Clustering - I

Prof. Dan A. Simovici

UMB

Definition

An *equivalence relation* on a set S is a relation ρ that is reflexive, symmetric, and transitive.

This means that

- $(x, x) \in \rho$ for every $x \in S$;
- $(x, y) \in \rho$ if and only if $(y, x) \in \rho$;
- $(x, y) \in \rho$ and $(y, z) \in \rho$ imply $(x, z) \in \rho$.

Example

Let U and V be two sets, and consider a function $f : U \longrightarrow V$. The relation $\mathbf{ker}(f) \subseteq U \times U$, called the *kernel* of f, is given by

$$\mathbf{ker}(f) = \{(u, u') \in U \times U \mid f(u) = f(u')\}.$$

In other words, $(u, u') \in \mathbf{ker}(f)$ if f maps both u and u' into *the same* element of V.

Example

Let $m \in \mathbb{N}$ be a positive natural number. Define the function $f_m : \mathbb{Z} \longrightarrow \mathbb{N}$ by $f_m(n) = r$ if r is the remainder of the division of n by m. The range of the function f_m is the set $\{0, \ldots, m-1\}$.
The relation $\mathbf{ker}(f_m)$ is usually denoted by $\equiv_m$. We have $(p, q) \in \equiv_m$ if and only if $p - q$ is divisible by m; if $(p, q) \in \equiv_m$, we also write $p \equiv q (\text{mod } m)$.

Definition

Let ρ be an equivalence on a set U and let $u \in U$.
The *equivalence class* of u is the set $[u]_\rho$, given by

$$[u]_\rho = \{y \in U \mid (u, y) \in \rho\}.$$

When there is no risk of confusion, we write simply $[u]$ instead of $[u]_\rho$.

Note that an equivalence class $[u]$ of an element u is never empty since $u \in [u]$ because of the reflexivity of ρ.

Theorem

Let ρ be an equivalence on a set U and let $u, v \in U$. The following three statements are equivalent:

1. $(u, v) \in \rho$;
2. $[u] = [v]$;
3. $[u] \cap [v] \neq \emptyset$.

Definition

Let S be a set and let $\rho \in \mathrm{EQ}(S)$. A subset U of S is *ρ-saturated* if it equals a union of equivalence classes of ρ.

It is easy to see that U is a ρ-saturated set if and only if $x \in U$ and $(x, y) \in \rho$ imply $y \in U$. It is clear that both $\emptyset$ and S are ρ-saturated sets.

Definition

Let S be a nonempty set. A *partition* of S is a nonempty collection $\pi = \{B_i \mid i \in I\}$ of nonempty subsets of S, such that $\bigcup\{B_i \mid i \in I\} = S$, and $B_i \cap B_j = \emptyset$ for every $i, j \in I$ such that $i \neq j$.
Each set B_i of π is a *block of the partition* π.
The set of partitions of a set S is denoted by $\mathrm{PART}(S)$. The partition of S that consists of all singletons of the form $\{s\}$ with $s \in S$ will be denoted by α_S; the partition that consists of the set S itself will be denoted by ω_S.

Example

For the two-element set $S = \{a, b\}$, there are two partitions: the partition $\alpha_S = \{\{a\}, \{b\}\}$ and the partition $\omega_S = \{\{a, b\}\}$.
For the one-element set $T = \{c\}$, there exists only one partition, $\alpha_T = \omega_T = \{\{t\}\}$.

Example

A complete list of partitions of a set $S = \{a, b, c\}$ consists of

$$\begin{array}{lllllll} \pi_0 & = & \{\{a\},\{b\},\{c\}\}, & \pi_1 & = & \{\{a,b\},\{c\}\}, \\ \pi_2 & = & \{\{a\},\{b,c\}\}, & \pi_3 & = & \{\{a,c\},\{b\}\}, \\ \pi_4 & = & \{\{a,b,c\}\}. & & & \end{array}$$

Clearly, $\pi_0 = \alpha_S$ and $\pi_4 = \omega_S$.

Definition

Let S be a set and let $\pi, \sigma \in \mathrm{PART}(S)$. The partition π is *finer* than the partition σ if every block C of σ is a union of blocks of π. This is denoted by $\pi \leqslant \sigma$.

Theorem

Let $\pi = \{B_i \mid i \in I\}$ *and* $\sigma = \{C_j \mid j \in J\}$ *be two partitions of a set* S*. For* $\pi, \sigma \in PART(S)$*, we have* $\pi \leqslant \sigma$ *if and only if for every block* $B_i \in \pi$ *there exists a block* $C_j \in \sigma$ *such that* $B_i \subseteq C_j$*.*

Proof

If $\pi \leqslant \sigma$, then it is clear for every block $B_i \in \pi$ there exists a block $C_j \in \sigma$ such that $B_i \subseteq C_j$.
Conversely, suppose that for every block $B_i \in \pi$ there exists a block $C_j \in \sigma$ such that $B_i \subseteq C_j$. Since two distinct blocks of σ are disjoint, it follows that for any block B_i of π, the block C_j of σ that contains B_i is unique. Therefore, if a block B of π intersects a block C of σ, then $B \subseteq C$.
Let $Q = \bigcup\{B_i \in \pi \mid B_i \subseteq C_j\}$. Clearly, $Q \subseteq C_j$. Suppose that there exists $x \in C_j - Q$. Then, there is a block $B_\ell \in \pi$ such that $x \in B_\ell \cap C_j$, which implies that $B_\ell \subseteq C_j$. This means that $x \in B_\ell \subseteq C$, which contradicts the assumption we made about x. Consequently, $C_j = Q$, which concludes the argument.

Note that $\alpha_S \leqslant \pi \leqslant \omega_S$ for every $\pi \in \mathrm{PART}(S)$.
Two equivalence classes either coincide or are disjoint. Therefore, starting from an equivalence $\rho \in \mathrm{EQ}(U)$, we can build a partition of the set U.

Definition

The *quotient set* of the set U with respect to the equivalence ρ is the partition U/ρ, where

$$U/\rho = \{[u]_\rho \mid u \in U\}.$$

An alternative notation for the partition U/ρ is π_ρ.

Theorem

Let $\pi = \{B_i \mid i \in I\}$ *be a partition of the set* U*. Define the relation* ρ_π *by* $(x, y) \in \rho_\pi$ *if there is a set* $B_i \in \pi$ *such that* $\{x, y\} \subseteq B_i$*. The relation* ρ_π *is an equivalence.*

Proof

Let B_i be the block of the partition that contains u. Since $\{u\} \subseteq B_i$, we have $(u, u) \in \rho_\pi$ for any $u \in U$, which shows that ρ_π is reflexive.
The relation ρ_π is clearly symmetric. To prove the transitivity of ρ_π, consider $(u, v), (v, w) \in \rho_\pi$. We have the blocks B_i and B_j such that $\{u, v\} \subseteq B_i$ and $\{v, w\} \subseteq B_j$. Since $v \in B_i \cap B_j$, we obtain $B_i = B_j$ by the definition of partitions; hence, $(u, w) \in \rho_\pi$.