Précision

Soit une séquence ou un ensemble de données indexé par j, où chaque élément x_j peut représenter un signal, un vecteur, un document ou un terme.

- $j \in \mathbb{N}$ lorsqu'il représente un indice dans une séquence.
- j désigne un terme lorsqu'on traite des documents (comme dans le calcul de fréquence documentaire).

La distance $\theta \in [-1, 1]$ représente la mesure de proximité entre ces éléments, et peut être ajustée ou modulée par une fonction $f(\theta)$.

Le seuil $R \in \mathbb{R}^+$ est utilisé pour les signaux ou séquences continues. Le seuil $N \in \mathbb{N}$ s'applique aux vecteurs, documents ou termes discrets.

Le symbole $*$ désigne un produit commutatif quelconque, utilisé pour exprimer la modulation du seuil dans les inégalités.

—

Équations avec seuil réel R (pour signaux, séquences) :

$$x_j \geq R * (1 - \theta), \quad f(x)_j \geq R * (1 - \theta), \quad f(x)_j \geq R * (1 - f(\theta))$$

Équations avec seuil entier N (pour vecteurs, documents, termes) :

$$x_j \geq N * (1 - \theta), \quad f(x)_j \geq N * (1 - \theta), \quad f(x)_j \geq N * (1 - f(\theta))$$

Cas particulier des documents, avec df_j la fréquence relative des documents contenant le terme j :

$$\mathrm{df}_j \geq N * (1 - \theta), \quad \mathrm{df}_j \geq N * (1 - f(\theta))$$

—

Variantes avec inégalité stricte :

$$x_j > N * (1 - \theta), \quad \mathrm{df}_j > N * (1 - \theta), \quad \mathrm{df}_j > N * (1 - f(\theta))$$

—

Notations et définitions clés :

- x_j : élément de séquence ou document ou signal indexé
- $f(x)_j$: évaluation ou transformation de l'élément x, associée à l'indice ou terme j
- df_j : fréquence documentaire du terme j
- θ : distance proprement dite entre les éléments
- $f(\cdot)$: fonction de modulation, appliquée aux données ou à la distance
- $R \in \mathbb{R}^+$: seuil réel, utilisé pour les séquences continues ou les signaux
- $N \in \mathbb{N}$: seuil entier, utilisé pour les données discrètes (documents, vecteurs, termes)
- $*$: produit commutatif générique, modulant l'intensité du seuil
- Les inégalités peuvent être strictes ($>$) ou larges ($\geq$) selon le niveau de précision exigé

Ces équations formelles fournissent un cadre pour l'analyse de la précision ou de la correspondance dans des structures variées (documents, signaux, vecteurs, séquences), tenant compte de la granularité de l'information, de la distance mesurée, et de la fonction de transformation choisie.