac{4}{4}.$$

Example

The recalls of the clusters are

$$\text{recall}_\theta(C_1) = \frac{2}{5}, \text{recall}_\theta(C_2) = \frac{2}{5}, \text{recall}_\theta(C_3) = \frac{4}{4}.$$

The F-score of C_1 is

$$F(C_1) = \frac{2\text{precision}(C_1) \cdot \text{recall}(C_1)}{\text{precision}(C_1) + \text{recall}(C_1)} = \frac{4}{9}.$$

Similarly, $F(C_2) = \frac{1}{2}$ and $F(C_3) = \frac{8}{9}$. The F-score for the cluster κ is the average of these scores, that is, $\frac{1}{3}(\frac{4}{9} + \frac{1}{2} + \frac{8}{9}) = \frac{11}{18}$.
A good score is usually close to 1.

Example

Consider the first two principal components of the objects in the `iris` data set and the projection of this data set on these components.
We begin by converting the class of the items in `iris` into numerical values by using the function `numClass`:

```
numClass <- function(){
   result <- vector(length=150)
   for(i in 1:150)
    if(iris[i,5]==''setosa'') result[i]=1
      else if (iris[i,5]==''versicolor'') result[i]=2
      else result[i]=3
return(result)
}
```

Example

The numerical class is saved in N with

```
N <- numClass()
```

Next, we represent the objects in the iris data set using the first two principal components by writing

```
> pcaIRIS <- PCA(iris[,1:4],graph=FALSE)
> M <- pcaIRIS$ind$coord[,1:2]
```

Recall that PCA is a function of the package FactMineR.
The k-means algorithm for $k = 3$ is applied to M:

```
> K1 <- kmeans(M,3)
```

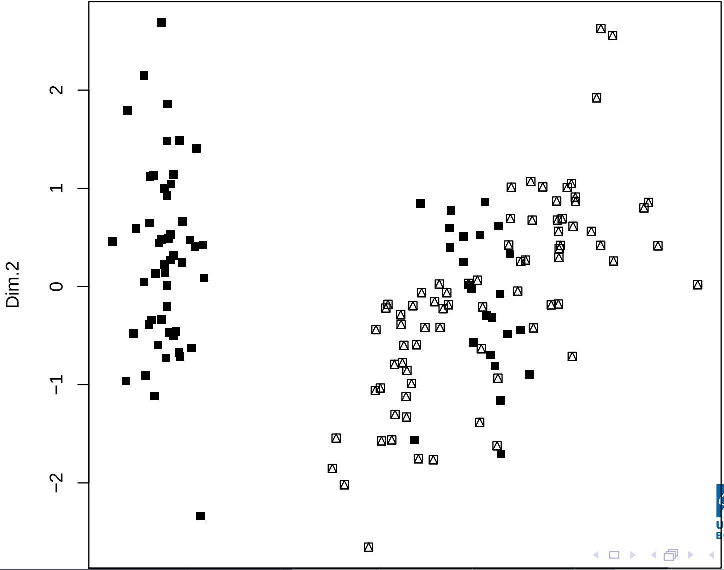
and the component cluster of K1 is compared to N. In case of equality, the object was placed in the correct cluster (and appears as a black rectangle); otherwise, the object was missclassified.

Example

The corresponding representation is obtained by writing:

```
> pdf(''pcaIRIS.pdf'')
> plot(M,pch = 14 + (K1$cluster==N))
> dev.off()
```

Representation of the `iris` data set



A contingency table can now be created by writing

```
> table(N,K1$cluster)
```

This produces

```
     1  2  3
  1 50  0  0
  2  0 11 39
  3  0 36 14
```

The first class `setosa` is perfectly classified; only 11 of the fifty plants of the `versicolor`, and only 14 of the fifty `virginica` are correctly placed.

- The notion of entropy is a probabilistic concept that lies at the foundation of information theory.
- Our goal is to define entropy in an algebraic setting by introducing the notion of entropy of a partition of a finite set.
- Entropy will allow us to compare clusterings regarded as partitions of finite sets.

From a probabilistic point of view Shannon's entropy is defined starting from a probability distribution $\mathbf{p} = (p_1, \ldots, p_n)$ with $p_i \geqslant 0$ for $1 \leqslant i \leqslant n$ and $\sum_{i=1}^n p_i = 1$, as $\mathcal{H}(\mathbf{p}) = \sum_{i=1}^n p_i \log_2 \frac{1}{p_i}$.
Since the function $f : (0, \infty) \longrightarrow \mathbb{R}$ given by $f(x) = -\log_2 x$ is convex by Jensen's inequality we have:

$$-\log_2 \left(\sum_{i=1}^n p_i x_i \right) \leqslant -\sum_{i=1}^n p_i \log_2 x_i.$$

If $x_i = \frac{1}{p_i}$, we obtain

$$-\log_2 n \leqslant -\sum_{i=1}^n p_i \log_2 \frac{1}{p_i}.$$

Thus $\mathcal{H}(p_1, \ldots, p_n) \leqslant \log_2 n$, which shows that the maximum Shannon entropy, $\log_2 n$ is obtained when $p_1 = \cdots = p_n = \frac{1}{n}$.

Definition

Let S be a finite set and let $\pi = \{B_1, \ldots, B_m\}$ be a partition of S. The *β-entropy of a partition* $\pi \in PART(S)$ is the number $\mathcal{H}_\beta(S, \pi)$ defined as

$$\mathcal{H}_\beta(S, \pi) = \frac{1}{1 - 2^{1-\beta}} \cdot \left(1 - \sum_{i=1}^{m} \left(\frac{|B_i|}{|S|}\right)^\beta\right).$$

When S is clear from context we write $\mathcal{H}_\beta(\pi)$ instead of $\mathcal{H}_\beta(S, \pi)$.

The *Shannon entropy of* π is the number

$$\mathcal{H}(\pi) = -\sum_{i=1}^{m} \frac{|B_i|}{|S|} \log_2 \frac{|B_i|}{|S|}.$$

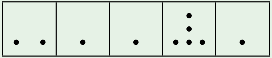
The *Gini index of* π is the number

$$gini(\pi) = 1 - \sum_{i=1}^{m} \left(\frac{|B_i|}{|S|} \right)^2.$$

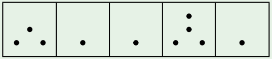
Both the Shannon entropy and the Gini index can be used to evaluate the uniformity of the distribution of the elements of S in the blocks of π because both values increase with the uniformity of the distribution of the elements of S.

Example

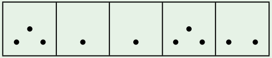
Entropy increasing with partition uniformity:



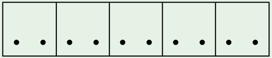
$\mathcal{H}(\pi_4) = 1.96$



$\mathcal{H}(\pi_3) = 2.04$

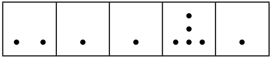


$\mathcal{H}(\pi_2) = 2.17$

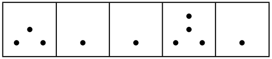


$\mathcal{H}(\pi_1) = 2.32$

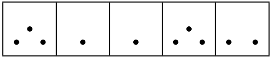
Gini index increasing with partition uniformity:



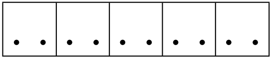
$gini(\pi_4) = 0.68$



$gini(\pi_3) = 0.72$



$gini(\pi_2) = 0.79$



$gini(\pi_1) = 0.80$

If $\beta = 2$, we obtain $\mathcal{H}_2(\pi)$, which is twice the Gini index,

$$\mathcal{H}_\beta(S, \pi) = 2 \cdot \left(1 - \sum_{i=1}^{m} \left(\frac{|B_i|}{|S|}\right)^2\right).$$

The *Gini index*, $gini(\pi) = 1 - \sum_{i=1}^{m} \left(\frac{|B_i|}{|S|}\right)^2$, is widely used in machine learning and data mining.

The limit case, $\lim_{\beta\to 1}\mathcal{H}_\beta(\pi)$, yields

$$\begin{aligned}\lim_{\beta\to 1}\mathcal{H}_\beta(S,\pi) &= \lim_{\beta\to 1}\frac{1}{1-2^{1-\beta}}\cdot\left(1-\sum_{i=1}^{m}\left(\frac{|B_i|}{|S|}\right)^{\beta}\right)\\ &= \lim_{\beta\to 1}\frac{1}{2^{1-\beta}\ln 2}\cdot\left(-\sum_{i=1}^{m}\left(\frac{|B_i|}{|S|}\right)^{\beta}\ln\frac{|B_i|}{|S|}\right)\\ &= -\sum_{i=1}^{m}\frac{|B_i|}{|S|}\log_2\frac{|B_i|}{|S|},\end{aligned}$$

which is the Shannon entropy of π.

Definition

Let S and T be disjoint sets, and let $\pi = \{B_1, \ldots, B_m\} \in PART(S)$ and $\sigma = \{C_1, \ldots, C_n\} \in PART(T)$. The *sum of the partitions* π and σ is the partition $\pi + \sigma$ of $S \cup T$ given by

$$\pi + \sigma = \{B_1, \ldots, B_m, C_1, \ldots, C_n\}$$

Theorem

Let $\beta \geqslant 1$. The following properties hold:

(i) *If $\pi, \pi' \in PART(S)$ and $\pi \leqslant \pi'$, then $\mathcal{H}_\beta(\pi') \leqslant \mathcal{H}(\pi)$.*

(ii) *If S, T are finite sets and $|S| \leqslant |T|$, then $\mathcal{H}_\beta(S, \alpha_S) \leqslant \mathcal{H}_\beta(T, \alpha_T)$.*

(iii) *If S and T are disjoint sets, $\pi \in PART(S)$ and $\sigma \in PART(T)$, then*

$$\begin{aligned} &\mathcal{H}_\beta(S \cup T, \pi + \sigma) \\ &= \left(\frac{|S|}{|S| + |T|}\right)^\beta \mathcal{H}_\beta(S, \pi) + \left(\frac{|T|}{|S| + |T|}\right)^\beta \mathcal{H}_\beta(T, \sigma) \\ &\quad + \mathcal{H}_\beta(S \cup T, \{S, T\}). \end{aligned}$$

Proof

To prove Part (i) it suffices to show that $\pi \prec \pi'$ implies $\mathcal{H}_\beta(\pi') \leqslant \mathcal{H}(\pi)$. Therefore, we may assume that $\pi = \{B_1, \ldots, B_{m-1}, B_m\}$ and $\pi' = \{B_1, \ldots, B_{m-1} \cup B_m\}$. The inequality to be proven amounts to showing that when $\beta \geqslant 1$ we have

$$|B_{m-1} \cup B_m|^\beta \geqslant |B_{m-1}|^\beta + |B_m|^\beta.$$

Since B_{m-1} and B_m are disjoint, this amounts to

$$(|B_{m-1}| + |B_m|)^\beta \geqslant |B_{m-1}|^\beta + |B_m|^\beta.$$

Note that for the function $\phi : [0,1] \longrightarrow \mathbb{R}$ defined by $\phi(t) = t^\beta + (1-t)^\beta$ we have $\phi(0) = \phi(1) = 1$ and the function has a minimum in $\frac{1}{2}$, $\phi\left(\frac{1}{2}\right) = \frac{1}{2^{1-\beta}}$. Therefore, $\phi(t) \geqslant 1$ when $t \in [0,1]$. Choosing $t = \frac{|B_m|}{|B_{m-1}|+|B_m|}$ in the last inequality we obtain the desired result.

Since $\mathcal{H}_\beta(S, \alpha_S) = \frac{1-|S|^{1-\beta}}{1-2^{1-\beta}}$, Part (ii) is immediate.
The last part follows from the definition of $\pi + \sigma$.

Definition

The *joint β-entropy* of partitions $\pi, \sigma \in PART(S)$ is the β-entropy $\mathcal{H}_\beta(\pi \wedge \sigma)$.

For $\pi = \{B_1, \ldots, B_m\}$ and $\sigma = \{C_1, \ldots, C_n\}$ in $PART(S)$ the joint β-entropy is:

$$\mathcal{H}_\beta(\pi \wedge \sigma) = \frac{1}{1 - 2^{1-\beta}} \left(1 - \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\frac{|B_i \cap C_j|}{|S|} \right)^\beta \right).$$

The entropies previously introduced generate corresponding *conditional entropies*.

Let $\pi \in PART(S)$ and let $C \subseteq S$. Denote by π_C the "trace" of π on C given by

$$\pi_C = \{B \cap C | B \in \pi \text{ such that } B \cap C \neq \emptyset\}.$$

Clearly, $\pi_C \in PART(C)$; also, if C is a block of π, then $\pi_C = \omega_C$.

Definition

Let $\pi, \sigma \in PART(S)$ and let $\sigma = \{C_1, \ldots, C_n\}$. The *$\beta$-conditional entropy* of partitions $\pi, \sigma \in PART(S)$ is the function $\mathcal{H}_\beta : PART(S)^2 \longrightarrow \mathbb{R}_{\geqslant 0}$ defined by

$$\mathcal{H}_\beta(\pi|\sigma) = \sum_{j=1}^{n} \left(\frac{|C_j|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_{C_j}).$$

Note that $\mathcal{H}_\beta(\pi|\omega_S) = \mathcal{H}_\beta(\pi)$ and that $\mathcal{H}_\beta(\omega_S|\pi) = \mathcal{H}_\beta(\pi|\alpha_S) = 0$ for every partition $\pi \in PART(S)$.

For $\pi = \{B_1, \ldots, B_m\}$ and $\sigma = \{C_1, \ldots, C_n\}$, the conditional entropy can be written explicitly as

$$\begin{aligned}\mathcal{H}_\beta(\pi|\sigma) &= \sum_{j=1}^{n} \left(\frac{|C_j|}{|S|}\right)^\beta \sum_{i=1}^{m} \frac{1}{1-2^{1-\beta}} \left[1 - \left(\frac{|B_i \cap C_j|}{|C_j|}\right)^\beta\right] \\ &= \frac{1}{1-2^{1-\beta}} \sum_{j=1}^{n} \left(\left(\frac{|C_j|}{|S|}\right)^\beta - \sum_{i=1}^{m} \left(\frac{|B_i \cap C_j|}{|S|}\right)^\beta\right). \qquad (1)\end{aligned}$$

For the special case when $\pi = \alpha_S$, we can write

$$\mathcal{H}_\beta(\alpha_S|\sigma) = \sum_{j=1}^{n} \left(\frac{|C_j|}{|S|}\right)^\beta \mathcal{H}_\beta(\alpha_{C_j}) = \frac{1}{1-2^{1-\beta}} \left(\sum_{j=1}^{n} \left(\frac{|C_j|}{|S|}\right)^\beta - \frac{1}{|S|^{\beta-1}}\right). \tag{2}$$

Example

By applying l'Hôpital rule the Shannon conditional entropy $\mathcal{H}(\pi|\sigma)$ is

$$\begin{aligned} \mathcal{H}(\pi|\sigma) &= \lim_{\beta \leftarrow 1} \frac{1}{1-2^{1-\beta}} \sum_{j=1}^{n} \left(\left(\frac{|C_j|}{|S|}\right)^\beta - \sum_{i=1}^{m} \left(\frac{|B_i \cap C_j|}{|S|}\right)^\beta \right) \\ &= \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{|B_i \cap C_j|}{|S|} \log \frac{|B_i \cap C_j|}{|C_j|}. \end{aligned}$$

Theorem

Let S be a finite set and let $\pi, \sigma \in PART(S)$. We have $\mathcal{H}_\beta(\pi|\sigma) = 0$ if and only if $\sigma \leqslant \pi$.

Proof.

Suppose that $\sigma = \{C_1, \ldots, C_n\}$. If $\sigma \leqslant \pi$, then $\pi_{C_j} = \omega_{C_j}$ for $1 \leqslant j \leqslant n$ and therefore

$$\mathcal{H}_\beta(\pi|\sigma) = \sum_{j=1}^{n} \left(\frac{|C_j|}{|S|}\right)^\beta \mathcal{H}_\beta(\omega_{C_j}) = 0.$$

Conversely, suppose that

$$\mathcal{H}_\beta(\pi|\sigma) = \sum_{j=1}^{n} \left(\frac{|C_j|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_{C_j}) = 0.$$

This implies $\mathcal{H}_\beta(\pi_{C_j}) = 0$ for $1 \leqslant j \leqslant n$, which means that $\pi_{C_j} = \omega_{C_j}$ for $1 \leqslant j \leqslant n$ by a previous remark. This means that every block C_j of σ is included in a block of π, so $\sigma \leqslant \pi$. □

The joint entropy of two partitions is linked to conditional entropy in the next statement.

Theorem

Let π and σ be two partitions of a finite set S. We have

$$\mathcal{H}_\beta(\pi \wedge \sigma) = \mathcal{H}_\beta(\pi|\sigma) + \mathcal{H}_\beta(\sigma) = \mathcal{H}_\beta(\sigma|\pi) + \mathcal{H}_\beta(\pi),$$

UMASS BOSTON

Proof

By a previous equality we have:

$$
\begin{aligned}
&\mathcal{H}_\beta(\pi \wedge \sigma) - \mathcal{H}_\beta(\pi|\sigma)\\
= \quad & \frac{1}{1-2^{1-\beta}} \left(1 - \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{|B_i \cap C_j|}{|S|}\right)^\beta\right)\\
& - \frac{1}{1-2^{1-\beta}} \sum_{j=1}^{n}\left(\left(\frac{|C_j|}{|S|}\right)^\beta - \sum_{i=1}^{m}\left(\frac{|B_i \cap C_j|}{|S|}\right)^\beta\right)\\
= \quad & \mathcal{H}_\beta(\sigma).
\end{aligned}
$$

The second equality has a similar proof.

Corollary

If $\mathcal{H}_\beta(\pi \wedge \sigma) = \mathcal{H}_\beta(\pi)$*, then* $\pi \leqslant \sigma$*.*

Proof.

Since $\mathcal{H}_\beta(\pi \wedge \sigma) = \mathcal{H}_\beta(\pi)$, we have $\mathcal{H}_\beta(\sigma|\pi) = 0$. Thus, we have $\pi \leqslant \sigma$. □

We saw that

$$\mathcal{H}_\beta(\pi) - \mathcal{H}_\beta(\pi|\sigma) = \mathcal{H}_\beta(\sigma) - \mathcal{H}_\beta(\sigma|\pi).$$

This justifies the following definition:

Definition

The *mutual information* of $\pi, \sigma \in PART(S)$ is the number:

$$I_\beta(\pi, \sigma) = \mathcal{H}_\beta(\pi) - \mathcal{H}_\beta(\pi|\sigma) = \mathcal{H}_\beta(\sigma) - \mathcal{H}_\beta(\sigma|\pi).$$

Taking into account the definition of the joint entropy we obtain a symmetric expression for the mutual information of π and σ as

$$I_\beta(\pi, \sigma) = \mathcal{H}_\beta(\pi) + \mathcal{H}_\beta(\sigma) - \mathcal{H}_\beta(\pi \wedge \sigma).$$

Therefore,

$$\mathcal{H}_\beta(\pi) - \mathcal{H}_\beta(\pi|\sigma) = \mathcal{H}_\beta(\sigma) - \mathcal{H}_\beta(\sigma|\pi).$$

Lemma

Let $w_1, \ldots, w_n$ be n positive numbers such that $\sum_{i=1}^{n} w_i = 1$, $a_1, \ldots, a_n \in [0, 1]$, and let $\beta \geqslant 1$. We have

$$1 - \left(\sum_{i=1}^{n} w_i a_i\right)^{\beta} - \left(\sum_{i=1}^{n} w_i (1 - a_i)\right)^{\beta} \geqslant \sum_{i=1}^{n} w_i^{\beta} \left(1 - a_i^{\beta} - (1 - a_i)^{\beta}\right).$$

Proof

It is easy to see that $x^\beta + (1-x)^\beta \leqslant 1$ for $x \in [0,1]$. This implies

$$w_i \left(1 - a_i^\beta - (1-a_i)^\beta\right) w_i^\beta \left(1 - a_i^\beta - (1-a_i)^\beta\right)$$

because $w_i \in (0,1)$ and $\beta \geqslant 1$.
By applying Jensen's inequality to the convex function $h(x) = x^\beta$ we have

$$\left(\sum_{i=1}^n w_i a_i\right)^\beta \leqslant \sum_{i=1}^n w_i a_i^\beta$$

$$\left(\sum_{i=1}^n w_i (1-a_i)\right)^\beta \leqslant \sum_{i=1}^n w_i (1-a_i)^\beta.$$

These inequalities allow us to write

$$1 - \left(\sum_{i=1}^n w_i a_i\right)^\beta - \left(\sum_{i=1}^n w_i (1-a_i)\right)^\beta$$

$$= \sum_{i=1}^n w_i - \left(\sum_{i=1}^n w_i a_i\right)^\beta - \left(\sum_{i=1}^n w_i (1-a_i)\right)^\beta$$

Theorem

Let S be a set, $\pi \in PART(S)$ and let C and D be two disjoint subsets of S. For $\beta \geqslant 1$, we have

$$|C \cup D|^{\beta} \mathcal{H}_{\beta}(\pi_{C \cup D}) \geqslant |C|^{\beta} \mathcal{H}_{\beta}(\pi_C) + |D|^{\beta} \mathcal{H}_{\beta}(\pi_D).$$

Proof

Let $\pi = \{B_1, \ldots, B_n\} \in PART(S)$. Define

$$w_i = \frac{|B_i \cap (C \cup D)|}{|C \cup D|}, a_i = \frac{|B_i \cap C|}{|B_i \cap (C \cup D)|}$$

for $1 \leqslant i \leqslant n$, so $1 - a_i = \frac{|B_i \cap D|}{|B_i \cap (C \cup D)|}$.
By a previous lemma we have:

$$1 - \left(\sum_{i=1}^{n} \frac{|B_i \cap C|}{|C \cup D|}\right)^{\beta} - \left(\sum_{i=1}^{n} \frac{|B_i \cap D|}{|C \cup D|}\right)^{\beta}$$

$$\geqslant \sum_{i=1}^{n} \left(\frac{|B_i \cap (C \cup D)|}{|C \cup D|}\right)^{\beta} \left(1 - \left(\frac{|B_i \cap C|}{|B_i \cap (C \cup D)|}\right)^{\beta} - \left(\frac{|B_i \cap D|}{|B_i \cap (C \cup D)|}\right)\right.$$

Proof cont'd

The last inequality is equivalent to

$$|C \cup D|^{\beta} - \sum_{i=1}^{n} |B_i \cap (C \cup D)|^{\beta}$$
$$\geqslant \quad |C|^{\beta} - \sum_{i=1}^{n} |B_i \cap C|^{\beta} + |D|^{\beta} - \sum_{i=1}^{n} |B_i \cap D|^{\beta}.$$

This last inequality leads immediately to the inequality of the theorem.

The β-conditional entropy is dually monotonic with respect to its first argument and is monotonic with respect to its second argument.

Theorem

Let $\pi, \sigma, \sigma' \in PART(S)$*, where* S *is a finite set. If* $\sigma \leqslant \sigma'$*, then* $\mathcal{H}_\beta(\sigma|\pi) \geqslant \mathcal{H}_\beta(\sigma'|\pi)$ *and* $\mathcal{H}_\beta(\pi|\sigma) \leqslant \mathcal{H}_\beta(\pi|\sigma')$*.*

Proof

Since $\sigma \leqslant \sigma'$, we have $\pi \wedge \sigma \leqslant \pi \wedge \sigma'$, so $\mathcal{H}_\beta(\pi \wedge \sigma) \geqslant \mathcal{H}_\beta(\pi \wedge \sigma')$. Therefore, $\mathcal{H}_\beta(\sigma|\pi) + \mathcal{H}_\beta(\pi) \geqslant \mathcal{H}_\beta(\sigma'|\pi) + \mathcal{H}_\beta(\pi)$, which implies $\mathcal{H}_\beta(\sigma|\pi) \geqslant \mathcal{H}_\beta(\sigma'|\pi)$.

Proof cont'd

For the second part of the theorem, it suffices to prove the inequality for partitions σ, σ' such that $\sigma \prec \sigma'$. Without loss of generality we may assume that $\sigma = \{C_1, \ldots, C_{n-2}, C_{n-1}, C_n\}$ and $\sigma' = \{C_1, \ldots, C_{n-2}, C_{n-1} \cup C_n\}$. Thus, we can write

$$\begin{aligned}
&\mathcal{H}_\beta(\pi|\sigma') \\
&= \sum_{j=1}^{n-2} \left(\frac{|C_j|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_{C_j}) + \left(\frac{|C_{n-1} \cup C_n|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_{C_{n-1} \cup C_n}) \\
&\geqslant \left(\frac{|C_j|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_{C_j}) + \left(\frac{|C_{n-1}|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_{C_{n-1}}) + \left(\frac{|C_n|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_{C_n}) \\
&= \mathcal{H}(\pi|\sigma).
\end{aligned}$$

Corollary

We have $\mathcal{H}_\beta(\pi) \geqslant \mathcal{H}_\beta(\pi|\sigma)$ *for every* $\pi, \sigma \in PART(S)$.

Proof.

We noted that $\mathcal{H}_\beta(\pi) = \mathcal{H}_\beta(\pi|\omega_S)$. Since $\omega_S \geqslant \sigma$, the statement follows. □

Corollary

Let ξ, θ, θ' be three partitions of a finite set S. If $\theta \geqslant \theta'$, then

$$\mathcal{H}_\beta(\xi \wedge \theta) - \mathcal{H}_\beta(\theta) \geqslant \mathcal{H}_\beta(\xi \wedge \theta') - \mathcal{H}_\beta(\theta').$$

Proof

By a previous result we have:

$$\mathcal{H}_\beta(\xi \wedge \theta) - \mathcal{H}_\beta(\xi \wedge \theta') = \mathcal{H}_\beta(\xi|\theta) + \mathcal{H}_\beta(\theta) - \mathcal{H}_\beta(\xi|\theta') - \mathcal{H}_\beta(\theta').$$

The monotonicity of $\mathcal{H}_\beta(|)$ in its second argument means that: $\mathcal{H}_\beta(\xi|\theta) - \mathcal{H}_\beta(\xi|\theta') \geqslant 0$, so $\mathcal{H}_\beta(\xi \wedge \theta) - \mathcal{H}_\beta(\xi \wedge \theta') \geqslant \mathcal{H}_\beta(\theta) - \mathcal{H}_\beta(\theta')$, which implies the desired inequality.

The behavior of β-conditional entropies with respect to the "addition" of partitions is discussed in the next statement.

Theorem

Let S be a finite set and π and θ be two partitions of S, where $\theta = \{D_1, \ldots, D_h\}$. If $\sigma_i \in PART(D_i)$ for $1 \leqslant i \leqslant h$, then

$$\mathcal{H}_\beta(\pi|\sigma_1 + \cdots + \sigma_h) = \sum_{i=1}^{h} \left(\frac{|D_i|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_{D_i}|\sigma_i).$$

If $\tau = \{F_1, \ldots, F_k\}$ and $\sigma = \{C_1, \ldots, C_n\}$ are two partitions of S, let $\pi_i \in PART(F_i)$ for $1 \leqslant i \leqslant k$. Then,

$$\mathcal{H}_\beta(\pi_1 + \cdots + \pi_k|\sigma) = \sum_{i=1}^{k} \left(\frac{|F_i|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_i|\sigma_{F_i}) + \mathcal{H}_\beta(\tau|\sigma).$$

Proof

Suppose that $\sigma_i = \{E_i^\ell \mid 1 \leqslant \ell \leqslant p_i\}$. The blocks of the partition $\sigma_1 + \cdots + \sigma_h$ are the sets of the collection $\bigcup_{i=1}^h \{E_i^\ell \mid 1 \leqslant \ell \leqslant p_i\}$. Thus, we have

$$\mathcal{H}_\beta(\pi|\sigma_1 + \cdots + \sigma_h) = \sum_{i=1}^{h} \sum_{\ell=1}^{p_i} \left(\frac{|E_i^\ell|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_{E_i^\ell}).$$

On the other hand, since $(\pi_{D_i})_{E_i^\ell} = \pi_{E_i^\ell}$, we have

$$\begin{aligned}\sum_{i=1}^{h} \left(\frac{|D_i|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_{D_i}|\sigma_i) &= \sum_{i=1}^{h} \left(\frac{|D_i|}{|S|}\right)^\beta \sum_{\ell=1}^{p_i} \left(\frac{|E_i^\ell|}{|D_i|}\right)^\beta \mathcal{H}_\beta(\pi_{E_i^\ell}) \\ &= \sum_{i=1}^{h} \sum_{\ell=1}^{p_i} \left(\frac{|E_i^\ell|}{|S|}\right)^\beta \mathcal{H}_\beta(\pi_{E_i^\ell}),\end{aligned}$$

which gives the first equality of the theorem.

Proof cont'd

To prove the second part, observe that $(\pi_1 + \cdots + \pi_k)_{C_j} = (\pi_1)_{C_j} + \cdots + (\pi_k)_{C_j}$ for every block C_j of σ. Thus, we have

$$\mathcal{H}_\beta(\pi_1 + \cdots + \pi_k | \sigma) = \sum_{j=1}^{n} \left(\frac{|C_j|}{|S|}\right)^\beta \mathcal{H}_\beta((\pi_1)_{C_j} + \cdots + (\pi_k)_{C_j}).$$

By applying a previous corollary to partitions $(\pi_1)_{C_j}, \ldots, (\pi_k)_{C_j}$ of C_j, we can write

$$\mathcal{H}_\beta((\pi_1)_{C_j} + \cdots + (\pi_k)_{C_j}) = \sum_{i=1}^{k} \left(\frac{|F_i \cap C_j|}{|C_j|}\right)^\beta \mathcal{H}_\beta((\pi_i)_{C_j}) + \mathcal{H}_\beta(\tau_{C_j}).$$

Thus,

$$
\begin{aligned}
&\mathcal{H}_\beta(\pi_1 + \cdots + \pi_k|\sigma)\\
&= \sum_{j=1}^{n}\sum_{i=1}^{k}\left(\frac{|F_i \cap C_j|}{|S|}\right)^{\beta}\mathcal{H}_\beta((\pi_i)_{C_j}) + \sum_{j=1}^{n}\left(\frac{|C_j|}{|S|}\right)^{\beta}\mathcal{H}_\beta(\tau_{C_j})\\
&= \sum_{i=1}^{k}\left(\frac{|F_i|}{|S|}\right)^{\beta}\sum_{j=1}^{n}\left(\frac{|F_i \cap C_j|}{|F_i|}\right)^{\beta}\mathcal{H}_\beta((\pi_i)_{F_i\cap C_j}) + \mathcal{H}_\beta(\tau|\sigma)\\
&= \sum_{i=1}^{k}\left(\frac{|F_i|}{|S|}\right)^{\beta}\mathcal{H}_\beta(\pi_i|\sigma_{F_i}) + \mathcal{H}_\beta(\tau|\sigma),
\end{aligned}
$$

which is the desired equality.

Theorem

Let π, σ, τ be three partitions of the finite set S. We have

$$\mathcal{H}_\beta(\pi|\sigma \wedge \tau) + \mathcal{H}_\beta(\sigma|\tau) = \mathcal{H}_\beta(\pi \wedge \sigma|\tau).$$

Proof

By a previous theorem we can write

$$\begin{aligned}\mathcal{H}_\beta(\pi|\sigma \wedge \tau) &= \mathcal{H}_\beta(\pi \wedge \sigma \wedge \tau) - \mathcal{H}_\beta(\sigma \wedge \tau)\\ \mathcal{H}_\beta(\sigma|\tau) &= \mathcal{H}_\beta(\sigma \wedge \tau) - \mathcal{H}_\beta(\tau).\end{aligned}$$

By adding these equalities we obtain the equality of the theorem.

Corollary

Let π, σ, τ be three partitions of the finite set S. Then, we have

$$\mathcal{H}_\beta(\pi|\sigma) + \mathcal{H}_\beta(\sigma|\tau) \geqslant \mathcal{H}_\beta(\pi|\tau).$$

Proof

By a previous result, the monotonicity of β-conditional entropy in its second argument, and the antimonotonicity of the same in its first argument, we can write

$$\begin{aligned} \mathcal{H}_\beta(\pi|\sigma) + \mathcal{H}_\beta(\sigma|\tau) &\geqslant \mathcal{H}_\beta(\pi|\sigma \wedge \tau) + \mathcal{H}_\beta(\sigma|\tau) \\ &= \mathcal{H}_\beta(\pi \wedge \sigma|\tau) \\ &\geqslant \mathcal{H}_\beta(\pi|\tau), \end{aligned}$$

which is the desired inequality.

The property of $\mathcal{H}_\beta$ described next is known as the *submodularity* of entropy.

Corollary

Let π and σ be two partitions of the finite set S. Then, we have

$$\mathcal{H}_\beta(\pi \vee \sigma) + \mathcal{H}_\beta(\pi \wedge \sigma) \leqslant \mathcal{H}_\beta(\pi) + \mathcal{H}_\beta(\sigma).$$

Proof

By a previous Corollary, we have $\mathcal{H}_\beta(\pi|\sigma) \leqslant \mathcal{H}_\beta(\pi|\tau) + \mathcal{H}_\beta(\tau|\sigma)$. Then, we obtain

$$\mathcal{H}_\beta(\pi \wedge \sigma) - \mathcal{H}_\beta(\sigma) \leqslant \mathcal{H}_\beta(\pi \wedge \tau) - \mathcal{H}_\beta(\tau) + \mathcal{H}_\beta(\tau \wedge \sigma) - \mathcal{H}_\beta(\sigma),$$

hence

$$\mathcal{H}_\beta(\tau) + \mathcal{H}_\beta(\pi \wedge \sigma) \leqslant \mathcal{H}_\beta(\pi \wedge \tau) + \mathcal{H}_\beta(\tau \wedge \sigma).$$

Choosing $\tau = \pi \vee \sigma$ implies immediately the inequality of the corollary.

The set of partitions $PART(S)$ can be equipped with a metric derived from the notion of entropy.

The initial result in this direction was obtained by L. de Mántaras in who proved that the mapping $d : PART(S)^2 \longrightarrow \mathbb{R}_{\geqslant 0}$ defined as

$$d(\pi, \sigma) = \mathcal{H}(\pi|\sigma) + \mathcal{H}(\sigma|\pi)$$

(for the Shannon entropy $\mathcal{H}$) is a metric on $PART(S)$. We show that this result holds for any β-entropy.

Theorem

The mapping $d_\beta : PART(S)^2 \longrightarrow \mathbb{R}_{\geqslant 0}$ *defined by*

$$d_\beta(\pi, \sigma) = \mathcal{H}_\beta(\pi|\sigma) + \mathcal{H}_\beta(\sigma|\pi).$$

is a metric on $PART(S)$*.*

Proof

A double application of a previous Corollary implies

$$\begin{aligned}\mathcal{H}_\beta(\pi|\sigma) + \mathcal{H}_\beta(\sigma|\tau) &\geqslant \mathcal{H}_\beta(\pi|\tau),\\ \mathcal{H}_\beta(\sigma|\pi) + \mathcal{H}_\beta(\tau|\sigma) &\geqslant \mathcal{H}_\beta(\tau|\pi),\end{aligned}$$

for all $\pi, \sigma, \tau \in PART(S)$. This implies the triangular inequality

$$d_\beta(\pi, \sigma) + d_\beta(\sigma, \tau) \geqslant d_\beta(\pi, \tau).$$

Proof cont'd

The symmetry of d_β is obvious and it is clear that $d_\beta(\sigma, \sigma) = 0$ for every $\sigma \in PART(S)$.
Suppose now that $d_\beta(\pi, \sigma) = 0$. Since the values of the β-entropy are non-negative we have $\mathcal{H}_\beta(\pi|\sigma) = \mathcal{H}_\beta(\sigma|\pi) = 0$. This implies both $\sigma \leqslant \pi$ and $\pi \leqslant \sigma$, so $\pi = \sigma$.

Note that d_β can also be written as

$$\begin{aligned} d_\beta(\pi,\sigma) &= \mathcal{H}_\beta(\pi \wedge \sigma) - \mathcal{H}_\beta(\sigma) + \mathcal{H}_\beta(\pi \wedge \sigma) - \mathcal{H}_\beta(\pi) \\ &= 2\mathcal{H}_\beta(\pi \wedge \sigma) - \mathcal{H}_\beta(\pi) - \mathcal{H}_\beta(\sigma). \end{aligned}$$

We can also write

$$d_\beta(\pi,\sigma) + I_\beta(\pi,\sigma) = \mathcal{H}_\beta(\pi \wedge \sigma),$$

an equality which relates the distance between partitions, the mutual information, and the joint entropy, three important measures that we introduced related to β-entropy.

For a partition $\pi \in PART(D)$ we write $x \equiv_\pi y$ if there is a block $B \in \pi$ such that $\{x, y\} \subseteq B$. It is immediate that "$\equiv_\pi$ is an equivalence relation. Let $\tau = \{T_1, \ldots, T_r\}$ be a *reference partition* of data set D (also known as the *ground-truth* partition) and let $\kappa = (C_1, \ldots, C_m)$ be a clustering.

The pairs of elements of D can be classified into four classes relative to the partitions τ and σ. Namely, a pair (x, y) with $x \neq y$ is

- a *true positive pair* if $x \equiv_\tau y$ and $x \equiv_\kappa y$;
- a *true negative pair* if $x \not\equiv_\tau y$ and $x \not\equiv_\kappa y$;
- a *false positive pair* if $x \not\equiv_\tau y$ and $x \equiv_\kappa y$;
- a *false negative pair* if $x \equiv_\tau y$ and $x \not\equiv_\kappa y$.

The number of true positive pairs is denoted by $\mathtt{TP}(\tau, \kappa)$, that of true negative pairs is $\mathtt{TN}(\tau, \kappa)$, the number of false positive pairs is $\mathtt{FP}(\tau, \kappa)$, and the number of false negative pairs is $\mathtt{FN}(\tau, \kappa)$.

For $|D| = n$ there are $\binom{n}{2}$ distinct pairs, hence

$$\binom{n}{2} = \mathtt{TP}(\tau,\kappa) + \mathtt{TN}(\tau,\kappa) + \mathtt{FP}(\tau,\kappa) + \mathtt{FN}(\tau,\kappa).$$

Let $G(\tau,\kappa) = (g_{ij}) \in \mathbb{R}^{r\times m}$ be the *contingency matrix* for the reference partition $\tau = \{T_1,\ldots,T_r\}$ and clustering $\kappa = \{C_1,\ldots,C_m\}$. We introduce the partial sums:

$$\begin{aligned} g_{i\cdot} &= \sum_{j=1}^{m} g_{ij} = |T_i|, \\ g_{\cdot j} &= \sum_{i=1}^{r} g_{ij} = |C_j|, \\ g_{\cdot\cdot} &= \sum_{i=1}^{r}\sum_{j=1}^{m} g_{ij} = |D|, \end{aligned}$$

for $1 \leqslant i \leqslant r$ and $1 \leqslant j \leqslant m$.

These notations are summarized by the following table:

		part. κ				
	class	C_1	C_2	$\cdots$	C_m	sums
	T_1	g_{11}	g_{12}	$\cdots$	g_{1m}	$g_{1\cdot}$
	T_2	g_{21}	g_{22}	$\cdots$	g_{2m}	$g_{2\cdot}$
part. τ	$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$
	T_r	g_{r1}	g_{r2}	$\cdots$	g_{rm}	$g_{r\cdot}$
	sums	$g_{\cdot 1}$	$g_{\cdot 2}$	$\cdots$	$g_{\cdot m}$	$g_{\cdot\cdot} = \lvert D \rvert$

All indices mentioned above can be computed in $O(rm)$ time because the contingency matrix $G(\tau, \kappa)$ can be computed in linear time.
We have

$$\begin{aligned}
\mathrm{TP}(\tau, \kappa) &= \sum_{i=1}^{r} \sum_{j=1}^{m} \binom{g_{ij}}{2} \\
&= \frac{1}{2} \left(\sum_{i=1}^{r} \sum_{j=1}^{m} g_{ij}^2 - \sum_{i=1}^{r} \sum_{j=1}^{m} g_{ij} \right) \\
&= \frac{1}{2} \left(\sum_{i=1}^{r} \sum_{j=1}^{m} g_{ij}^2 - n \right).
\end{aligned}$$

The number of pairs that belong to the same block of the reference partition is $\sum_{i=1}^{r} \binom{g_{i\cdot}}{2}$. If we eliminate from these pairs the true positive pairs we obtain the number of false negative pairs:

$$\begin{aligned}
\mathtt{FN}(\tau,\kappa) &= \sum_{i=1}^{r} \binom{g_{i\cdot}}{2} - \mathtt{TP}(\tau,\kappa) \\
&= \frac{1}{2}\sum_{i=1}^{r} g_{i\cdot}^2 - \frac{1}{2}\sum_{i=1}^{r} g_{i\cdot} - \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{m} g_{ij}^2 + \frac{n}{2} \\
&= \frac{1}{2}\left(\sum_{i=1}^{r} g_{i\cdot}^2 - \sum_{i=1}^{r}\sum_{j=1}^{m} g_{ij}^2\right)
\end{aligned}$$

because $\sum_{i=1}^{r} g_{i\cdot} = n$.

The number of false positive pairs is obtained by subtracting from the number of pairs that belong to the same cluster the number of true positive pairs:

$$
\begin{aligned}
\mathrm{FP}(\tau,\kappa) &= \sum_{j=1}^{m} \binom{g_{\cdot j}}{2} - \mathrm{TP}(\tau,\kappa) \\
&= \frac{1}{2}\sum_{j=1}^{m} g_{\cdot j}^2 - \frac{1}{2}\sum_{j=1}^{m} g_{\cdot j} - \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{m} g_{ij}^2 + \frac{n}{2} \\
&= \frac{1}{2}\left(\sum_{j=1}^{m} g_{\cdot j}^2 - \sum_{i=1}^{r}\sum_{j=1}^{m} g_{ij}^2\right)
\end{aligned}
$$

because $\sum_{j=1}^{m} g_{\cdot j} = n$.

The number of true negative pairs is

$$\mathtt{TN}(\tau, \kappa) \quad = \quad \frac{1}{2}\left(g_{\cdot\cdot}^2 - \sum_{i=1}^{r} g_{i\cdot}^2 - \sum_{j=1}^{m} g_{\cdot j}^2 + \sum_{i=1}^{r}\sum_{j=1}^{m} g_{ij}^2 \right).$$

These numbers can be used, in turn, to compute efficiently several numerical characteristics of the pair (τ, κ).
Let ρ_τ and ρ_κ the equivalences that correspond to the partitions τ and κ. These equivalences are sets of pairs in $D \times D$. Therefore, it makes sense to consider their Jaccard coefficient:

$$J(\rho_\tau, \rho_\kappa) = \frac{|\rho_\tau \cap \rho_\kappa|}{|\rho_\tau \cup \rho_\kappa|},$$

which evaluates the similarity between the reference partition τ and the clustering κ. It is clear that

$$|\rho_\tau \cap \rho_\kappa| = \frac{|\mathtt{TP}(\tau, \kappa)|}{|\mathtt{TP}(\tau, \kappa)| + |\mathtt{FN}(\tau, \kappa)| + |\mathtt{FP}(\tau, \kappa)|}.$$

The *Rand coefficient* is

$$R(\tau, \kappa) = \frac{|\mathtt{TP}(\tau, \kappa)| + |\mathtt{TN}(\tau, \kappa)|}{\binom{n}{2}},$$

and represents the fraction of objects where the reference partition and the clustering agree. When $R(\tau, \kappa) = 1$ the two partitions are identical.

The notions of precision and recall previously introduced are reformulated for pairs of objects.

The *precision* for τ and κ is

$$\text{precision}(\tau, \kappa) = \frac{\mathtt{TP}(\tau, \kappa)}{\mathtt{TP}(\tau, \kappa) + \mathtt{FP}(\tau, \kappa)}$$

and reflects the size of the set of correctly classified pairs of objects vs. the size of the sets of pairs of objects that reside in the same cluster. We have $\text{precision}(\tau, \kappa) = 1$ if and only if no false positive pairs.

The *recall* for τ and κ is

$$\text{recall}(\tau, \kappa) = \frac{\mathtt{TP}(\tau, \kappa)}{\mathtt{TP}(\tau, \kappa) + \mathtt{FN}(\tau, \kappa)}$$

Recall evaluates the fraction of correctly classiffied pairs of objects compared to all pairs of objects that inhabit the same block of reference partition.

We have $\text{recall}(\tau, \kappa) = 1$ if $\mathtt{FN}(\tau, \kappa) = 0$, that is, if there are no pairs in ρ_τ whose components belong to two distinct clusters.

The *Fowlkes-Mallows* coefficient $\mathtt{FM}(\tau, \kappa)$ is the geometric average of recall and precision, that is,

$$\begin{aligned}
\mathtt{FM}(\tau, \kappa) &= \sqrt{\text{precision}(\tau, \kappa) \cdot \text{recall}(\tau, \kappa)} \\
&= \frac{\mathtt{TP}(\tau, \kappa)}{\sqrt{(\mathtt{TP}(\tau, \kappa) + \mathtt{FP}(\tau, \kappa))(\mathtt{TP}(\tau, \kappa) + \mathtt{FN}(\tau, \kappa))}}.
\end{aligned}$$