TD : Généralités.

1. Preuve de la distance de Jaccard.

Soit x, y des variables binaires à K entrées.

Ex : $x = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$ $y = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$

- Indice de Jaccard : $J(x,y) = \dfrac{M_{11}}{K - M_{00}}$
- Distance de Jaccard : $d(x,y) = 1 - J(x,y)$

$$= \frac{M_{10} + M_{01}}{M_{01} + M_{10} + M_{11}}$$

- Positivité : tous les termes au numérateur et au dénominateur sont ≥ 0. $\Rightarrow d(x,y) \geq 0$.
- Séparation : supposons que $d(x,y) = 0$.

Alors $M_{10} + M_{01} = 0$. Donc il n'y a pas de variables qui 0 pour x et 1 pour y et inversement. Comme $M_{01} + M_{10} + M_{11} > 0$, on a $x = y$.

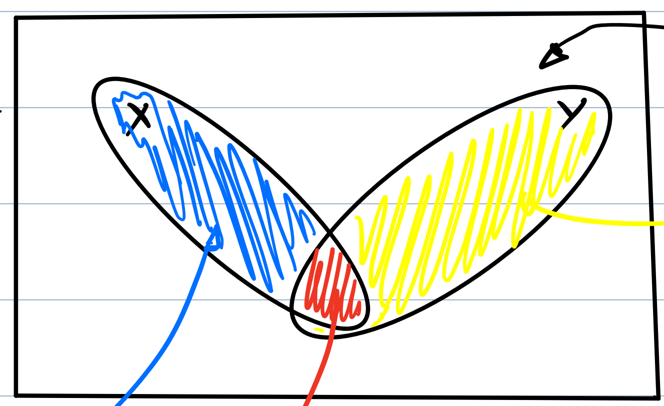
Supposons $x = y$. Alors $M_{01} = M_{10} = 0$ donc $d(x,y) = 0$

- Symétrie : L'indice de Jaccard est symétrique donc $d(x,y) = d(y,x)$

$X = \{ i \in \{1, \dots, K\} \mid x_i = 1 \} \to$ ensemble des indices tel que $x_i = 1$

$Y = \{ i \in \{1, \dots, K\} \mid y_i = 1 \} \to$ ensemble des indices tel que $y_i = 1$.

1.



$\to \{1, \dots, K\}$

$M_{01} \to \begin{cases} i \in Y \\ i \notin X \end{cases}$

$M_{10} \to \begin{cases} i \in X \\ i \notin Y \end{cases}$ $\quad M_{11} \to \begin{cases} i \in X \\ i \in Y \end{cases}$

On note $|A| = \text{card}(A)$ le cardinal de l'ensemble A (i.e. le nb d'éléments $a \in A$)

2. $$d(x,y) = \frac{M_{01} + M_{10}}{M_{01} + M_{10} + M_{11}}$$

$M_{11} = |X \cap Y|$

$M_{10} = |X \setminus Y|$ et $M_{01} = |Y \setminus X|$

donc $M_{10} + M_{01} = |X \setminus Y \cup Y \setminus X| = |X \Delta Y|$

et $M_{10} + M_{01} + M_{11} = |X \cup Y|$

donc $d(x,y) = \dfrac{|X \Delta Y|}{|X \cup Y|}$.

3. $A = \{1, 2\}$, $B = \{2, 3\}$ et $C = \{3, 4\}$

$d(A,B) = \dfrac{|\{1\} \cup \{3\}|}{\quad} = \dfrac{|\{1,3\}|}{\quad} = \dfrac{2}{\quad}$

$|\{1,2,3\}| \qquad |\{1,2,3\}| \qquad 3$

$$d(B,C) = \frac{|\{2\} \cup \{4\}|}{|\{2,3,4\}|} = \frac{2}{3}$$

$$d(A,C) = \frac{|\{1,2\} \cup \{3,4\}|}{|\{1,2,3,4\}|} = 1$$

4. $Z = \{ i \in \{1,\dots,k\} \mid z_i = 1 \}$.

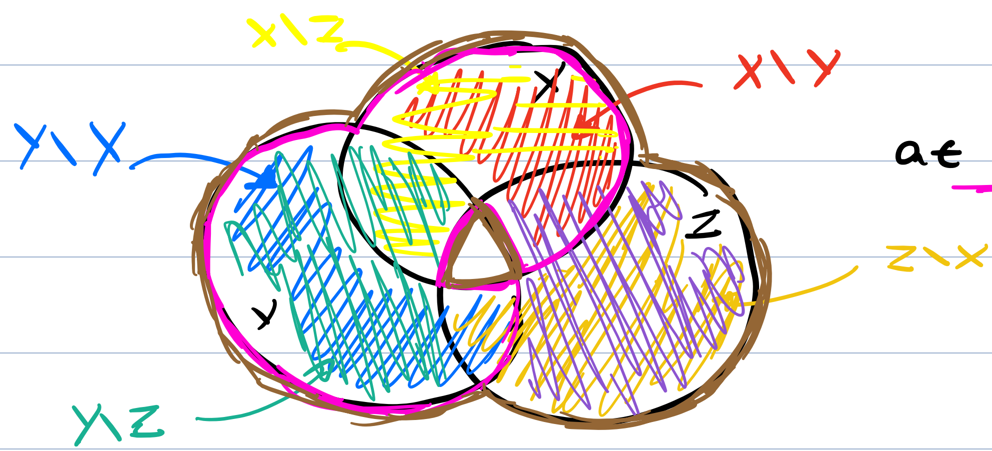
$$d(x,y) \leq d(x,z) + d(z,y)$$

$$\Rightarrow \frac{|X \Delta Y|}{|X \cup Y|} \leq \frac{|X \Delta Z|}{|X \cup Z|} + \frac{|Z \Delta Y|}{|Z \cup Y|}$$

5. Mq $X \Delta Y \subseteq (X \Delta Z) \cup (Y \Delta Z)$.

Soit $a \in X \Delta Y$. Mq $a \in (X \Delta Z) \cup (Y \Delta Z)$.

$a \in X \Delta Y \Leftrightarrow a \in X \setminus Y \cup Y \setminus X$.



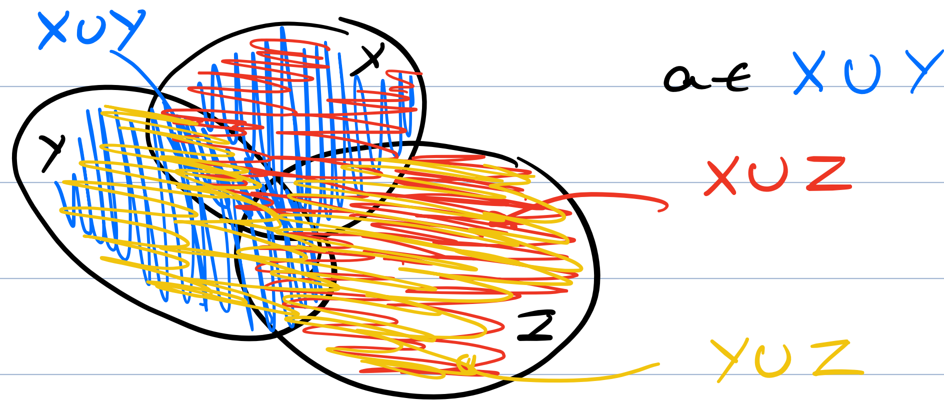
$a \in X \setminus Y \cup Y \setminus X$

$(X \Delta Z) \cup (Y \Delta Z) = X \setminus Z \cup Z \setminus X \cup Y \setminus Z \cup Z \setminus Y$

donc $a \in X \Delta Z \cup Y \Delta Z$

6. Mq $X \cup Y \subseteq (X \cup Z) \cup (Y \cup Z)$

Soit $a \in X \cup Y$. Mq $a \in (X \cup Z) \cup (Y \cup Z)$



donc $a \in X\cup Z \cup Y \cup Z$.

7. Soit A et B des ensembles, on a :

1. Modularité : $|A\cup B| + |A\cap B| = |A| + |B|$
2. Monotonie : Si $A \subseteq B$, alors $|A| \leq |B|$
3. Non négativité : $|A| \geqslant 0$.

→ Pour tous ensembles X, Y, Z, on a.

(*) $|X\cap Z|.|Y\cup Z| + |X\cup Z|.|Y\cap Z| \leq |Z|.\left[|X|+|Y|\right]$.

<u>Preuve</u> :

• $|X\cap Z|.|Y\cup Z| \overset{\text{modularité}}{=} |X\cap Z|.\left[|Y|+|Z|-|Y\cap Z|\right]$

$= |X\cap Z|.|Y| + |X\cap Z|.|Z| - |X\cap Z|.|Y\cap Z|$

$X\cap Z \subseteq Z$ donc $|X\cap Z| \leq |Z|$

$\leq |Z|.|Y| + |X\cap Z|.|Z| - |Z|.|Y\cap Z|$.

$= |Z|.\left[|Y| + |X\cap Z| - |Y\cap Z|\right]$

• $|X\cup Z|.|Y\cap Z| \overset{\text{modularité}}{=} |Y\cap Z|.\left[|X|+|Z|-|X\cap Z|\right]$

$= |Y\cap Z|.|X| + |Y\cap Z|.|Z| - |Y\cap Z|.|X\cap Z|$

$Y\cap Z \subseteq Z$ donc $|Y\cap Z| \leq |Z|$

$\leq |Z|.|X| + |Y\cap Z|.|Z| - |Z|.|X\cap Z|$

$= |Z|.\left[|X| + |Y\cap Z| - |X\cap Z|\right]$

- Donc

$$|X\cap Z|.|Y\cup Z| + |X\cup Z|.|Y\cap Z|$$
$$\leq |Z|.\left[|Y| + \cancel{|X\cap Z|} - \cancel{|Y\cap Z|} + |X| + \cancel{|Y\cap Z|} - \cancel{|X\cap Z|}\right]$$
$$= |Z|.\left[|X| + |Y|\right] \qquad \blacksquare$$

→ Pour tout ensembles S et T, on a

$$(\circledast)\quad |S\cap T|.|S\cup T| \leq |S|.|T|.$$

<u>Preuve</u>: Prendre $X=S$, $Y=S$ et $Z=T$ dans $(\circledast)$

$$|S\cap T|.|S\cup T| + |S\cup T|.|S\cap T| \leq |T|.\,2|S|$$
$$\iff |S\cap T|.|S\cup T| \leq |S|.|T|.$$

→ Pour tout ensembles X, Y, et Z, $d(X,Y) \leq d(X,Z) + d(Y,Z)$

où d est la distance de Jaccard définie par

$$d(X,Y) = \frac{|X\Delta Y|}{|X\cup Y|} = 1 - \frac{|X\cap Y|}{|X\cup Y|}.$$

<u>Preuve</u>: On doit montrer que :

$$1 - \frac{|X\cap Y|}{|X\cup Y|} \leq 1 - \frac{|X\cap Z|}{|X\cup Z|} + 1 - \frac{|Y\cap Z|}{|Y\cup Z|}.$$

$$\iff \frac{|X\cap Z|}{|X\cup Z|} + \frac{|Y\cap Z|}{|Y\cup Z|} \leq 1 + \frac{|X\cap Y|}{|X\cup Y|} \underset{\text{modularité}}{=} \frac{|X| + |Y|}{|X\cup Y|}$$

$$\frac{|X\cap Z|}{|X\cup Z|}+\frac{|Y\cap Z|}{|Y\cup Z|}=\frac{|X\cap Z|.|Y\cup Z|+|Y\cap Z|.|X\cup Z|}{|X\cup Z|.|Y\cup Z|}$$

$$(\ast)\quad \leq \frac{|Z|.\left[|X|+|Y|\right]}{|X\cup Z|.|Y\cup Z|}$$

$(\ast)$

$S = X\cup Z$

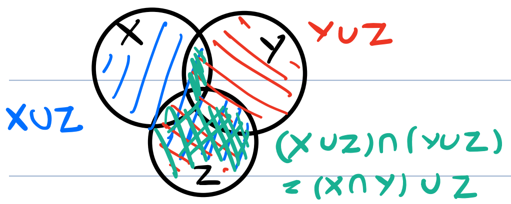
$T = Y\cup Z$

$$\leq \frac{|Z|.\left[|X|+|Y|\right]}{|(X\cup Z)\cap(Y\cup Z)|.|X\cup Y\cup Z|}$$

$X\cup Y \subseteq X\cup Y\cup Z$

donc $|X\cup Y| \leq |X\cup Y\cup Z|$

$$\leq \frac{|Z|}{|(X\cap Y)\cup Z|}\;\frac{|X|+|Y|}{|X\cup Y|}$$



$Z \subseteq (X\cap Y)\cup Z$

donc $|Z| \leq |(X\cap Y)\cup Z|$

$$\leq \frac{|X|+|Y|}{|X\cup Y|}$$

ce qui conclut la démonstration !